Compression des images et des vidéos numériques : dix années de recherches au CNRS

Marc Antonini

COMPRESSION DES IMAGES ET DES VIDEOS NUMERIQUES
DIX ANNEES DE RECHERCHES AU CNRS

Marc ANTONINI

Habilitation à Diriger des Recherches présentée à l'Université de Nice-Sophia Antipolis le 20 octobre 2003 devant le jury composé de

Claude Labit, ***Directeur de Recherche à l'INRIA,*** *président*
Robert M. Gray, ***Professeur à l'Université de Stanford,*** *Rapporteur*
Jean-Marc Chassery, ***Directeur de Recherche au CNRS,*** *Rapporteur*
Benoît Macq, ***Professeur à l'Université Catholique de Louvain-la-Neuve,*** *Rapporteur*
Michel Barlaud, ***Professeur à l'Université de Nice-Sophia Antipolis,*** *membre*

spécialité AUTOMATIQUE TRAITEMENT DU SIGNAL ET DES IMAGES
Université de Nice-Sophia Antipolis

Couverture dessinée par Anaïs Antonini

à Anaïs,
Renata, et mes parents

Mon travail de recherche a été réalisé au sein du laboratoire I3S, UMR 6070 du CNRS et de l'Université de Nice-Sophia Antipolis, dans l'équipe CReATIVe dirigée par le Professeur Michel Barlaud. Je tiens tout d'abord à remercier les différents directeurs qui ont dirigé ce laboratoire durant ces dix dernières années pour m'avoir fait bénéficier d'excellentes conditions de travail qui m'ont permis de mener à bien mes travaux de recherche.

Je tiens à remercier Monsieur Claude Labit, Directeur de Recherche à l'INRIA, de faire partie de mon jury et de me faire l'honneur d'en avoir accepté la présidence.

Monsieur Robert M. Gray, Professeur à l'Université de Stanford aux Etats-Unis, et spécialiste mondial en théorie de l'information m'a fait le grand honneur d'accepter de participer au jury et d'être rapporteur de mes travaux. Qu'il me soit ici permis de lui exprimer ma profonde reconnaissance.

Monsieur Jean-Marc Chassery, Directeur de recherche au CNRS, m'a fait le grand plaisir de juger mon travail de recherche. Je tiens à le remercier pour l'attention apportée lors de la lecture de ce document. J'ai véritablement apprécié ses remarques et ses conseils depuis plusieurs années, au cours de nombreuses discussions enrichissantes. Je lui sais gré d'avoir accepté d'être rapporteur de mon HDR.

Monsieur Benoît Macq, Professeur à l'Université Catholique de Louvain-la-Neuve, m'a fait la grande joie de juger mon travail et d'être un des rapporteurs. Je le remercie pour le temps qu'il a consacré à la lecture et l'analyse détaillée de ce document ainsi que pour avoir accepté spontanément cette tâche.

Je réserve une place particulière aux remerciements à l'attention de Monsieur Michel Barlaud, Professeur à l'Université de Nice-Sophia Antipolis. Michel m'a formé à la recherche depuis mon entrée en thèse en 1988. Je veux saluer ici la volonté et le dynamisme dont il fait preuve pour animer son équipe de recherche et je le remercie pour l'intérêt qu'il a toujours porté sur mon travail ainsi que pour la confiance qu'il m'a accordée tout au long de ces années. Qu'il sache que je lui en suis reconnaissant et que mon Amitié lui est acquise.

J'associe dans ces remerciements tous les membres du laboratoire I3S et particulièrement Madame Micheline Hagnéré, assistante de l'équipe CReATIVe, pour son devouement, son aide précieuse et son efficacité dans la résolution des problèmes administratifs de l'équipe.

Enfin, tout acte de recherche n'étant effectivement que la concrétisation personnelle d'un travail collectif et le fruit de nombreuses collaborations, je tiens à remercier ici toutes les personnes, collaborateurs, étudiants avec qui j'ai eu le plaisir de travailler de manière directe ou indirecte durant ces dix dernières années. Qu'ils y trouvent une expression de ma reconnaissance.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Avant-propos

J'ai été admis au CNRS en tant que Chargé de Recherche deuxième classe en Octobre 1993 au laboratoire I3S (Informatique Signaux et Système de Sophia Antipolis) à Sophia Antipolis, UMR 6070 du CNRS et de l'Université de Nice-Sophia Antipolis. J'ai été promu première classe en Octobre 1997. J'effectue ma recherche au sein de l'équipe CReATIVe (Compression Reconstruction Adaptées au Traitement des Images et des Vidéos - http ://www.i3s.unice.fr/~barlaud/Creative.htm) dirigée par le professeur Michel Barlaud.

Ce document comporte trois parties qui synthétisent environ dix années de mes recherches. J'ai voulu mettre en valeur mes activités de recherche et d'administration de la recherche principales en faisant apparaître les collaborations que j'ai pu avoir avec d'autres laboratoires en France ou à l'étranger, ainsi que les travaux des doctorants que j'ai eu l'occasion de co-encadrer. Il est structuré de la façon suivante. Dans la première partie du document je présente un rapport de mes activités nationales et internationales ainsi qu'une liste complète de mes publications. Dans la deuxième partie je développe les travaux de recherches les plus importants que j'ai effectué depuis mon entrée au CNRS. Enfin, dans la troisième partie sont insérées quelques unes de mes publications significatives utiles pour la compréhension de mes travaux.

Chapitre 2

Résumé synthétique de mes activités

Mots clés Images, images 3D, vidéos, multimédia, Internet, canal radio-mobile, compression, quantification vectorielle, codage à bas et très bas débit, contraintes entropiques, contraintes spatiales, ondelettes, multirésolution, schémas *lifting*, échantillonnage quinconce, optimisation de bancs de filtres, problèmes inverses, décodage optimal, filtrage non-linéaire, indexage, codage source/canal conjoint, images satellites, images médicales, images multispectrales, maillages 3D, compression géométrique.

Activités de recherche Mon domaine d'activité est le traitement du signal appliqué à l'image et mes travaux de recherche portent principalement sur un des domaines du traitement numérique de l'image : la compression et le codage. Parmi les points forts de mon activité de recherche, on peut souligner les travaux que j'ai menés en commun avec I. Daubechies (Princeton University – Etats-Unis). Ces travaux commencés pendant ma thèse ont abouti sur la construction des bases dites "9-7" qui fournissent à l'heure actuelle les meilleurs résultats en compression d'images. Ces filtres ont déjà fait l'objet d'une implantation sur circuit intégré commercialisé par Analog Device sous le nom de ADV601 ainsi que d'une implantation sur DSP par Texas Instrument. Ils sont de plus retenus par toutes les propositions de pointe pour la future norme de compression d'images fixes JPEG-2000. Ils constituent un des filtres de référence pour cette nouvelle norme. L'article IEEE Image Processing qui a résulté de nos travaux est actuellement parmi les plus cités dans le domaine de la compression des images (cité 408 fois en référence dans la base de données NEC en date du 22 septembre 2003 – cf. site http ://citeseer.nj.nec.com/cs).

Par la suite, et dans le cadre d'une application de compression d'images satellitaires suscitée par le Centre National d'Etudes Spatiales (CNES) de Toulouse, nous avons été amenés à développer un algorithme de transformation

multirésolution "au fil de l'eau". En effet, l'acquisition d'image par le satellite ainsi que la mémoire de masse présente à bord de celui-ci sont telles qu'il n'est pas envisageable de stocker entièrement une image acquise (de l'ordre de 24 000 pixels par ligne et de plusieurs centaines de lignes dans une fauchée). Ces travaux ont donnés lieu en 1995 à un algorithme de transformée en ondelettes au "fil de l'eau" capable de traiter des fauchées de taille infinie et permettant de générer une transformée en ondelettes exacte, comme si la totalité de l'image était connue. L'algorithme d'allocation des ressources binaires pour la compression que nous avons développé dans ce cadre multirésolution est basé sur des modèles théoriques de distorsion et de débit. Il permet soit un contrôle du débit binaire, soit un contrôle de la "qualité" image en terme d'erreur quadratique moyenne ou de rapport signal-à-bruit. La méthode de compression que nous avons proposée a été adaptée à la compression d'images satellites d'observation de la TERRE et a été retenue par le CNES pour être embarquée sur les futurs satellites d'observation de la génération PLEIADE. Un brevet est en cours de dépôt conjointement avec le CNES et le CNRS.

Cette méthode nous a servi comme brique de base pour construire un algorithme de codage source/canal par descriptions multiples efficace pour la transmission de données images ou vidéos sur des réseaux Internet ou encore de troisième génération (UMTS). De plus, nous avons montré qu'il était aussi efficace pour la compression de maillages 3D et pouvait concurrencer en terme de compromis DEBIT-DISTORSION-COMPLEXITE les meilleures méthodes de compression actuelles telles que le standard JPEG-2000. Parallèlement à cette approche scalaire, nous avons développé des travaux sur la quantification vectorielle par réseaux réguliers de points. Nos travaux se sont orientés suivant plusieurs objectifs et principalement, nous avons proposé des solutions pour le dénombrement des vecteurs dans un réseau régulier et pour l'indexage de ces vecteurs dans le cas d'une distribution Gaussienne généralisée.

L'aspect décodage est aussi très important. Les travaux que nous avons effectué sur ce sujet de recherche ont eu pour objectif de remettre en cause les filtres linéaires à reconstruction parfaite. Un point novateur de notre approche a été l'introduction dans la chaîne de traitement du bruit de quantification et du bruit électronique caractérisés par des bruits bornés non stationnaires. Contrairement à un post-traitement "classique", notre méthode de décodage permet la conservation du train binaire initial JPEG, M-JPEG ou MPEG1, MPEG2. Nous avons déposé récemment un brevet avec le CNRS sur la méthode développée.

L'aspect valorisation est aussi un point fort de mon activité de recherche. J'assure le transfert d'un savoir-faire technologique auprès de différents industriels (CNES, THALES, France Telecom, Opteway, IMSTAR) au travers de nombreux contrats.

Publications (157 papiers dont 43 IEEE) Les résultats obtenus sur les travaux que j'ai effectués depuis 1993 ont été publiés dans différentes revues internationales (IEEE, Electronic Letters, EURASIP JASP, Traitement du signal...), dans différents congrès nationaux et internationaux (IEEE ICIP, IEEE ICASSP, SPIE VCIP, GRETSI...) et dans divers séminaires et workshops nationaux et internationaux. Je totalise aujourd'hui 15 publications dans des revues spécialisées Internationales (dont 7 IEEE) et 2 publications soumises (2 IEEE), 4 chapitres d'ouvrage, 1 revue spécialisée, 4 conférences invitées (dont 1 IEEE), 2 brevets (dont 1 en cours de dépôt), 74 conférences avec comité de lecture (dont 33 IEEE), 22 conférences sans comité de lecture, 20 rapports de contrats, 13 rapports internes. Soit, un total de 155 papiers déjà parus (dont 41 IEEE) et 2 papiers soumis (2 IEEE).

Première partie

Mon rapport d'activités

Chapitre 1

Curriculum Vitae

Nom :	ANTONINI
Prénom :	Marc
Date de Naissance :	29 août 1965
Lieu :	Nice
Tél. Professionnel :	04-92-94-27-18
Email :	am@i3s.unice.fr
Nationalité :	Française

SITUATION PROFESSIONNELLE

Grade :	*Chargé de Recherche 1ère classe au CNRS* *Titulaire - Echelon 5 - Indice 672*
Section du comité national :	*07*
Intitulé :	*Sciences et Technologies de l'Information*
Date de nomination au CNRS :	*octobre 1993*
dans le présent grade :	*octobre 1997*
Unité d'affectation :	*Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S)* *UMR-6070* *CNRS-Université de Nice-Sophia Antipolis*
Directeur :	*Professeur J.M. Fedou*

FORMATION

Post-doctorat	*au CNES de Toulouse de 1991 à 1993*
Titres Universitaires	- *Thèse de doctorat en "Sciences de l'Ingénieur"* septembre 1991 - Université de Nice-Sophia Antipolis "Transformée en ondelettes et compression numérique des images"
	- *Diplôme d'Etudes Approfondies (DEA)* juin 1988 - Université de Nice-Sophia Antipolis

Chapitre 2

Activités d'encadrement et d'enseignement

2.1 Encadrement de doctorants (dix doctorants encadrés)

2.1.1 Thèses soutenues (six thèses soutenues)

J'ai co-encadré (à 50%) six étudiants en thèse à l'Université de Nice-Sophia Antipolis avec le Professeur M. Barlaud :

- *J.M. Moureaux* - années 1993-1994 (bourse NENRT)
  "Quantification vectorielle algébrique pour la compression d'images. Application à l'imagerie radar à synthèse d'ouverture (SAR)"
  Thèse soutenue le 2 décembre 1994.
  Aujourd'hui J.M. Moureaux est maître de conférence à l'Université de Nancy I.

- *P. Raffy* - années 1994-1997 (NENRT)
  "Modélisation, optimisation et mise en œuvre de quantificateurs bas débits pour la compression d'images utilisant une transformée en ondelettes"
  Thèse soutenue le 12 décembre 1997.
  Aujourd'hui P. Raffy est ingénieur dans la Silicon Valley aux Etats-Unis.

- *S. Tramini* - années 1996-1999 (bourse NENRT)
  "Problèmes inverses et EDP pour le décodage et la déconvolution d'images"
  Thèse soutenue le 29 novembre 1999.

- *J. Jung* - années 1997-2000 (bourse NENRT)
  "OMD : une méthode de décodage optimal orientée objet pour les séquences d'images"
  Thèse soutenue le 17 novembre 2000.
  Aujourd'hui J. Jung est ingénieur au Laboratoire d'Electronique de Philips (Philips Research France).

- *A. Gouze* - années 1998-2002 (bourse NENRT)
  "Schéma lifting quinconce pour la compression d'images"
  Thèse soutenue le 12 décembre 2002.
  Aujourd'hui A. Gouze est en post-doctorat à l'Université Catholique de Louvain-la-Neuve en Belgique.

- *C. Parisot* - années 1998-2003 (Financement CNES et CNRS)
  "Allocations basées modèles et transformée en ondelettes au fil de l'eau pour le codage des images et des vidéos"
  Thèse soutenue le 20 janvier 2003.

2.1.2 Thèses en cours (quatre en cours)

Je co-encadre actuellement (à 50%) quatre étudiants en thèse à l'Université de Nice-Sophia Antipolis avec le Professeur M. Barlaud :

- *M. Pereira* - depuis octobre 2000 (Bourse PRAXIS XXI - Portugal)
  "Compression par descriptions multiples pour la transmission d'images et de vidéos sur canaux bruités"

- *M. Cagnazzo* - depuis janvier 2003 (Bourse de l'Université Federico II - Italie)
  en collaboration avec le Professeur Poggi de l'Université de Naples
  "Compression de vidéos à très bas débit"

- *T. André* - à partir d'octobre 2003 (Bourse MESR)
  "Compression scalable de vidéos par transformée en ondelettes 3D au fil de l'eau compensée en mouvement"

J'encadre actuellement un étudiant en thèse à 100% grâce à une dérogation de l'Université de Nice-Sophia Antipolis :

- *F. Payan* - depuis octobre 2000 (Bourse Région/Entreprise - Opteway)
  "Compression multirésolution de maillages géométriques 3D"

2.2 Encadrement de post-doctorants (deux post - doctorants encadrés)

J'ai co-encadré (à 50%) deux post-doctorants avec le Professeur M. Barlaud :

- J.E. Fowler - année 1997 (Bourse post-doctorale de l'Université de l'Ohio USA)
  "Quantification vectorielle pour la compression de séquence d'images"
  Aujourd'hui, J.E. Fowler est "Associate Professor" à l'Université du Mississippi.

- S. Tramini - années 1999/2000 et 2000/2001 (Bourse post-doctorale du CNES de Toulouse)
  "Décodage optimal d'images satellitaires à haute résolution"

2.3 Encadrement de DEA et ingénieurs (quarante stagiaires encadrés)

J'ai encadré quinze étudiants de DEA, six stagiaires Européens (Projet ERASMUS avec l'Université Polytechnique de Catalogne [UPC] – Barcelone, Espagne) et dix neuf stagiaires d'écoles d'ingénieurs (Ecole Supérieure en Sciences Informatiques de Sophia Antipolis (ESSI), Ecole Supérieure d'Ingénieurs de Sophia Antipolis (ESINSA), Massachusetts Institute of Technology (MIT) aux Etats-Unis, Ecole Nationale Supérieure d'Hydraulique, d'Electronique et d'Informatique de Toulouse (E.N.S.H.E.I.T.) à Toulouse, Ecole Nationale Supérieure des Mines de Paris, Ecole Nationale Supérieure d'Ingénieurs de Sfax en Tunisie).

2.4 Activités d'enseignement

2.4.1 Participation à la formation doctorale STIC

- Je suis **co-responsable du DEA Image-Vision** de l'Université de Nice-Sophia Antipolis depuis le 01 octobre 2003.

- Je participe à l'enseignement dans le DEA Image-Vision de l'Université de Nice-Sophia Antipolis (UNSA) depuis 1993 : **6 heures de cours** par an dans le module "Compression des images".

2.4.2 Participation à l'enseignement en Ecoles d'Ingénieurs

Je suis responsable et j'ai monté les cours des enseignements suivants :

- "Technique de compression des images" en 3ième année du cycle d'ingénieur de l'Ecole Supérieure d'Ingénieur de Sophia Antipolis (ESINSA) de l'UNSA depuis 1995 : **18 heures de cours**, **15 heures de TD** et **15 heures de TP** par an. Ce module est commun avec le DEA SICOM de l'UNSA ;

- "Codage Source/Canal" en 3ième année du cycle d'ingénieur de l'Ecole Supérieure d'Ingénieur de Sophia Antipolis (ESINSA) depuis 2000 : **6 heures de cours** par an ;

- "Normes en compresssion d'images et de vidéos" à l'Institut Supérieur d'Informatique et d'Automatique (ISIA) de l'Ecole des mines de Paris depuis 2000 : **4 heures de cours** chaque deux ans ;

- "Compression des images et des vidéos" en 3ième année du cycle d'ingénieur de l'Ecole Supérieure en Sciences Informatiques (ESSI) de l'UNSA depuis 1997 (responsable depuis 2002) : **15 heures de cours** par an et **16 heures de TD**.

2.5 Participation à des jurys

J'ai été membre de 7 jurys de thèse en France et à l'étranger :

- J'ai été membre du jury de la thèse de Mr. J.M. Moureaux soutenue en Décembre 1994 à l'Université de Nice-Sophia Antipolis en France.

- J'ai été membre du jury et rapporteur de la thèse de Mme SHI Hongqin soutenue en novembre 1995 à l'Université Catholique de Louvain-la-Neuve en Belgique.

- J'ai été membre du jury de la thèse de Mr. P. Raffy soutenue en décembre 1997 à l'Université de Nice-Sophia Antipolis en France.

- J'ai été membre du jury de la thèse de Mr. S. Tramini soutenue en novembre 1999 à l'Université de Nice-Sophia Antipolis en France.

- J'ai été membre du jury de la thèse de Mr. J. Jung soutenue en novembre 2000 à l'Université de Nice-Sophia Antipolis en France.

– J'ai été membre du jury de la thèse de Mlle. A. Gouze soutenue en décembre 2002 à l'Université de Nice-Sophia Antipolis en France.

– J'ai été membre du jury de la thèse de Mr. C. Parisot soutenue en janvier 2003 à l'Université de Nice-Sophia Antipolis en France

J'ai également participé aux jurys de différents stagiaires de DEA (Image-Vision), d'Ecoles d'Ingénieur (ESSI, ESINSA, EURECOM) et de stagiaires ERASMUS.

Chapitre 3

Activités nationales et internationales

3.1 Participation à des contrats industriels

Les applications de compression/codage des images, liées à la transmission et à l'archivage, font l'objet à l'heure actuelle d'un enjeu industriel très important. Je participe à la mise en place de contrats avec différents partenaires industriels :

- **6 contrats avec le Centre National d'Etudes Spatiales (CNES) de Toulouse :**
  1. J'ai été co-responsable avec M. Barlaud d'un contrat avec le CNES Toulouse (N° 896/95/CNES/1379/00, 1995-1996) pour la "compression d'images satellites optiques à haute résolution".

     Ce travail avait pour but d'étudier la transformée en ondelettes et la quantification des coefficients d'ondelettes sous des contraintes d'implantation bord. Il a conduit à l'élaboration d'un algorithme "au fil de l'eau" pour l'analyse d'images par ondelettes [140], [141].
  2. J'ai été co-responsable avec M. Barlaud d'un contrat avec le CNES Toulouse (N° 896/CNES/96/0639, 1997-1998).

     Ce contrat constituait une suite logique à l'étude précédente. Cette nouvelle étude avait pour but l'évaluation de différentes tranformées en ondelettes dans le cadre de la compression d'images satellites optiques à haute résolution [143], [144].
  3. J'ai été co-responsable avec M. Barlaud d'un contrat avec le CNES Toulouse (N° 1/96/CNES/96/0761, 1999-2000) et en collaboration avec la société CRIL Technologies de Toulouse, pour la "compression d'images satellites à haute résolution".

Cette étude constituait la suite des deux études précédentes réalisées avec le CNES. Elle avait pour but d'optimiser l'algorithme de compression déjà développé pour le CNES et de l'adapter aux images "supermode" (échantillonnage quinconce) ainsi que d'améliorer les performances du codeur utilisé. L'algorithme développé devrait être utilisé pour les missions PLEIADE *après SPOT5. L'aspect "fil de l'eau" de l'algorithme a aussi fait l'objet d'une soumission au comité de normalisation JPEG-2000 (normalisation "grand public") ainsi qu'au CCSDS (normalisation spatiale)* [147], [151].

Un brevet relatif à ce procédé de compression va être déposé conjointement par le CNES et le CNRS.

4. J'ai participé à un contrat avec le CNES pour le "décodage optimal d'images SPOT" (N° 762/98/CNES/7422/00, 1998-2000).

   Cette étude consistait à remettre en cause les filtres linéaires à reconstruction parfaite utilisés de manière générale par les algorithmes lors de l'opération de décodage. Pour cela, nous avons étudié et proposé un algorithme de décodage adapté à la problématique globale posée : prise en compte de la chaîne globale de compression / décompression en intégrant la connaissance du système d'acquisition de l'image par le satellite (connaissance de la fonction de transfert de l'optique et du bruit de numérisation) [146].

5. J'ai été co-responsable avec M. Barlaud d'un contrat CNES Toulouse (avenant au contrat N° 1/96/CNES/96/0761, octobre 2000 - mars 2001), en collaboration avec la société CRIL Technologies de Toulouse.

   Ce contrat fait suite aux différentes études menées avec le CNES. Il a pour but de finaliser l'algorithme de compression "au fil de l'eau" que nous avons développé pour le CNES dans l'optique d'une intégration à bord des futurs satellites d'observation de la Terre, PLEIADE [153].

6. J'ai été co-responsable avec M. Barlaud d'un contrat CNES Toulouse (N° 713/01/ CNES/8710/00, novembre 2001), en collaboration avec la société CRIL Technologies de Toulouse.

   Ce contrat avait pour but d'évaluer l'asservissement (méthode pour la régulation de débit) et la quantification optimale avec zone morte pour la compression multirésolution PLEIADE [154], [155].

– **1 contrat avec ACATEL à Cannes :**

J'ai été co-responsable avec M. Barlaud d'un contrat avec ACATEL à Cannes, (N° UPM/3/98, 1997), sur "l'allocation de débits dans un schéma de compression d'images satellitaires multispectrales".

L'objet de cette étude était de déterminer la stabilité de l'algorithme de compression d'images satellitaires développé dans les contrats CNES, pour le codage d'images provenant de différentes bandes spectrales (images multi-spectrales et hyper-spectrales), mais représentant un type de scène identique [142].

– **1 contrat avec la société OPTEWAY à Sophia Antipolis :**

Je suis responsable d'un contrat avec la société Opteway à Sophia Antipolis dans le cadre de la bourse de thèse cofinancée Région/Entreprise de F. Payan, dont je suis le directeur de thèse (réf. CNRS 00052, octobre 2000 – octobre 2003).

Les objectifs de cette étude de thèse sont de définir une méthode de compression géométrique performante pour les maillages 3D.

– **2 contrats RNRT :**

1. Je suis responsable de la partie technique du projet exploratoire COSOCATI (COdage conjoint SOurce-CAnal pour la Transmission d'Images) dans le cadre du RNRT (décision no. 00S0016) sur une durée de 36 mois (mai 2000 – mai 2003).

   Les partenaires sont : l'ASPI, le CNES Toulouse, l'ENST Bretagne, l'ENST Paris, le L2S, France Telecom R&D et I3S.

   L'objectif de ce projet est la spécification détaillée d'un algorithme de codage source canal confirmé pour la transmission d'images fixes ou de séquences d'images. Les fonctions de codage source et de codage canal ne sont plus conçues séparément mais optimisées, ou adaptées, ou encore, effectivement conçues globalement, en tenant compte des caractéristiques précises de la source et du canal. Pour illustrer la validité de cette approche différents types d'applications sont envisagées en relation avec la transmission d'images dans le canal satellite et dans le canal radio-mobile [152], [156].

2. Je suis co-responsable avec M. barlaud du projet précompétitif EIRE (Etudes d'optImisations algoRithmiques de JPEG-2000) dans le cadre du RNRT sur une durée de 24 mois (novembre 2001 - novembre 2003).

   Les partenaires sont : THALES COMMUNICATIONS, IRCOMSIC, ENSTA, CRIL TECHNOLOGY, INRIA Rennes et I3S.

   L'objectif du projet est de proposer des améliorations au schéma de base de JPEG-2000, de réduire la complexité algorithmique, d'améliorer la qualité tant à l'encodage qu'au décodage, et d'assurer une

utilisation optimale de JPEG-2000, y compris dans un contexte d'interopérabilité avec d'autres schémas de codage. Ces travaux permettront de fournir la meilleure qualité de service dans un contexte limité en ressources (bande passante, capacité de traitement...), et de démontrer la possibilité d'implémenter efficacement JPEG 2000. Les applications visées concerneront principalement l'imagerie de type scientifique (satellite et médical) et la télésurveilance (dans un contexte de transmission vidéo sur IP) [157].

– **1 contrat RNTS :**

Je suis co-responsable avec M. barlaud du projet PATHNET (Un système automatique pour "microscopie virtuelle" et télétransmission à haut débit de lames histologiques à visée diagnostique en Anatomo-Cyto Pathologie) dans le cadre du RNTS sur une durée de 24 mois (2003 - 2005).

Les partenaires sont : IMSTAR, 3 services d'Anatomo-Cyto-Pathologie des Hôpitaux de l'Assistance Publique de Paris (La Pitié, Kremlin-Bicêtre, Henri Mondor) et un laboratoire Européen (Gratz, Autriche) et I3S.

Le développement technologique, réalisé en collaboration étroite avec IMSTAR, PME innovante, pilote du projet, devrait permettre de disposer d'un microscope dédié à "l'imagerie numérique virtuelle" et à la télétransmission sur ADSL, facile à utiliser et à un coût accessible pour les Anatomopathologistes du secteur publique et privé. Ce projet s'inscrit dans le cadre des objectifs prioritaires du Réseau National des Technologies de la Santé, sur deux thématiques d'innovation : la Télémédecine (Diagnostic à distance) et l'Imagerie Médicale (Histologique) de haute résolution spatiale.

– **3 actions soutenues par le GDR-PRC ISIS :**

1. J'ai participé au projet ACCORD financé par le CNET/CCETT (N° CNET 936B005, février 1993 – février 1995).

   Les partenaires sont : IRISA, LSS, TIMC, INSA-Creatis, IRESTE et I3S

   L'objectif de ce projet était l'élaboration d'une plateforme de tests en compression d'images [138], [139].

2. J'ai été responsable d'une Action Recherche/Industrie du GDR-PRC ISIS, Action CISAT (Compression d'images Satellitaires - http : // www-isis.enst.fr / NEW / Ori.html#CISAT) avec un financement du CNES de Toulouse (N° 713 / CNES / 97 / 6966 / 00). Cette étude avait une durée de 18 mois (mars 1998 – juillet 1999).

Les partenaires sont : I3S, le CRAN Nancy.

L'objet de l'étude consistait à évaluer les performances en terme de qualité / complexité de différentes approches de quantification vectorielle, pour la compression d'images satellite. Nous avons envisagé deux grandes classes de quantificateurs : la quantification vectorielle par apprentissage et la quantification vectorielle algébrique. Ces méthodes ont été comparées à la méthode référence du CNES définie pour les compresseurs PLEIADE *après SPOT5* [145], [148], [149], [150].

3. J'ai été responsable d'une Action GDR-PRC ISIS sur la "compression de séquences d'images pour la transmission de vidéo interactive sur ADSL" (2000-2001 - http : // www-isis.enst.fr / Kiosque / FourreTout / resultat_appel projets.html).

   Les partenaires sont : I3S, le CRAN Nancy.

   Le but principal est de proposer une chaîne de compression efficace (ou un ensemble d'algorithmes) permettant d'effectuer de la vidéo interactive sur support ADSL et qui offre une bonne qualité image pour des débits de transmission inférieurs à 8 Mbits/s. Le terme vidéo interactive est pris ici au sens large ; il s'agit de toutes les applications de type vidéo avec interactivité du système récepteur par rapport au système émetteur [80], [123].

– **1 action Télécom soutenue par le CNRS :**

J'ai été responsable d'une Action Télécom CNRS : "Codage Source / Canal conjoint par descriptions multiples pour la transmission d'images et de vidéos sur le réseau Internet" (N° TL 99012). Cette étude avait une durée 24 mois (novembre 1999 – novembre 2001).

Ce projet, qui s'inscrit dans le cadre du codage conjoint source/canal, a pour but le développement des applications multimédias aux usagers.

– **2 projets européens :**

1. J'ai participé à une action du Projet Européen SUMARE (Survey of Marine Resources - http ://www.mumm.ac.be/SUMARE/) en collaboration avec le projet SAM du laboratoire I3S (Proposal number IST-1999-10836, 2000-2001)

   L'objectif du projet SUMARE est la démonstration de l'intérêt de capteurs autonomes intelligents (des plateformes robotiques autonomes équipées de capteurs) dans le contexte de la surveillance et la gestion de ressources naturelles, en particulier de ressources sous-marines. Le but principal de l'étude est de réaliser la segmentation

d'images de fonds sous-marins acquises par un robot, de façon à extraire la ligne de séparation entre une zone de corail vivant et une zone de corail mort et permettre ainsi le guidage du robot pour le repérage de ces différentes zones.

2. Je participe au Réseau d'Excellence Européen SCHEMA ("Network of Excellence in Content-Based Semantic Scene Analysis and Information Retrieval") en collaboration avec l'Université de Tampéré en Finlande, l'Université de Louvain-la-Neuve en Belgique, l'Université Polytechnique de Catalogne en Espagne, l'Université Queen Mary & Westfield College de Londres en Angleterre, L'Ecole Polytechnique Fédérale de Lausanne en Suisse, l'Université de Munich en Allemagne et l'Université de Thessalonique en Grèce.

   L'objectif du réseau SCHEMA est de permettre l'échange de chercheurs et de doctorants entre ces différents laboratoires et d'effectuer ainsi un transfert de savoir faire.

Depuis mon entrée au CNRS j'ai travaillé sur plusieurs contrats industriels et projets. Le montant total de ces contrats et projets sur la période 1995-2003 s'élève à 736 646 euros HT.

3.2 Valorisation : dépot de brevets

– Les activités de recherches relatives à la thèse de J. Jung que j'ai coencadrée avec M. Barlaud ont permis de déposer, avec le CNRS, un brevet dont je suis l'un des coauteurs avec J. Jung et M. Barlaud. Le titre de ce brevet est "Décodeur vidéo optimal basé sur les standards de type MPEG" [22], il a été déposé le 11 juin 1999 :

  - en France, numéro 99/07443 ;
  - aux Etats-Unis, numéro 09/406,673.

  Et étendu le 09 juin 2000

  - en Europe, numéro 00940488.0 ;
  - au Canada, numéro PCT/FR00/01613 ;
  - au Japon, numéro PCT/FR00/01613.

– Suite aux activités de recherche et développement, que nous avons eu dans le cadre de différents contrats, nous sommes en train de déposer un brevet avec le Centre National d'Etudes Spatiales (CNES) de Toulouse et le CNRS [23]. Ce brevet est relatif à un procédé d'allocation de débit dynamique "temps réel" pour des images satellitaires d'observation de la Terre et pour des systèmes embarqués à bord de satellites.

3.3 Collaborations

3.3.1 Par le GDR-PRC ISIS du CNRS

La participation au GDR Traitement du Signal et Images du CNRS permet des échanges d'idées et de points de vue avec d'autres laboratoires nationaux ainsi qu'avec des industriels, de ce fait elle est très importante.

Je participe essentiellement dans le thème D (Télécommunications : compression, transmission, protection). Depuis 1993, nos travaux ont fait l'objet de plusieurs présentations au GDR-PRC ISIS Traitement du Signal et Image [103], [106], [105], [107], [109], [108], [113], [115], [114], [117], [116], [121], [124], [123].

Nous avons monté récemment une Action Spécifique AS 190 (RTP 25 et GDR ISIS) intitulée "Compression scalable et robuste de signaux vidéos" en collaboration avec l'IRISA, le LABRI et l'ENST.

3.3.2 Au niveau national

- Collaboration avec C. LABIT (IRISA - Rennes) pour divers aspects relatifs à la compression d'images fixes et de séquences d'images [138], [139] ;

- Collaboration avec J.M. CHASSERY (LIS - Grenoble) pour la compression d'images par fractals [5], [108] ;

- Collaboration avec P. SOLE (I3S – Université de Nice-Sophia Antipolis) pour la quantification vectorielle d'images par réseaux réguliers de points [3], [37], [38], [40], [125] ;

- Collaboration avec J.M. MOUREAUX (CRAN/CNRS Université de Nancy 1) sur divers points de recherche liés à la quantification vectorielle. Principalement sur :
  - l'indexage de vecteurs dans les réseaux réguliers de point pour la transmission de données image et l'implantation de l'algorithme sur DSP [6], [7], [13], [50], [55], [68], [114], [134] ;
  - la quantification vectorielle dans le cadre d'une action Recherche/Industrie du GDR-PRC ISIS en partenariat avec le CNES (action CISAT) [145], [148], [149], [150] ;
  - la compression de vidéos interactives pour leur transmission sur lignes téléphoniques par modulation ADSL dans le cadre d'une action du GDR-PRC ISIS [80], [123].

- Collaboration avec P. LOYER (ALCATEL à Cannes) sur les problèmes théoriques d'indexation de vecteurs dans les réseaux réguliers de point avec une application à la compression des images [6], [13], [50], [68], [134] ;

– Collaboration avec G. AUBERT (laboratoire Dieudonné de l'Université de Nice-Sophia Antipolis) sur les problèmes inverses pour le décodage optimal d'images [71].

3.3.3 Au niveau Européen

– Une collaboration Européenne soutenue par le **CNRS** et **NWO** a été engagée en 1990/1993 avec le Professeur J. BIEMOND (Université de DELFT aux Pays-Bas). Les thèmes de recherche en traitement des images de son laboratoire concernent en particulier les aspects de compression de séquences d'images en vue d'une application pour la Télévision à Haute Définition (TVHD). En effet, les recherches sur la TVHD représentent une principale motivation Européenne à l'heure actuelle. Cette collaboration a permis un échange de chercheurs entre les deux laboratoires.

– Une collaboration Européenne soutenue par le **CNRS** et **FNRS** a été engagée en 1994 avec le Professeur B. MACQ (Laboratoire de Télécommunication et de Télédétection de l'Université de Louvain la Neuve en Belgique).
J'ai accueilli le Professeur B. Macq dans l'équipe CReATIVe durant son séjour de 1 mois en Avril 2000.
Les thèmes de recherche de son équipe concernent la compression d'images et de vidéos et les bancs de filtres numériques.

– Nous avons mis en place un réseau d'excellence Européen (SCHEMA) regroupant le laboratoire I3S, l'Université de Tampéré en Finlande, l'Université de Louvain-la-Neuve en Belgique, l'Université Polytechnique de Catalogne en Espagne, l'Université Queen Mary & Westfield College de Londres en Angleterre, L'Ecole Polytechnique Fédérale de Lausanne en suisse, l'Université de Munich en Allemagne et l'Université de Thessalonique en Grèce. L'objectif est de permettre l'échange de chercheurs et de doctorants entre ces différents laboratoires.

3.3.4 Au niveau international

– J'ai été co-responsable avec M. Barlaud d'une collaboration internationale soutenue par le **CNRS** et **NSF** (Etats-Unis) engagée en 1994/1997 avec le Professeur N. FARVARDIN (Université de Maryland aux Etats-Unis).
J'ai accueilli le Professeur N. Farvardin durant plusieurs de ses séjours dans l'équipe CReATIVE et j'ai visité son laboratoire en septembre 1994.
Les recherches de l'équipe du Professeur N. Farvardin sont principalement orientées sur le codage de source et de canal.

- J'ai été co-responsable avec M. Barlaud d'une collaboration internationale soutenue par le **CNRS** et **NSF** (Etats-Unis) engagée en 1997/2000 avec le Professeur R.M. GRAY (Université de Stanford aux Etats-Unis).
  J'ai accueilli le Professeur R.M. GRAY dans l'équipe CReATIVe, 1 semaine en 1997, une semaine en 1999 et une semaine en 2001.
  Le professeur R.M. Gray est mondialement connu dans la communauté traitement du signal et des images comme étant le principal inventeur de la quantification vectorielle pour la compression des données.

- J'ai été co-responsable avec M. Barlaud d'une collaboration avec le Professeur J. Vaisey (Université Simon Fraser au Canada).
  Le Professeur J. Vaisey a effectué un séjour de 6 mois, en tant que "Professeur Invité", dans l'équipe CReATIVe en 1997/1998.
  Ces activités de recherche concerne la compression des images.

- J'ai été responsable d'une collaboration internationale soutenue par le **CNRS** et le **CNPq** (Brésil) engagée depuis septembre 2000 pour une durée de 2 ans avec le Professeur P.RAMIREZ DINIZ (Université Fédérale de Rio de Janeiro - UFRJ). La dotation était de 1 voyage avec séjour de 4 semaines par an pour chacun des deux partenaires.
  Les travaux que nous avons entrepris concernent l'introduction de filtres ondelettes perceptuels par "schéma lifting" pour la compression orientée MPEG-4 de vidéos.

- Nous avons démarré une collaboration internationale soutenue par le **CNRS** et **NSF** (Etats-Unis) depuis 2003 pour une durée de 3 ans avec le Professeur J. KONRAD de l'Université de Boston (Etats-Unis).
  J'ai accueilli le Professeur J. Konrad durant son séjour dans l'équipe CReATIVe en juillet 2003.
  Nous avons entrepris des travaux sur la transformée en ondelette par schéma lifting compensé en mouvement.

- Nous avons mené des études en commun avec I. DAUBECHIES (Princeton University aux Etats-Unis). Nos travaux ont porté sur la transformée en ondelettes et ses applications en codage d'image. Ces travaux ont conduit en 1992 à la création des filtres ondelettes dits "9-7" qui aujourd'hui sont les filtres utilisés par la nouvelle norme de compression des images fixes numériques, JPEG-2000 (http ://jj2000.epfl.ch) [2], [31].

3.4 Activité internationale

J'ai participé à de nombreux congrès, workshops et séminaires en France. Ils ne sont pas présentés ici, mais les différentes publications relatives à ces

congrès et workshops peuvent être trouvées dans la liste des publications jointe en annexe. Les congrès ainsi que les séjours à l'étranger sont importants dans l'activité d'un chercheur car ils permettent de faire connaître mondialement ses activités scientifiques. Ils permettent aussi de nouer de nouveaux contacts en vue de nouvelles collaborations internationales. C'est ce qui fait aujourd'hui la renommée mondiale du CNRS. C'est dans cet objectif que j'ai effectué différents séjours dans des pays étrangers, soit dans le cadre de congrès, soit dans le cadre de collaborations entre laboratoires.

3.4.1 Participation à des congrès internationaux

- Mission en octobre 1990 en Espagne : présentation au congrès international EUSIPCO (European Signal Processing Conference) à Barcelone [32] ;
- Mission en mai 1991 au Canada : présentation au congrès international IEEE ICASSP (International Conference on Acoustics, Speech, and Signal Processing) à Toronto [34], [35] ;
- Mission en mars 1992 aux Etats-Unis : présentation au congrès international IEEE ICASSP (International Conference on Acoustics, Speech, and Signal Processing) à San Francisco [38] ;
- Mission en novembre 1992 aux Etats-Unis : présentation au congrès international SPIE VCIP (Visual Communication and Signal Processing) à Boston [39] ;
- Mission en septembre 1994 aux Etats-Unis : présentation au congrès international SPIE VCIP (Visual Communication and Image Processing) à Chicago [45].
- Mission en octobre 1995 aux Etats-Unis : présentation au congrès IEEE ICIP (International Conference on Image Processing) à Washington, DC [47] ;
- Mission en septembre 1996 en Suisse : présentations au congrès IEEE ICIP (International Conference on Image Processing) à Lausanne [50], [51] ;
- Mission en octobre 1997 aux Etats-Unis : présentation au congrès IEEE ICIP (International Conference on Image Processing) à Santa Barbara, Californie [56] ;
- Mission en juillet 2000 aux Etats-Unis : présentation invitée dans une session spéciale au congrès international IEEE IGARSS (International Geoscience and Remote Sensing Symposium) à Honolulu, Hawaii [100] ;
- Mission en octobre 2001 en Grèce : présentations au congrès IEEE ICIP (International Conference on Image Processing) à Thessalonique [75], [76] ;
- Mission en avril 2003 en France : présentations au congrès PCS (Picture Coding Symposium) à Saint-Malo [89], [90].

- Mission en juillet 2003 au Portugal : présentations au congrès IEEE HSNMC (High Speed Networks and Multimedia Communications) à Estoril [92], [93]. *J'ai été à cette occasion "chaiman" de la session "video".*
- Mission en septembre 2003 en Espagne : présentations au congrès IEEE ICIP (International Conference on Image Processing) à Barcelone [94], [95], [96].

3.4.2 Séjours à l'étranger

- Visite en mars 1992 à l'Université de Stanford de l'équipe du Professeur R.M. GRAY à l'occasion du congrès ICASSP 1992 à San Francisco (Etats-Unis).
- Séjour en septembre 1994 à l'Université de Maryland à Washington (Etat-Unis) : séjour d'une semaine dans l'équipe du Professeur N. FARVARDIN dans le cadre de la collaboration CNRS/NSF.
- Séjour en janvier 1995 à Louvain la Neuve en Belgique : séjour d'une semaine dans l'équipe du Professeur B. MACQ (laboratoire de télécommunication et de télédétection de l'Université de Louvain la Neuve) dans le cadre de la collaboration CNRS/FNRS.
- Séjour en février 2001 à Rio de Janeiro au Brésil : séjour de trois semaines dans l'équipe du Professeur P. RAMIREZ DINIZ (Université Fédérale de Rio de Janeiro – UFRJ) dans le cadre de la collaboration CNRS/CNPq.

3.4.3 Accueil de chercheurs étrangers

- Professeur R.M. Gray (Université de Stanford - Etats-Unis) 3 séjours d'une semaine entre 1997 et 2001 ;
- Professeur N. Farvardin (Université du Maryland – Etats-Unis) plusieurs séjours courts (3 jours) entre 1994 et 2000 ;
- Professeur J. Vaisey (Université Simon Fraser – Canada) 6 mois en 1998
- Professeur J. Fowler (Université du Mississippi – Etats-Unis) 3 semaines en 1998 ;
- Professeur B. Macq (Université Catholique de Louvain-La-Neuve – Belgique) 1 mois en 2000 ;
- Professeur A. Hero (Université du Michigan – Etats-Unis) 1 mois en 2001 ;
- Professeur J. Konrad (Université de Boston - Etats-Unis) 10 jours en 2003.

3.5 Valorisation culturelle

3.5.1 Rédaction de chapitres de livres

Je suis le coauteur de 4 chapitres de livres :

- Chapitre : "Digital image compression using vector quantization and the wavelet transform"
  Livre :"Wavelets and Applications : Proceedings of the International Conference, Marseille 1989" [24].
  Editeur : Yves Meyer, Editions Masson, Paris 1992, pp.160-174.
  Coauteurs : M. Antonini, M. Barlaud, P. Mathieu.
- Chapitre : "Wavelet transform and image coding" [19]
  Livre : "Wavelet in Image Communication"
  Editeur : M. Barlaud, Editions Elsevier, North Holland, Vol.5 1994, pp.65-188.
  Coauteurs : M. Antonini, T. Gaidon, M. Barlaud, P. Mathieu.
- Chapitre : "Quantification" [20].
  Livre : "Compression des Images et Vidéos" du Traité IC2.
  Editeurs : M. Barlaud, C. Labit, Edition Hermès, janvier 2002, pp.45-72.
  Coauteurs : M. Antonini, V. Ricordel.
- Chapitre : "Transformée en ondelettes" [21].
  Livre : "Compression des Images et Vidéos" du Traité IC2.
  Editeurs : M. Barlaud, C. Labit, Edition Hermès, janvier 2002, pp.73-96.
  Coauteurs : M. Barlaud, M. Antonini.

3.5.2 Publications et conférences vulgarisatrices

Nos travaux ont fait l'objet :

- d'un article dans le "Courrier du CNRS" numéro 77 en juin 1991, pp.50-51 [24] ;
- d'une présentation dans une conférence vulgarisatrice du comité technique de l'image et du son (CST) à La Villette à Paris le 16 février 1999 [99].

3.6 Administration de la recherche

3.6.1 Responsabilités au sein du GDR-PRC ISIS

- Je suis rédacteur en chef de la Gazette du GDR-PRC ISIS depuis juillet 2000 (http :// www - isis.enst.fr/Gazette/).

La gazette est destinée à véhiculer sous forme condensée les informations présentes sur le site du GDR et constitue ainsi une alternative à celui-ci : elle s'adresse aux personnes qui ne souhaitent pas être submergées par le flot des informations et/ou qui n'ont pas le temps de les consulter. Elle est publiée au rythme d'une par an ;

– Je suis **co-responsable du thème D** (Télécommunications : compression, transmission, protection) du GDR-PRC ISIS depuis mai 2003 (http ://www-isis.enst.fr/Membre2/structure.php ?sub=TH)

3.6.2 Responsabilités internationales

– J'ai participé à l'organisation du workshop IEEE MultiMedia Signal Processing (MMSP) qui s'est déroulé à Cannes du 03 au 05 octobre 2001 (http ://mmsp01.eurecom.fr/).

– J'ai été "chairman" de la session "vidéo" organisée au congrès IEEE HSNMC (High Speed Networks and Multimedia Communications) qui s'est déroulé à Estoril au Portugal du 23 au 25 juillet 2003.

3.6.3 Review d'articles

Je suis reviewer pour différentes revues :

– la revue IEEE Transactions on Signal Processing ;
– la revue IEEE Transactions on Image Processing ;
– la revue IEEE Transactions on Information Theory ;
– la revue IEE Electronics Letters ;
– la revue Traitement du Signal ;
– la revue Signal Processing de Elsevier Science Publisher.
– le National Science Foundation (aux USA) ;

Je participe aux comités scientifiques de différents congrès :

– IEEE ICIP : International Conference on Image Processing ;
– IEEE ICASSP : International Conference on Acoustics, Speech, and Signal Processing ;
– PCS : Picture Coding Symposium

3.6.4 Activités au sein du laboratoire I3S UMR 6070

– Je suis membre de la Commission de Spécialistes (CS) de la 61ième section de l'Université de Nice-Sophia Antipolis ;

– Je suis membre du “Comité des Projets” du laboratoire Informatique Signaux et Systèmes de Sophia Antipolis (I3S) UMR 6070 auquel j'appartiens. Ce comité à pour rôle de prendre des décisions d'ordre scientifiques pour le bon fonctionnement du laboratoire.

Chapitre 4

Liste de mes publications

Publications avec comité de lecture

[1] P. Mathieu, M. Barlaud, and M. Antonini, "Compression d'Image par transformée en ondelette et quantification vectorielle," *Traitement du Signal*, vol. 7, pp. 101–115, juin 1990.

[2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, avril 1992.

[3] M. Barlaud, P. Solé, T. Gaidon, M. Antonini, and P. Mathieu, "Lattice vector quantization for multiscale image coding," *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 367–381, juillet 1994.

[4] J.M. Moureaux, M. Antonini, and M. Barlaud, "Counting lattice points on ellipsoids : Application to image coding," *Electronics Letters*, vol. 31, no. 15, pp. 1224–1225, juillet 1995.

[5] F. Davoine, M. Antonini, J.M. Chassery, and M. Barlaud, "Fractal image compression based on delaunay triangulation and vector quantization," *IEEE Transactions on Image Processing, special issue on vector quantization*, vol. 5, no. 2, pp. 338–346, février 1996.

[6] J.M. Moureaux, P. Loyer, and M. Antonini, "Low complexity indexing method for z^n and d_n lattice quantizers," *IEEE Transactions on Communications*, vol. 46, no. 12, pp. 1602–1609, décembre 1998.

[7] J.M. Moureaux, P. Nus, and M. Antonini, "A multi-DSP architecture for real-time lattice quantization indexing," *SPIE Journal of Electronic Imaging, special section on image/video compression and processing for visual communications*, 1998.

[8] P. Raffy, M. Antonini, and M. Barlaud, "Optimal subband bit allocation procedure for very low bit rate image coding," *Electronic Letters*, vol. 34, no. 7, pp. 647–648, avril 1998.

[9] S. Tramini, M. Antonini, and M. Barlaud, "Intraframe image decoding based on a nonlinear variational approach," *International Journal of Imaging Systems and Technology*, vol. 9, no. 5, pp. 369–380, octobre 1998.

[10] J. Jung, M. Antonini, and M. Barlaud, "A new approach for video sequence restoration," *Annales des Télécommunications*, vol. 55, no. 3-4, pp. 101–107, mars-avril 2000.

[11] P. Raffy, M. Antonini, and M. Barlaud, "Distortion-rate models for entropy coded lattice vector quantization," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2006–2017, décembre 2000.

[12] C. Parisot, M. Antonini, and M. Barlaud, "3d scan based wavelet transform and quality control for video coding," *EURASIP Journal on Applied Signal Processing, Special issue on Multimedia Signal Processing*, vol. 2003, no. 1, pp. 521–528, janvier 2003.

[13] P. Loyer, J.M. Moureaux, and M. Antonini, "Lattice codebook enumeration for generalized gaussian source," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 521–528, février 2003.

[14] J. Jung, M. Antonini, and M. Barlaud, "Optimal decoder for block-transform based video coders," *IEEE Transactions on Multimedia*, vol. 5, no. 2, pp. 145–160, juin 2003.

[15] M. Pereira, M. Antonini, and M. Barlaud, "Multiple description image and video coding for wireless channels," *Session spéciale EURASIP : Image Communication Special Issue on Recent Advances in Wireless Video*, vol. 18, no. 10, pp. 925–945, novembre 2003.

[16] A. Gouze, M. Antonini, M. Barlaud, and B. Macq, "Signal-adapted multidimensionnal lifting scheme," *en 2nd review dans la revue IEEE Transactions on Image Processing*, mai 2003.

[17] F. Payan and M. Antonini, "Mean square error for biorthogonal m-channel wavelet coder," *soumis à IEEE Transactions on Image Processing*, décembre 2003.

Publications dans des ouvrages de synthèse

[18] M. Antonini, M. Barlaud, and P. Mathieu, *Digital Image Compression Using Vector Quantization and the Wavelet Transform*, pp. 160–174, Masson, Paris, 1989.

[19] M. Antonini, T. Gaidon, M. Barlaud, and P. Mathieu, *Wavelet Transform and Image Coding*, vol. 5, Elsevier, North Holland, 1994.

[20] M. Antonini and V. Ricordel, *Chapitre "Quantification"*, pp. 45–72, Traité IC2. Hermès, Paris, janvier 2002.

[21] M. Barlaud and M. Antonini, *Chapitre "Transformée En Ondelettes"*, pp. 73–96, Traité IC2. Hermès, Paris, janvier 2002.

Valorisation

[22] M. Antonini, M. Barlaud, and J. Jung, "Décodeur vidéo optimal basé sur les standards de type MPEG," BREVET no. 99/07443 (France), no. 09/406,673 (Etats-Unis), 11 juin 1999 et extension no. 00940488.0 (Europe), no. PCT/FR00/01613 (Canada) et no. PCT/FR00/01613 (Japon) le 9 juin 2000.

[23] M. Antonini, C. Parisot, M. Barlaud, and C. Lambert-Nebout, "Une méthode rapide de calcul du paramètre de lagrange pour les problèmes d'allocation de débit et de qualité en compression d'images et de vidéos par transformée en ondelettes," 2003, BREVET en cours de dépot.

Publications dans des revues spécialisées

[24] M. Antonini, M. Barlaud, and P. Mathieu, *Ondelettes et Compression Numérique Des Images*, pp. 50–51, Number 77. Courrier du CNRS, Paris, 1991.

Communications dans des conférences avec comité de lecture

[25] M. Barlaud, L. Blanc-Féraud, P. Mathieu, J. Menez, and M. Antonini, "2d linear predictive image coding with vector quantization," in *"Signal Processing IV : Theories and Applications", Vol.2, Elsevier Science publisher, EUSIPCO*, Grenoble, France, 5-8 septembre 1988, pp. 1637–1640.

[26] M. Barlaud, P. Mathieu, L. Blanc-Féraud, J. Menez, and M. Antonini, "Predictive image coding with vector quantization using mirror images," in *IEEE International Symposium on Information Theory*, Kobe, Japon, 19-24 juin 1988, pp. 203–204.

[27] P. Mathieu, M. Barlaud, and M. Antonini, "Compression d'Images par transformée en ondelette," in *12ème Colloque GRETSI*, Juan-les-Pins, 12-16 juin 1989, pp. 781–784.

[28] M. Antonini, M. Barlaud, and P. Mathieu, "Codebook optimal et nouvelle stratégie de quantification vectorielle d'image," in *12ème Colloque GRETSI*, Juan les Pins, France, 12-16 juin 1989, pp. 605–608.

[29] M. Antonini, M. Barlaud, and P. Mathieu, "Compression numérique des images par quantification vectorielle dans l'Espace des transformées

en ondelette," in *Wavelet and some of their Applications, Workshop Marseille-Luminy*, 29 mai - 3 juin 1989.

[30] M. Barlaud, P. Mathieu, and M. Antonini, "Wavelet transform image coding using vector quantization," in *Proc. Of the 6th IEEE Multidimensional Signal Processing Workshop*, Monterey California, Etats-Unis, septembre 1989, pp. 103–104.

[31] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using vector quantization in the wavelet transform domain," in *Proc. Of the IEEE Int. Conf. On Acoust., Speech and Signal Processing*, Albuquerque, New Mexico, Etats-Unis, 1990, vol. 4, pp. 2297–2300.

[32] M. Antonini, M. Barlaud, and P. Mathieu, "Predictive interscale image coding using vector quantization," in *"Signal Processing V : Theories and Applications", Elsevier Science Publisher, EUSIPCO*, Barcelone, Espagne, 1990, pp. 1091–1094.

[33] M. Antonini, M. Barlaud, P. Mathieu, and J.C. Feauveau, "Multiscale image coding using the kohonen neural network," in *Visual Communication and Image Processing*, Lausanne, Suisse, 1-4 octobre 1990, pp. 14–26.

[34] M. Antonini, M. Barlaud, and P. Mathieu, "Image coding using lattice vector quantization of wavelet coefficients," in *Proc. Of the IEEE Int. Conf. On Acoust., Speech and Signal Processing*, Toronto, canada, 1991, vol. 4, pp. 2273–2276.

[35] J.C. Feauveau, P. Mathieu, M. Barlaud, and M. Antonini, "Recursive biorthogonal wavelet transform for image coding," in *Proc. Of the IEEE Int. Conf. On Acoust., Speech and Signal Processing*, Toronto, canada, 1991, vol. 4, pp. 2649–2652.

[36] M. Antonini, M. Barlaud, and P. Mathieu, "Image coding using lattice vector quantization of wavelet coefficients," in *Picture Coding Symposium, VISICOM*, Tokyo, JAPON, 2-4 septembre 1991, p. 155.

[37] M. Barlaud, M. Antonini, P. Mathieu, and P. Solé, "Entropy-constrained lattice VQ for image coding using wavelet transform," in *Workshop on Computer Vision and Image Processing for Spaceborn Applications, ESA*, Noordwijk, juin 1991.

[38] M. Barlaud, P. Solé, M. Antonini, and P. Mathieu, "A pyramidal scheme for lattice vector quantization of wavelet transform coefficients applied to image coding," in *Proc. Of the IEEE Int. Conf. On Acoust., Speech and Signal Processing*, San Francisco, California, Etats-Unis, 1992, vol. 4, pp. 401–404.

[39] M. Antonini, M. Barlaud, and T. Gaidon, "Adaptive entropy-constrained lattice vector quantization for multiresolution image co-

ding," in *Visual Communication and Image Processing*, Boston, Massachusetts, Etats-Unis, 18-20 novembre 1992, pp. 441–457.

[40] M. Barlaud, P. Solé, J.M. Moureaux, M. Antonini, and P. Gauthier, "Elliptical codebook for lattice vector quantization," in *Proc. Of the IEEE Int. Conf. On Acoust., Speech and Signal Processing*, Minneapolis, Minnesota, Etats-Unis, 1993, vol. 5, pp. 590–593.

[41] M. Antonini, M. Barlaud, B. Rougé, and C. Lambert-Nebout, "Allocation optimale des débits binaires pour le codage multirésolution d'Images satellites," in *14ième Colloque GRETSI sur Le Traitement Du Signal et Des Images*, Juan les Pins, France, 13-16 septembre 1993, pp. 455–458.

[42] M. Antonini, M. Barlaud, B. Rougé, and C. Lambert-Nebout, "Weighted optimum bit allocation for multiresolution satellite image coding," in *CCECE*, Vancouver, Canada, 14-17 septembre 1993.

[43] M. Barlaud, M. Antonini, and L. Blanc-Féraud, "Ondelettes et multirésolution en traitement d'Images," in *Colloque TOM*, Lyon, France, 9-11 mars 1994, pp. 2.1–2.10.

[44] J.M. Moureaux, M. Antonini, and M. Barlaud, "Codebook design for elliptical source statistics in image coding," in *Picture Coding Symposium*, 25-28 septembre 1994.

[45] J.M. Moureaux, M. Antonini, and M. Barlaud, "Lattice vector quantization of image using a product-code form and a new labelling method," in *SPIE, Visual Communication and Image Processing*, Chicago, Etats-Unis, 25-28 septembre 1994, pp. 422–433.

[46] J. Roca Oliver, M. Antonini, and M. Barlaud, "Optimisation de quantificateurs vectoriels à entropie contrainte pour le codage d'Images multirésolution," in *15ième Colloque GRETSI sur Le Traitement Du Signal et Des Images*, Juan-les Pins, 18-22 septembre 1995, pp. 733–736.

[47] M. Antonini, P. Raffy, and M. Barlaud, "Towards entropy constrained lattice vector quantization for image coding," in *IEEE International Conference on Image Processing (ICIP)*, Washington, DC, Etats-Unis, 22-25 octobre 1995, pp. 121–124.

[48] P. Raffy, M. Barlaud, and M. Antonini, "Quantificateurs à contrainte spatiale pour le codage d'Images multirésolution," in *2ièmes Journées d'Etudes et d'Echanges "COmpression et REprésentation Des Signaux Audiovisuels (CORESA)*, CNET Grenoble, France, 15-16 février 1996, pp. 115–120.

[49] P. Raffy, M. Barlaud, and M. Antonini, "A new model involving spatial constraints for edge preserving quantization," in *Ninth Image and Multidimensional Signal Processing Workshop (IMDSP)*, Belize City, Belize, 3-6 mars 1996, pp. 38–39.

[50] J.M. Moureaux, P. Loyer, and M. Antonini, "Efficient indexing method for lattice quantization applications," in *IEEE International Conference on Image Processing (ICIP)*, 16-19 septembre 1996, pp. 447–450.

[51] P. Raffy, M. Antonini, and M. Barlaud, "Multiresolution edge adaptive algorithm for low bit rate image coding," in *IEEE International Conference on Image Processing (ICIP)*, Lausane, Suisse, 16-19 septembre 1996, pp. 657–660.

[52] P. Raffy, M. Antonini, and M. Barlaud, "Zerotree edge adaptive coder for low bit rate image transmission," in *SPIE, Visual Communication and Image Processing*, San Jose, Etats-Unis, février 1997, pp. 1067–1076.

[53] S. Tramini, M. Antonini, and M. Barlaud, "Approche variationnelle non-linéaire en compression d'Images," in *3ièmes Journées d'Etudes et d'Echanges "COmpression et REprésentation Des Signaux Audiovisuels (CORESA)*, CNET Issy les Moulineaux, France, 26-27 mars 1997, pp. 235–242.

[54] S. Tramini, M. Antonini, and M. Barlaud, "Filtrage non-linéaire dynamique pour la compression des images," in *16ième Colloque GRETSI sur Le Traitement Du Signal et Des Images*, Grenoble, 15-19 septembre 1997.

[55] J.M. Moureaux, P. Nus, J.M. Van Binneveld, and M. Antonini, "Implantation sur DSP d'une méthode rapide d'Indexage pour la quantification vectorielle algébrique," in *16ième Colloque GRETSI sur Le Traitement Du Signal et Des Images*, Grenoble, 15-19 septembre 1997.

[56] S. Tramini, M. Antonini, and M. Barlaud, "Non-linear dynamic filtering for image compression," in *IEEE International Conference on Image Processing (ICIP)*, Santa-Barbara Californie, Etats-Unis, octobre 1997.

[57] S. Tramini, M. Antonini, and M. Barlaud, "Quantization noise removal for optimal transform decoding," in *IEEE International Conference on Image Processing (ICIP)*, Chicago, Etats-Unis, octobre 1998.

[58] J. Jung, M. Antonini, and M. Barlaud, "Optimal JPEG decoding," in *IEEE International Conference on Image Processing (ICIP)*, Chicago, Etats-Unis, octobre 1998.

[59] J. Vaisey, M. Barlaud, and M. Antonini, "Multispectral image coding using lattice VQ and the wavelet transform," in *IEEE International Conference on Image Processing (ICIP)*, Chicago, Etats-Unis, octobre 1998.

[60] A. Gouze, M. Antonini, and M. Barlaud, "Quincunx filtering lifting scheme for image coding," in *SPIE, Visual Communication and Image Processing*, San Jose, Etats-Unis, janvier 1999.

[61] S. Tramini, M. Antonini, and M. Barlaud, "Optimal joint decoding / deblurring method for optical images," in *IEEE International Conference on Image Processing (ICIP)*, Kobe, Japon, octobre 1999.

[62] S. Tramini, M. Antonini, and M. Barlaud, "Prise en compte de la chaîne complète d'Acquisition/Compression pour le décodage optimal d'Images," in *Colloque GRETSI*, Vannes, France, 13-17 septembre 1999.

[63] J. Jung, M. Antonini, and M. Barlaud, "Décodage MPEG orienté objet," in *Colloque GRETSI*, Vannes, France, 13-17 septembre 1999.

[64] S. Tramini, M. Antonini, and M. Barlaud, "Décodage et déconvolution conjoint optimal d'images," in *Colloque CORESA*, Eurecom, Sophia Antipolis, 14-15 juin 1999.

[65] J. Jung, M. Antonini, and M. Barlaud, "Supression des effets de bloc dans les séquences DV et MPEG-2 par décodage optimal orienté objet," in *Colloque CORESA*, Eurecom, Sophia Antipolis, France, 14-15 juin 1999.

[66] J. Jung, M. Antonini, and M. Barlaud, "Restauration de films anciens par une approche orientée objet," in *Colloque RFIA 2000*, Paris, France, janvier 2000.

[67] J. Jung, M. Antonini, and M. Barlaud, "Removing blocking effects and dropouts in DCT-based video sequences," in *Colloque Electronic Imaging 2000*, San-Jose CA , Etats-Unis, janvier 2000.

[68] P. Loyer, J.M. Moureaux, and M. Antonini, "Solving lattice codebook enumeration problem for generalized gaussian sources," in *IEEE Information Theory*, Sorrento, Italie, juin 2000.

[69] A. Gouze, M. Antonini, and M. Barlaud, "Quincunx lifting scheme for lossy image compression," in *IEEE International Conference on Image Processing (ICIP)*, Vancouver, Canada, 10-13 septembre 2000.

[70] C. Parisot, M. Antonini, and M. Barlaud, "EBWIC : A low complexity and efficient rate constrained wavelet image coder," in *IEEE International Conference on Image Processing (ICIP)*, Vancouver, Canada, 10-13 septembre 2000.

[71] S. Tramini, M. Antonini, M. Barlaud, and G. Aubert, "Spatio-frequency noise distribution a priori for satellite image joint Denoising/Deblurring," in *IEEE International Conference on Image Processing (ICIP)*, Vancouver, Canada, 10-13 septembre 2000.

[72] C. Lambert-Nebout, C. Latry, G. Moury, M. Antonini, M. Barlaud, and C. Parisot, "On-board optical image compression for future high resolution space remote sensing systems," in *SPIE 2000 : Special Session on Remote Sensing Compression*, San Diego, Etats-Unis, 30 juillet - 4 août 2000.

[73] J. Jung, M. Antonini, and M. Barlaud, "Une approche variationnelle orientée objet pour la gestion des pertes de cellules ATM lors de la transmission de séquences MJPEG," in *Colloque CORESA*, Futuroscope, Poitier, France, octobre 2000.

[74] J. Jung, M. Antonini, and M. Barlaud, "A new object-based variational approach for MPEG2 data recovery over lossy packet networks," in *Colloque SPIE Electronic Imaging 2001*, San-Jose CA , Etats-Unis, janvier 2001.

[75] C. Parisot, M. Antonini, M. Barlaud, S. Tramini, C. Latry, and C. Lambert-Nebout, "Optimization of the joint coding/decoding structure," in *IEEE International Conference on Image Processing (ICIP)*, Thessalonique, Grèce, 7-10 octobre 2001.

[76] A. Gouze, M. Antonini, M. Barlaud, and B. Macq, "Optimized lifting scheme for two-dimensional quincunx sampling images," in *IEEE International Conference on Image Processing (ICIP)*, Thessalonique, Grèce, 7-10 octobre 2001.

[77] C. Parisot, S. Tramini, M. Antonini, M. Barlaud, C. Latry, and C. Lambert-Nebout, "Optimization d'une chaîne image de télédétection : De la compression embarquée aux post-traitements sol," in *GRETSI*, Toulouse, France, 10-13 septembre 2001.

[78] C. Parisot, M. Antonini, and M. Barlaud, "3d scan-based wavelet transform for video coding," in *Workshop IEEE Multimedia Signal Processing (MMSP)*, Cannes, France, 3-5 octobre 2001.

[79] C. Parisot, M. Antonini, and M. Barlaud, "Optimal nearly uniform scalar quantizer design for wavelet coding," in *SPIE, Visual Communication and Image Processing (VCIP)*, San José, Californie, Etats-Unis, 20-25 janvier 2002.

[80] M. Antonini, J.M. Moureaux, and V. Lecuire, "Optimal multi-tone bit allocation for fixed rate video transmission over ADSL," in *SPIE, Visual Communication and Image Processing*, San José, Californie, 20-25 janvier 2002.

[81] F. Payan and M. Antonini, "3d mesh wavelet coding using efficient model-based bit allocation," in *Workshop IEEE 3D Data Processing Visualization Transmission*, Padoue, Italie, 19-21 juin 2002.

[82] M. Pereira, M. Antonini, and M. Barlaud, "Channel adapted multiple description video coding," in *IEEE International Conference on Multimedia and Expo (ICME)*, Lausanne, Suisse, 26-29 août 2002.

[83] C. Parisot, M. Antonini, and M. Barlaud, "High performance coding using a model-based bit allocation with EBCOT," in *EUSIPCO 2002, XI European Signal Processing Conference*, Toulouse, France, 3-6 septembre 2002.

[84] M. Pereira, M. Antonini, and M. Barlaud, "Low complexity multiple description coding scheme using wavelet transform," in *EUSIPCO 2002, XI European Signal Processing Conference*, Toulouse, France, 3-6 septembre 2002.

[85] C. Parisot, M. Antonini, and M. Barlaud, "Stripe-based MSE control in image coding," in *IEEE International Conference on Image Processing (ICIP)*, Rochester, New York, Etats-Unis, 22-25 septembre 2002.

[86] M. Pereira, M. Antonini, and M. Barlaud, "Channel adapted multiple description coding scheme using wavelet transform," in *IEEE International Conference on Image Processing (ICIP)*, Rochester, New York, Etats-Unis, 22-25 septembre 2002.

[87] F. Payan and M. Antonini, "Multiresolution 3d mesh compression," in *IEEE International Conference on Image Processing (ICIP)*, Rochester, New York, Etats-Unis, 22-25 septembre 2002.

[88] M. Pereira, M. Antonini, and M. Barlaud, "Multiple description coding for noisy-varying channels," in *Data Compression Conference (DCC)*, Snowbird, Utah, Etats-Unis, 25-27 mars 2003.

[89] M. Pereira, M. Antonini, and M. Barlaud, "Multiple description video coding for UMTS," in *IEEE EURASIP Picture Coding Symposium (PCS) Session Spéciale Codage Robuste et Codage Conjoint Pour l'Image et la Vidéo*, Saint-Malo, France, 23-25 avril 2003.

[90] V. Valentin, M. Cagnazzo, M. Antonini, and M. Barlaud, "Scalable context-based motion vector coding for video compression," in *IEEE EURASIP Picture Coding Symposium (PCS) Session Spéciale Codage Video*, Saint-Malo, France, 23-25 avril 2003.

[91] F. Payan and M. Antonini, "Weighted bit allocation for multiresolution 3d mesh geometry compression," in *SPIE, Visual Communication and Image Processing (VCIP)*, Lugano, Suisse, 8-11 juillet 2003.

[92] M. Pereira, M. Antonini, and M. Barlaud, "Multiple description coding for video streaming over wireless networks," in *6th IEEE International Conference on High Speed Networks and Multimedia Communications (HSNMC)*, Estoril, Portugal, 23-25 juillet 2003.

[93] M. Pereira, A. Gouze, M. Antonini, and M. Barlaud, "Multiple description coding for quincunx images. application to satellite transmission," in *6th IEEE International Conference on High Speed Networks and Multimedia Communications (HSNMC)*, Estoril, Portugal, 23-25 juillet 2003.

[94] M. Pereira, M. Antonini, and M. Barlaud, "Multiple description coding for internet video streaming," in *IEEE International Conference on Image Processing (ICIP)*, Barcelone, Espagne, 14-17 septembre 2003.

[95] F. Payan and M. Antonini, "3d multiresolution context-based coding for geometry compression," in *IEEE International Conference on Image Processing (ICIP)*, Barcelone, Espagne, 14-17 septembre 2003.

[96] A. Gouze, C. Parisot, M. Antonini, and M. Barlaud, "Optimal weighted model-based bit allocation for quincunx sampled image," in *IEEE International Conference on Image Processing (ICIP)*, Barcelone, Espagne, 14-17 septembre 2003.

[97] A. Gouze, M. Antonini, M. Barlaud, J. Meessen, Y. Verschueren, and B. Macq, "Efficient navigation through quincunx mega images," in *Third International Workshop on Content-Based Multimedia Indexing (CBMI)*, IRISA, Rennes, France, 22-24 septembre 2003.

[98] M. Cagnazzo, V. Valentin, M. Antonini, and M. Barlaud, "Motion vector estimation and encoding for motion compensated DWT," in *International Workshop VLBV03 Theme : Visual Content Processing and Representation*, Madrid, Espagne, 18-19 septembre 2003.

Communications invités dans des conférences avec comité de lecture

[99] J. Jung, M. Antonini, and M. Barlaud, "Restauration de séquences DV," in *La Villette*, 16 février 1999.

[100] C. Parisot, M. Antonini, M. Barlaud, C. Lambert-Nebout, C. Latry, and G. Moury, "On-board stripe-based wavelet image coding for future space missions," in *IEEE IGARSS 2000 : Special Session on Remote Sensing Compression*, Honolulu, Hawaii, 24-28 juillet 2000.

[101] C. Parisot, M. Antonini, and M. Barlaud, "Scan-based quality control for JPEG2000 using r-d models," in *EUSIPCO 2002, XI European Signal Processing Conference*, 3-6 septembre 2002.

[102] C. Parisot, M. Antonini, and M. Barlaud, "Motion-compensated scan-based wavelet transform for video coding," in *Proceedings of Tyrrhenian International Workshop on Digital Communications (IWDC'02)*, Capri, Italie, septembre 2002.

Conférences sans comité de lecture

[103] M. Barlaud, M. Antonini, and P. Mathieu, "Ondelettes et traitement numérique des images," in *GDR134 Traitement Du Signal et Image*, ENST Paris, 23 novembre 1989.

[104] M. Barlaud, M. Antonini, and P. Mathieu, "Compression d'Images par transformée en ondelette," in *Journée d'Etude sur Les Ondelettes À L'observatoire de Nice, Séminaire No.788, "Introduction À l'Analyse En Ondelettes"*, 20 juin 1989.

[105] M. Antonini, "Ondelettes et compression d'Images," in *GDR134 Traitement Du Signal et Image*, ENST Paris, 26 octobre 1990.

[106] M. Barlaud, M. Antonini, J.M. Bruneau, T. Gaidon, and P. Mathieu, "Application des ondelettes biorthogonales en traitement numérique des images," in *GDR134 Traitement Du Signal et Image*, ENST Paris, 26 juin 1990.

[107] M. Antonini, T. Gaidon, and M. Barlaud, "Quantification vectorielle et réseaux réguliers de points," ENST Paris, 10 décembre 1992, GDR134 Traitement du Signal et Image - Action compression.

[108] F. Davoine, M. Antonini, J.M. Chassery, and M. Barlaud, "Les fractals et la quantification vectorielle pour la compression d'Images," ENST Paris, 21 octobre 1994, GDR134 Traitement du Signal et Image - Action compression.

[109] J.M. Moureaux, M. Antonini, and M. Barlaud, "Quantification vectorielle sur réseaux : Application en compression d'Images," ENST Paris, 3 février 1994, GDR134 Traitement du Signal et Image - Action compression.

[110] M. Antonini and M. Barlaud, "Techniques futures de compression pour les images satellites," Toulouse, 13-14 novembre 1996, Atelier CNES Compression de données : état de l'art et perspectives.

[111] M. Antonini, M. Barlaud, D. Mainard, and C. Lambert-Nebout, "Compression multirésolution pour l'après SPOT5," Toulouse, 13-14 novembre 1996, Atelier CNES Compression de données : état de l'art et perspectives.

[112] M. Moureaux, P. Gauthier, M. Barlaud, P. Raffy, and M. Antonini, "Etude de la quantification vectorielle des données brutes RSO," Toulouse, 13-14 novembre 1996, Atelier CNES Compression de données : état de l'art et perspectives.

[113] P. Raffy, M. Barlaud, and M. Antonini, "Quantificateur scalaire à contraintes spatiales," ENST Paris, 11 janvier 1996, GDR-PRC ISIS - Action compression GT8.

[114] J.M. Moureaux, P. Nus, and M. Antonini, "Méthode d'Indexage pour la quantification vectorielle algébrique - implantation sur DSP," 17 mars 1997, GDR-PRC ISIS - GT6.

[115] S. Tramini, M. Antonini, and M. Barlaud, "Filtrage inverse non-linéaire dynamique pour la compression des images," ENST Paris, 31 janvier 1997, GDR-PRC ISIS - GT8.

[116] J. Jung, M. Antonini, and M. Barlaud, "Décodage optimal de JPEG," ENST Paris, 28 mai 1998, GDR-PRC ISIS - GT8.

[117] P. Raffy, M. Antonini, and M. Barlaud, "Modélisation, optimisation et mise en œuvre de quantificateurs bas débis pour la compression d'Images utilisant une transformée en ondelettes," ENST Paris, 9 janvier 1998, GDR-PRC ISIS - GT8.

[118] M. Antonini and J.M. Moureaux, "Quantification vectorielle sous contrainte de débit fixe," Toulouse, 22-23 juin 1999, Atelier CNES, Compression embarquée d'images fixes.

[119] M. Antonini, C. Parisot, M. Barlaud, C. Lambert-Nebout, D. Rozzonelli, and F. Pelleau, "Compression multi-résolution pour l'Observation de la terre haute résolution," Toulouse, 22-23 juin 1999, Atelier CNES, Compression embarquée d'images fixes.

[120] A. Gouze, M. Antonini, and M. Barlaud, "Compression d'images sans perte par lifting quinconce," Toulouse, 22-23 juin 1999, Atelier CNES, Compression embarquée d'images fixes.

[121] S. Tramini, M. Antonini, and M. Barlaud, "Décodage et déconvolution conjoints," ENST-Paris, 11 mars 1999, GDR-PRC ISIS - GT8.

[122] S. Tramini, M. Antonini, and M. Barlaud, "Décodage optimal d'Images satellite," Toulouse, 22-23 juin 1999, Atelier CNES, Compression embarquée d'images fixes.

[123] M. Antonini, O. Salom, J.M. Moureaux, and V. Lecuire, "Compression de séquences d'images pour la transmission de vidéos interactives sur ADSL," Hourtin, France, 11 janvier 2001, GDR-PRC ISIS - GT8.

[124] C. Parisot, M. Antonini, M. Barlaud, C. Lambert-Nebout, C. Latry, and G. Moury, "Compression d'images satellitaires au fil de l'eau au moyen d'EBWIC," ENST Paris, 11 janvier 2001, GDR-PRC ISIS - GT8.

Rapports internes de recherches

[125] M. Barlaud, P. Solé, T. Gaidon, M. Antonini, and P. Mathieu, "A pyramidal scheme for lattice vector quantization of wavelet transform coefficients applied to image coding," *Rapport interne I3S de 20 pages*, , no. 91-26, 1991.

[126] T. Gaidon, M. Antonini, M. Barlaud, and P. Mathieu, "Compression d'Images fixes par transformée en ondelettes et quantification vectorielle basée sur les treillis," *Rapport interne I3S de 64 pages*, , no. 92-59, octobre 1992.

[127] M. Antonini, "Compression d'Images satellites," *Rapport Post-Doctoral CNES - Rapport interne I3S de 46 pages*, , no. 93-73, septembre 1993.

[128] M. Antonini and M. Barlaud, "Weighted optimum bit allocation for multiresolution image coding," *Rapport interne I3S de 16 pages*, , no. 93-29, avril 1993.

[129] H. Munier, M. Antonini, and M. Barlaud, "Quantification vectorielle algébrique. application au codage d'images fixes," *Rapport interne I3S de 63 pages*, , no. 94-44, juillet 1994.

[130] J.M. Moureaux, M. Antonini, and M. Barlaud, "Fast encoding algorithm for lattice vector quantization," *Rapport interne I3S de 19 pages*, , no. 94-42, mars 1994.

[131] J.M. Moureaux, M. Antonini, and M. Barlaud, "Lattice vector quantization design for image coding," *Rapport interne I3S de 35 pages*, , no. 94-42, juillet 1994.

[132] M. Antonini, J.M. Moureaux, and M. Barlaud, "Etudes des performances d'un code produit appliqué à la quantification vectorielle algébrique," *Rapport interne I3S de 11 pages*, , no. 94-52, décembre 1994.

[133] J. Roca Oliver, M. Antonini, and M. Barlaud, "Allocation multirésolution optimale des débits binaires pour le codage d'Images," *Rapport interne I3S de 144 pages*, mai 1995.

[134] J.M. Moureaux, P. Loyer, and M. Antonini, "Low complexity indexing method for zn and dn lattice quantizers," *Rapport interne I3S de 31 pages*, novembre 1996.

[135] S. Tramini, M. Antonini, and M. Barlaud, "Intraframe image decoding based on a nonlinear variational approach," *Rapport Interne I3S*, , no. 97-26, novembre 1997.

[136] M. Antonini and D. Garravé Pont, "Compression d'Images satellites par quantification vectorielle," *Rapport Interne I3S*, , no. 98-07, juillet 1998.

[137] M. Antonini and C. Parisot, "Compression d'Images satellites haute résolution par quantification vectorielle algébrique et allocation de débits optimale," *Rapport Interne I3S*, , no. 98-19, septembre 1998.

Rapports de contrats

[138] M. Antonini and M. Barlaud, "Quantification vectorielle à l'aide de réseaux réguliers de points (lattices). aspects théoriques," *Rapport Contrat GDR/CNET*, , no. 936B005, février 1994.

[139] M. Antonini and M. Barlaud, "Quantification vectorielle algébrique pour la compression d'Images fixes," *Rapport de Fin de Contrat GDR/CNET*, janvier 1995.

[140] P. Charbonnier, M. Antonini, and M. Barlaud, "Implantation d'une transformée en ondelettes 2d dyadique au fil de l'eau," *Rapport Contrat CNES/TBS*, , no. 896/95/CNES/1379/00, octobre 1995.

[141] M. Antonini and M. Barlaud, "Compression bord multiresolution," *Rapport Final Contrat CNES/TBS*, , no. 896/95/CNES/1379/00, juillet 1996.

[142] M. Antonini, J. Jung, and M. Barlaud, "Etude de l'Allocation de débit dans un schéma de compression multispectrale TO/KLT," *Rapport Final Contrat AEROSPATIALE*, novembre 1997.

[143] M. Antonini, M. Barlaud, J. Jung, and R. Goglio, "Compression image bord multirésolution. lot1 - ondelettes quinconces," *Rapport Final Contrat CNES/CRIL*, , no. 896/CNES/96/0639/00, janvier 1998.

[144] M. Antonini, M. Barlaud, J. Jung, and R. Goglio, "Compression image bord multirésolution. lot3 - allocation dynamique," *Rapport Final Contrat CNES/CRIL*, , no. 896/CNES/96/0639/00, janvier 1998.

[145] M. Antonini and J.M. Moureaux, "Quantification vectorielle pour la compression d'Image bord. LOT1 - état de l'art," *Rapport Contrat CNES*, , no. 713/CNES/97/6966/00, juillet 1998.

[146] M. Antonini, S. Tramini, J. Jung, and M. Barlaud, "Décodage optimal d'Images spot," *Rapport Intermédiaire Contrat CNES*, , no. 762/98/CNES/7422/00, décembre 1998.

[147] M. Antonini, M. Barlaud, A. Gouze, and C. Parisot, "Compression bord multirésolution 99," *Rapport Contrat CNES*, , no. 1/96/CNES/96/0761, juin 1999.

[148] M. Antonini and J.M. Moureaux, "Quantification vectorielle pour la compression d'Image bord. LOT2 - quantification vectorielle par apprentissage," *Rapport Contrat CNES*, , no. 713/CNES/97/6966/00, février 1999.

[149] M. Antonini and J.M. Moureaux, "Quantification vectorielle pour la compression d'Image bord. LOT3 - quantification vectorielle algébrique," *Rapport Contrat CNES*, , no. 713/CNES/97/6966/00, juillet 1999.

[150] M. Antonini and J.M. Moureaux, "Quantification vectorielle pour la compression d'Image bord. LOT4 - synthèse et spécifications algorithmiques," *Rapport Contrat CNES*, , no. 713/CNES/97/6966/00, juillet 1999.

[151] M. Antonini, M. Barlaud, A. Gouze, and C. Parisot, "Compression bord multirésolution 99," *Rapport Final Contrat CNES*, , no. 1/96/CNES/96/0761, juin 2000.

[152] M. Antonini and M. Barlaud, "Une approche variationnelle orientée objet pour la détection et la suppression des pertes lors de la transmission de séquences d'images sur des canaux bruités," *Rapport Annuel RNRT COSOCATI*, mars 2001.

[153] M. Antonini, M. Barlaud, and C. Parisot, "Etude de l'Optimisation de l'Algorithme de compression bord multirésolution," *Bon de Commande Nř 13 Du Contrat CNES*, , no. 1/96/CNES/0761, avril 2001.

[154] M. Antonini, M. Barlaud, and C. Parisot, "Evaluation de l'asservissement et de la quantification optimale avec zone morte pour la compression multirésolution pléiades," *Rapport Final Contrat CNES*, , no. 713/01/CNES/8710/00, novembre 2001.

[155] M. Antonini, M. Barlaud, and C. Parisot, "Note d'expertise concernant l'asservissement pour la compression multirésolution pléiades," *Contrat CNES*, , no. 713/01/CNES/8710/00, décembre 2001.

[156] M. Antonini, M. Barlaud, L. Brunel, C. Boin, and T. Makram, "Transmission de vidéos à bas débits sur des canaux radio-mobiles dans le cadre de la future norme UMTS," *Rapport Annuel RNRT COSOCATI*, mars 2002.

[157] C. Parisot, M. Antonini, and M. Barlaud, "Solutions d'optimisation qualitative," *Rapport Annuel RNRT EIRE*, février 2003.

Thèse de doctorat

[158] M. Antonini, *Transformée en ondelettes et compression numérique des images*, thèse présentée à l'Université de Nice-Sophia Antipolis, 17 septembre 1991.

Deuxième partie

Mes travaux de recherche

Chapitre 1

Introduction générale

La compression des images et des vidéos n'est pas un domaine de recherche nouveau. Il existe déjà de nombreuses normes de compression d'images fixes telles que JPEG et JPEG-2000, et de vidéos telles que la famille MPEG-1,2 et bientôt H264 et MPEG-4. Cet état de fait présente un danger : penser que le problème de la compression est résolu et cesser toute activité dans ce domaine. Ce mouvement de pensée était déjà présent à la fin des années 80 et au début des années 90. Et pourtant depuis, nous avons vu apparaître les standards MPEG-1 et MPEG-2 avec les succès commerciaux que nous leurs connaissons aujourd'hui. En effet, de nombreuses techniques impliquant l'image et la vidéo ont vu le jour : le développement d'Internet, l'apparition des appareils-photos et des camescopes numériques ou encore de la télévision numérique, les images de la terre ou de l'espace que nous renvoient les satellites, les images médicales tridimensionnelles, ne sont que quelques exemples parmi de nombreux. Face à cette demande, les supports digitaux ont également évolué (vidéodisque, DVD...), et exigent à présent une qualité d'image exemplaire. Mais une image de qualité est une image qui est lourde à stocker, et sachant que par exemple la vidéo compte 25 images par seconde, que les maillages tridimensionnels peuvent être échantillonnés dans un intervalle compris entre 10 millions de points à plus d'un milliard de points, on réalise rapidement le problème : d'une part le stockage va être coûteux et d'autre part le temps de transmission sera excessif. Une telle quantité de données constitue un réel défi à la fois technologique et scientifique et plusieurs verrous doivent encore être levés. Technologique d'une part car les volumes de mémoire et de calculs requis dépassent les capacités des machines actuelles. Scientifique d'autre part car l'enjeu consiste à modéliser, traiter et exploiter des données gigantesques dans leur globalité tout en ayant un accès local à ces données. Par ailleurs, la tendance est à la diversification des infrastructures plutôt qu'à une unification des systèmes : assistants personnels, ordinateurs portables, stations de travail, systèmes de navigation. On assiste également à un changement des usages de l'informatique avec le développement rapide du nomadisme et de l'informatique en réseau, qui requièrent dans

leur version ultime l'ubiquité des ressources en calcul et des données afin d'en assurer l'exploitation sur des infrastructures hétérogènes.

L'apparition d'environnements de communication hétérogènes, inscrite dans un mouvement de convergence des secteurs des télécommunications, de l'audio-visuel et de l'information, pose en outre de nouveaux problèmes de représentation et de compression des signaux audio-visuels. Les applications multimédia communiquantes sont en effet confrontées à des problèmes de transmission de flux volumineux sur des réseaux aux performances variables. Il est alors toujours nécessaire de faire évoluer les méthodes de compression de l'information de façon à s'adapter à l'évolution des applications et des media de transmission ou de stockage. Un des enjeux est de développer des outils théoriques et algorithmiques afin de permettre à un concepteur d'accéder à une base de données et de l'exploiter aussi bien sur une puissante station de travail, depuis un ordinateur portable ou encore à partir d'un assistant personnel au cours d'une mission. Il est entendu que dans un tel scénario un assistant personnel ne fournira pas la même puissance de calcul ni les mêmes fonctionnalités qu'une station de travail, mais plutôt un accès et une vision globale suffisante pour la prise de décision et le travail collaboratif. La *scalabilité* des données comprimées est donc un point clé. Il n'y a pas encore de solutions scalables efficaces, permettant d'adapter au mieux la transmission des signaux aux diverses ressources disponibles (réseaux ou terminaux). Les solutions de codage scalable dites à grain fin existantes, essentielles pour une adaptation dynamique du contenu, sont loin d'être satisfaisantes et souffrent encore de faibles performances en compression. D'autre part, avec avec l'apparition des réseaux mobiles la conception des systèmes de compression ne peut plus se faire en supposant une qualité de service transport garantie. L'hypothèse de taux d'erreur résiduel quasi nul n'est plus vraie dans les réseaux sans fils et mobiles dont les caractéristiques des canaux varient dans le temps (canaux non stationnaires). Il reste donc nécessaire de développer des solutions de compression *scalable* et *robuste* des images et des vidéos. C'est dans cette optique que s'est orienté au cours des années mon travail de recherche en compression.

Le point de départ des techniques de compression est l'idée selon laquelle la simple numérisation des images ne permet pas un codage économique de l'information qu'elles contiennent. Autrement dit, la description numérique est fortement redondante et peu rentable, et il donc possible de représenter les images ou vidéos numériques sous une forme plus dense que la simple description numérique, tout en conservant sensiblement le même message visuel. Les méthodes de compression des images et des vidéos ont ainsi convergé naturellement vers le schéma de compression présenté à la figure 1.1 et de décompression présenté figure 1.2. En effet, la compression de l'information nécessite une transformation adaptée des données en les projetant sur une base de fonctions, et un codage des coefficients transformés. Les images naturelles peuvent être considérées comme une combinaison de régions homogènes et texturelles ainsi que

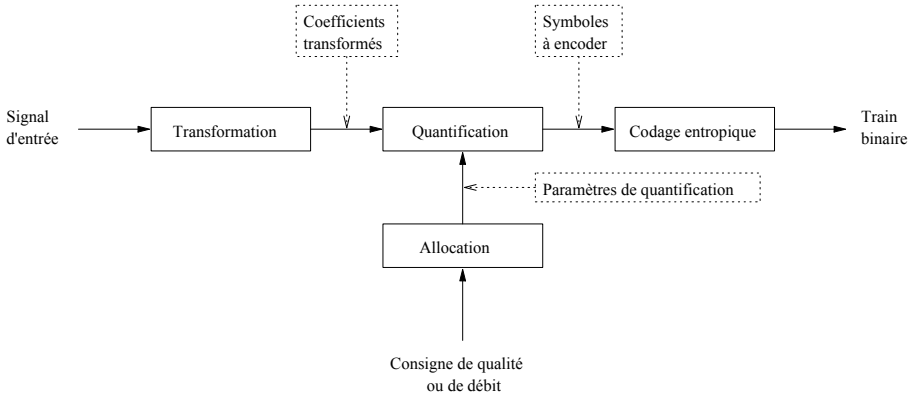


FIG. 1.1 – Schéma général de COMPRESSION des images et des vidéos.

d'une partie géométrique. Trois grands types d'approches sont alors apparus pour représenter au mieux ce type de signaux :

- Les approches purement fréquentielles telles que la DCT qui ont permi le développement des codeurs JPEG, MPEG-1 et MPEG-2. Cependant, ce type d'approche ne permet pas d'exploiter la redondance spatiale contenue par une image ;
- Les approches spatio-fréquentielles basées sur la transformée en ondelettes ou les paquets d'ondelettes qui ont suscité le développement de JPEG-2000 [282]. Les codeurs basés sur la transformée en ondelettes fournissent une représentation efficace des zones homogènes (en utilisant des arbres de zéros par exemple), et des textures (en utilisant une méthode de quantification appropriée telle que la quantification scalaire ou vectorielle) ;
- Les méthodes fractales [192] ou encore des approches emergentes telles que les ondelettes géométriques ("wedgelet" [226], "bandelet" [224], "curvelet" [168] ou encore "ridgelet" [181]) qui outre les informations relatives aux régions homogènes et aux textures, tentent de prendre en compte l'information géométrique de contours contenue dans l'image.

C'est dans le deuxième type d'approche que mon travail de recherche s'est axé durant ces dix dernières années. Notre effort de recherche porte d'une part sur le choix de la transformée à utiliser ainsi que sur la méthode de codage appropriée à cette transformée et au canal de transmission, et d'autre part sur la méthode de décodage optimal pour la restitution du signal comprimé. Dans cette partie du document je développe les travaux de recherches les plus importants que j'ai effectué depuis mon entrée au CNRS en faisant référence

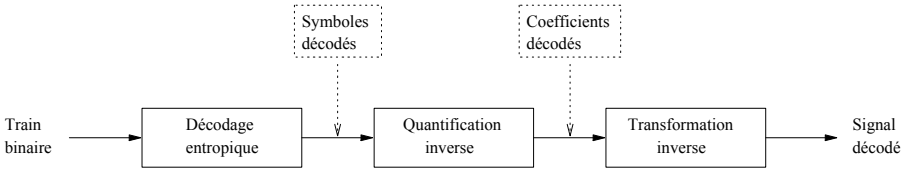


FIG. 1.2 – Schéma général de DECOMPRESSION des images et des vidéos.

dans certains cas à quelques unes de mes publications que j'ai annexé au document. Elle se décompose en quatre chapitres principaux, chacun traitant d'un domaine particulier de recherche. Dans le chapitre 2, je présente les travaux que j'ai effectués sur la transformée en ondelettes et son insertion dans un schéma de compression. Dans le chapitre 3, je présente la méthode d'allocation des ressources binaires que j'ai développé dans le cadre multirésolution et étendu au cas des descriptions multiples pour la transmission sur des canaux bruités ainsi qu'aux maillages tridimensionnels pour la compression d'objets volumétriques. Le chapitre 4 est consacré à la quantification vectorielle algébrique des coefficients d'ondelettes. Enfin, le chapitre 5 présente la méthode de décodage-restauration-débruitage que nous avons proposé pour les images fixes et la vidéo. La conclusion ainsi que mes perspectives de recherches futures sont données dans le chapitre 6.

Chapitre 2

La transformée en ondelettes

Dans ce chapitre je développe les activités de recherche que j'ai effectuées ou que j'effectue encore dans le domaine lié à la transformée en ondelettes. Le plan de ce chapitre est le suivant. Tout d'abord j'introduis dans le paragraphe 2.1 la notion d'ondelettes et leur nécessité en compression des images. Dans le paragraphe 2.2 je présente la construction des filtres "9-7" que nous avons développé avec I. Daubechies, une extension dans le cas 2D quinconce d'une implémentation par schéma lifting et enfin la transformation "au fil de l'eau". Le paragraphe 2.3 aborde le problème de la transformation en ondelettes pour les vidéos et principalement la transformée en ondelettes compensée en mouvement qui est un sujet de recherche sur lequel ma recherche porte actuellement. Enfin, dans le paragraphe 2.4 je présente les premiers travaux que nous avons effectués dans le domaine de la transformée en ondelettes sur des maillages 3D.

2.1 Pourquoi les ondelettes ?

Les ondelettes sont des fonctions qui ont été récemment introduites et développées en mathématiques. Le mot ondelette lui-même ne date que d'une vingtaine d'années, et la première ondelette a été introduite par Grossmann et Morlet en 1984 pour modéliser des signaux sismiques [187]. Les ondelettes sont des fonctions générées par translations et dilatations à partir d'une fonction appelée *ondelette mère* ψ. L'ensemble de ces fonctions d'ondelettes forme une famille de fonctions $\psi_{a,b}$ de $L^2(\mathbb{R})$ permettant d'analyser le signal dans un espace transformé. Une base d'ondelettes est donc définie par :

$$\psi_{a,b}\ (x) = |a|^{-\frac{1}{2}}\ \psi\left(\frac{x-b}{a}\right) \quad \text{avec } (a,b) \in \mathbb{R}^2 \text{ et } a \neq 0, \tag{2.1}$$

où les coefficients a et b désignent respectivement le facteur d'échelle pour la dilatation de ψ et le coefficient de translation. En outre, de façon à assurer

l'inversibilité de la transformée, la fonction ψ doit vérifier la condition d'admissibilité suivante :

$$\int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \tag{2.2}$$

où $\Psi(\omega)$ denote la transformée de Fourier de ψ. Dans l'espace $L^2(\mathbf{R})$, cette condition contraint l'ondelette mère à être une fonction de moyenne nulle :

$$\int_{-\infty}^{+\infty} \psi(x)\, dx = 0. \tag{2.3}$$

Cette propriété entaîne que l'ondelette est une fonction oscillante. L'oscillation de ψ et l'adjonction d'une propriété de décroissance liée à la régularité de l'ondelette ont donné le nom d'*ondelettes* ou encore "*petites ondes*" [176].

Historiquement, la plus ancienne base d'ondelettes connue est la base de Haar qui date du début du 20ième siècle (1910) [189]. Plus tard, dans les années 1980 plusieurs bases orthonormales de $L^2(\mathbb{R})$ ont été construites [177]. La première construction est due à Stromberg en 1982 ; les ondelettes qu'il proposa sont dans C^k (k arbitraire mais fini) et ont une décroissance exponentielle [232]. Ensuite, Meyer proposa en 1985 dans [216] des ondelettes telles que leur transformée de Fourier Ψ est à support compact et donc telles que ψ appartient à C^∞. Peu de temps après, Tchamitchian a construit en 1987 le premier exemple d'ondelettes non orthogonales [238]. Se basant sur ces travaux, Battle en 1987 [165] et Lemarié en 1988 [205] ont utilisé des méthodes différentes pour construire des familles d'ondelettes orthogonales identiques avec ψ dans C^k (k arbitraire mais fini). Enfin, les travaux de Daubechies [175] en 1988 et ceux de Cohen, Daubechies et Feauveau [171] en 1992 ont permis d'introduire la génération d'ondelettes adaptée au traitement des images.

En 1985, Y. Meyer a montré [216], [217] qu'il existait des fonctions ondelettes mère ψ telles que pour $a = 2^m$ et $b_0 = n2^m$ la famille de fonctions

$$\begin{aligned} &\psi_{m,n}(x) = 2^{-\frac{m}{2}} \psi(2^{-m}x - n) \quad \text{pour tout } (m,n) \in \mathbb{Z}^2 \\ &\text{avec } c_{m,n}(f) = \langle \psi_{m,n}, f \rangle = \int f(x)\, \overline{\psi}_{m,n}(x)\, dx \end{aligned} \tag{2.4}$$

constitue une base orthonormale de $L^2(\mathbb{R})$. Ces travaux mettent en évidence une transformée en ondelettes dyadiques, le terme *dyadique* se référant au facteur d'échelle qui dilate l'ondelette d'un facteur puissance de 2. L'existence des bases orthonormales est conditionnée par la propriété de régularité donnée dans [217], [175] et [177]. Les ondelettes qui constituent des bases de $L^2(\mathbb{R})$ correspondent à des *analyses multirésolutions*. L'analyse multirésolution consiste, pour un signal donné, à définir des approximations de ce signal à différentes échelles (résolutions) en le projetant sur une base d'ondelettes. Ce concept

introduit par Mallat en 1989 [421] est un outil mathématique bien adapté à l'utilisation de bases d'ondelettes en analyse d'images. Pour définir l'analyse multirésolution, on introduit une fonction d'échelle ϕ de la même manière que l'ondelette et directement liée à l'ondelette, avec [217], [211] :

$$\phi_{m,n}(x) = 2^{-\frac{m}{2}} \phi\left(2^{-m}x - n\right) \text{ pour tout } (m,n) \in \mathbb{Z}^2 \tag{2.5}$$

Depuis ces travaux, les scientifiques portent un intérêt de plus en plus important pour ces fonctions. En fait, la philosophie des ondelettes ne résulte que d'une synthèse de plusieurs idées déjà existantes dans différents domaines tels que le traitement du signal (codage en sous-bandes, filtres Filtres Miroirs en Quadrature - FMQ [173] et Filtres Conjugués en Quadrature - FCQ [231], [244]), la physique, les mathématiques. D'un autre point de vue, les ondelettes peuvent être considérées comme un outil mathématique dont l'utilisation et les applications sont multiples. Notamment, en traitement des images, de nombreuses applications très intéressantes sont développées actuellement (compression/codage, détection de contours, ...) et suscitent l'intérêt de nombreux chercheurs.

Pourquoi utiliser les ondelettes, et par là même l'analyse multirésolution, en traitement des images ? La raison en est simple, la transformée en ondelettes est une transformation qui admet la non-stationnarité, qui est bien localisée à la fois en espace et en fréquence et qui peut être obtenue par un algorithme rapide mis en œuvre à l'aide de filtres numériques (RIF ou RII) [421] (cf. figure 2.1) . De plus, l'analyse multirésolution d'une image permet d'obtenir un ensemble de sous-bandes ayant de meilleures caractéristiques que l'image d'origine : réorganisation de l'information et de l'énergie, plus faible entropie, contours orientés...

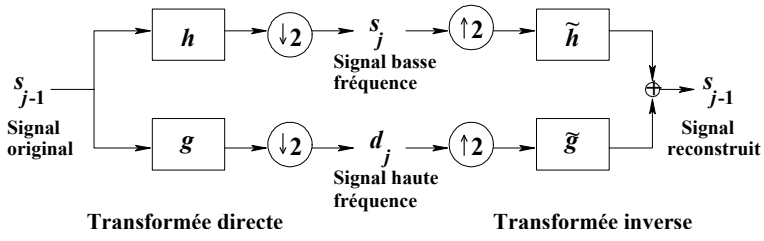


FIG. 2.1 – La transformée en ondelettes est mise en oeuvre par un banc de filtres numériques.

Cependant, les ondelettes ne sont pas directement et facilement utilisables sur le signal image. En effet, des problèmes subsistent notamment sur la définition de la forme des ondelettes, leur degré d'oscillation, leur régularité, leur nombre de moments nuls ainsi que sur le type d'ondelette à utiliser dans un

schéma de compression. Actuellement il n'existe pas de solution "miracle" pour régler ces paramètres et adapter les ondelettes aux images. Toutefois, en analyse d'images il semble préférable d'avoir des ondelettes présentant une certaine *régularité*, qui constituent des *bases orthogonales* de $L^2(\mathbb{R})$ et qui soient à *reconstruction exacte*. De plus, dans le but de diminuer la charge de calcul, les filtres associés aux ondelettes doivent être à *support compact* (réponse impulsionnelle courte). Enfin, il est bien connu en traitement d'images que les filtres doivent être à *phase nulle*, c'est-à-dire symétriques [164]. Malheureusement, toutes ces conditions ne peuvent être satisfaites simultanément. En effet, il n'existe pas de filtres RIF à phase linéaire et à support compact qui engendrent des ondelettes régulières orthonormales et qui permettent la reconstruction exacte [231].

Notre contribution dans ce domaine a consisté a introduire des ondelettes qui optimisent le compromis entre la régularité et les oscillations permettant de diminuer les artéfacts du type écho sur les contours des images traitées. Ces travaux ont permis la construction des bases dites "9-7" qui fournissent à l'heure actuelle les meilleurs résultats en compression d'image [249]. De plus, il sont retenus par toutes les propositions de pointe pour la future norme de compression d'images fixes JPEG-2000 et constituent un des filtres de référence pour cette nouvelle norme [282]. Nous avons développé des travaux sur les ondelettes biorthognales dans les cas séparable et non séparable (quinconce) et mes travaux de recherche portent actuellement sur l'utilisation des ondelettes et de l'analyse multirésolution pour la compression géométrique de maillages 3D.

2.2 Les ondelettes pour les images 2D

2.2.1 La biorthogonalité

Des bases *biorthogonales d'ondelettes régulières* ont été construites, de façon simultanée mais indépendante, par Cohen, Daubechies et Feauveau [183], [171] et par Herley et Vetterli [245], [246]. L'article [171] contient une étude mathématique détaillée de la construction des bases biorthogonales avec les preuves que, sous certaines conditions, les ondelettes biorthogonales constituent des bases numériquement stables (bases de Riesz). Une discussion sur les conditions nécessaires et suffisantes de régularité est aussi développée dans cet article. Les bases d'ondelettes biorthogonales sont une généralisation des bases d'ondelettes orthogonales. Dans le cas biorthogonal, il existe deux bases duales $\psi_{m,n}$ et $\widetilde{\psi}_{m,n}$, chacune étant construite par dilatations et translations d'une unique fonction mère ψ ou $\widetilde{\psi}$. Il en est de même pour les fonctions d'échelle $\phi_{m,n}$ et $\widetilde{\phi}_{m,n}$ générées à partir de dilatations et translations d'une unique fonction mère ϕ ou $\widetilde{\phi}$. P. Tchamitchian [238] a été le premier à construire en 1987

de telles paires de bases duales non orthogonales.

L'orthogonalité entre $\phi_{m,n}$ et $\widetilde{\phi}_{m,n}$ pour m fixé entraîne la relation suivante pour les filtres associés :

$$\sum_k h(k)\,\widetilde{h}(k+2n) = \delta_{n,0} \tag{2.6}$$

Cette condition de reconstruction exacte peut s'écrire en termes de polynôme trigonométriques $H(\omega)$ et $\widetilde{H}(\omega)$ (transformées de fourier des filtres h et $\widetilde{h}$). Pour des filtres symétriques l'équation (2.6) devient :

$$H(\omega)\,\widetilde{H}(\omega) + H(\omega+\pi)\,\widetilde{H}(\omega+\pi) = 1 \tag{2.7}$$

Le fait que les fonctions ψ et $\widetilde{\psi}$ sont respectivement $(k-1)$ et $(\widetilde{k}-1)$ fois continûment dérivable [171], [2] entraîne que les polynômes $H(\omega)$ et $\widetilde{H}(\omega)$ sont divisibles par $(1+\exp(-j\omega))^k$ et $(1+\exp(-j\omega))^{\widetilde{k}}$. Les filtres h et $\widetilde{h}$ doivent donc être de longueurs respectives supérieures à k et $\widetilde{k}$. Cette propriété et la relation (2.7) conduisent à la relation suivante (la démonstration est donnée dans [171] et [177]) :

$$H(\omega)\,\widetilde{H}(\omega) = \cos^{2l}\left(\frac{\omega}{2}\right)\left[\sum_{p=0}^{l-1}\binom{l-1+p}{p}\sin^{2p}\left(\frac{\omega}{2}\right) + \sin^{2l}\left(\frac{\omega}{2}\right)R(\omega)\right] \tag{2.8}$$

où, $R(\omega)$ est un polynôme impair en $\cos(\omega)$ (cf. proposition 6.1.2 page 171 du livre d'Ingrid Daubechies [177]) et $2l = k + \widetilde{k}$ (la symétrie des filtres h et $\widetilde{h}$ force la somme $k+\widetilde{k}$ à être paire).

2.2.2 Nos travaux avec Ingrid Daubechies et les filtres "9-7"

Suite à notre rencontre avec Ingrid Daubechies en 1989, nous nous sommes intéressés à l'utilisation de la transformée en ondelettes pour la compression des images et plus particulièrement à la transformée en ondelettes biorthogonales. En effet, en relachant la contrainte d'orthogonalité la notion de *biorthogonalité* permet de construire des ondelettes symétriques et donc à phase nulle, caractéristique très importante pour l'analyse des images. Notre intérêt commun à porté principalement sur la construction de bases d'ondelettes biorthogonales proches de bases orthogonales qui auraient de bonnes performances en compression des images [2]. Nous avons effectué des comparaisons sur différentes ondelettes au niveau de la régularité, des moments nuls, du degré d'oscillation des fonctions de base, de la longueur des filtres associés...). Nous avons choisi des ondelettes qui optimisent le compromis entre la régularité et les oscillations.

Les artéfacts du type écho n'apparaissent alors plus sur les contours des images traitées. Ces travaux, commencés au cours de ma thèse ont permis la construction des bases dites "9-7" [2], [31], [171], [177] qui fournissent à l'heure actuelle les meilleurs résultats en compression d'images [249]. Ces filtres ont fait l'objet d'une implantation sur circuit intégré commercialisé par Analog Device sous le nom de ADV601 ainsi que d'une implantation sur DSP par Texas Instrument. Ils sont de plus retenus par toutes les propositions de pointe pour la future norme de compression d'images fixes JPEG-2000. Ils constituent un des filtres de référence pour cette nouvelle norme [282].

Ces travaux ont été publiés dans la revue IEEE Transactions on Image Processing en 1992 : "Image coding using wavelet transform" [2]. Cet article est donné en annexe A du document.

2.2.2.1 Filtres splines de longueurs peu différentes

Pour construire une famille de filtres *spline* de longueurs peu différentes, on choisit $R \equiv 0$ dans la formule (2.8). Il est alors possible de trouver deux filtres h et $\widetilde{h}$ de longueurs proches en choisissant une factorisation appropriée du polynôme en $\sin(\omega/2)$ de degré $l-1$, en un produit de deux polynômes en $\sin(\omega/2)$ avec des coefficients réels [177]. Un polynôme sera affecté à $H(\omega)$ et l'autre à $\widetilde{H}(\omega)$. Ce choix fourni les plus petits filtres h et $\widetilde{h}$ dans cette famille ; il correspond à prendre $l=4$ et $k=4$. Dans ce cas, la relation (2.8) devient :

$$H(\omega)\,\widetilde{H}(\omega) = \cos^8\left(\frac{\omega}{2}\right)\left[\sum_{p=0}^{3}\binom{3+p}{p}\sin^{2p}\left(\frac{\omega}{2}\right)\right] \tag{2.9}$$

$$= 20\underbrace{\left(\cos^4(\frac{\omega}{2})\sin^2\left(\frac{\omega}{2}\right) + 0,342\,38\right)}_{(1)}\times \tag{2.10}$$

$$\underbrace{\cos^4(\frac{\omega}{2})\left(\sin^4\left(\frac{\omega}{2}\right) + 0,157\,62\sin^2\left(\frac{\omega}{2}\right) + 0,146\,03\right)}_{(2)}. \tag{2.11}$$

En introduisant la transformée en z, les termes (1) et (2) s'écrivent :

$$(1) = -\frac{1}{64z^3} - \frac{9,\,851\,3\times10^{-3}}{z^2} + \frac{0,101\,22}{z} + 0,190\,89 + 0,101\,22z - 9,\,851\,3\times10^{-3}z^2 - \frac{1}{64}z^3 \tag{2.12}$$

$$(2) = \frac{1}{256z^4} - \frac{2,\,462\,8\times10^{-3}}{z^3} - \frac{1,\,142\,4\times10^{-2}}{z^2} + \frac{0,038\,97}{z} + 0,088\,05 + 0,038\,97z - 1,\,142\,4\times10^{-2}z^2 - 2,\,462\,8\times10^{-3}z^3 + \frac{1}{256}z^4$$

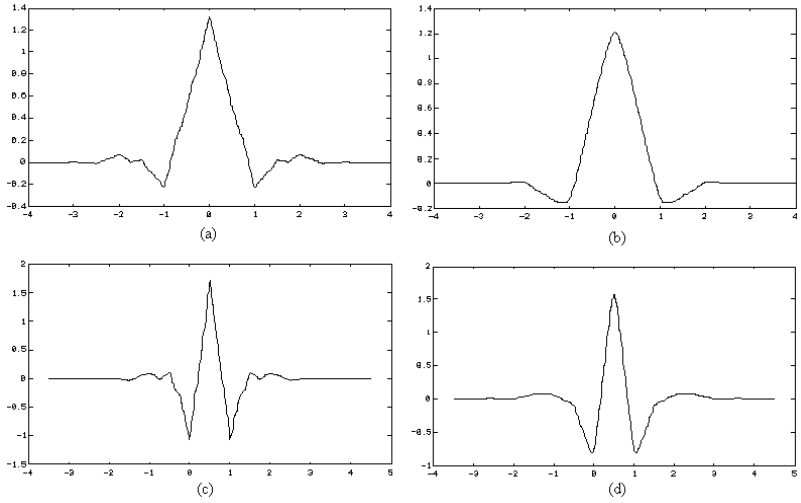


FIG. 2.2 – Fonctions d'échelles et ondelettes associées aux filtres "9-7". (a) fonction d'échelle ϕ ; (b) fonction d'échelle $\widetilde{\phi}$; (c) ondelette ψ ; (d) ondelette $\widetilde{\psi}$.

Le tableau 2.1 donne les coefficients des plus petits filtres obtenus dans cette famille (cf. équations **??**) et normalisés à 1. Du fait de leurs longueurs, ces filtres ont été appelés "9-7". Les fonctions d'échelles et ondelettes associées sont représentées sur la figure 2.2.

n	$h(n)$	$\widetilde{h}(n)$
0	0,602949018236	0,557543526229
± 1	0,266864118443	0,295635881557
± 2	-0,078223266529	-0,028771763114
± 3	-0,016864118443	-0,045635881557
± 4	0,026748757411	0

TAB. 2.1 – Filtres 9-7 : coefficients des filtres passe-bas h et $\tilde{h}$ ($l = 4$ et $k = 4$) normalisés à 1.

Notons qu'il existe plusieurs exemples où $R \neq 0$. En particulier, on peut choisir R tel que h et $\widetilde{h}$ soient très proches l'un de l'autre et tous les deux proches d'une base orthonormale d'ondelettes. Nous avons donné en 1992 [2] le premier exemple dans cette famille. Il correspond à prendre le filtre de Burt issu de la pyramide Laplacienne[1] [166] comme filtre h et à choisir $R(\omega) = 48\cos(\omega)/175$

[1] Un panorama des différentes techniques de représentations pyramidales des images est

avec $l = 2$ et $k = 2$ donnant 2 moments nuls à ψ et $\widetilde{\psi}$. Cet exemple est présenté dans l'article [2] en annexe A du document. Il est connu sous le nom de filtres "5-7".

2.2.2.2 Discussion sur l'optimalité des filtres "9-7"

Le choix d'un banc de filtres pour la compression des images par transformée en ondelettes est un point important qui influence la qualité de l'image décodée ainsi que la conception du système de compression. Bien que la régularité de l'ondelette est un critère qui a été suggéré pour l'évaluation des filtres [225], le nombre de moments nuls de l'ondelette analysante est aussi une caractéristique à prendre en compte[2]. Cependant, ces critères mathématiques ne sont pas très significatifs dans le choix de la base puisque le seul critère impartial est la qualité visuelle des images décodées. C'est dans ce sens que Villasenor a mené ses expériences en 1995 dans [249] et montré que le filtre "9-7" présentait les meilleures performances en terme de compromis débit-distorsion. Le choix de la structure "9-7", c'est-à-dire 9 coefficients pour le filtre h et 7 coefficients pour le filtre $\widetilde{h}$ n'est pas anodine. En effet, nos expériences ont montré qu'inverser les filtres conduisait à des résultats débit-distorsion moins bons [2]. Il semblerait que mettre un nombre de moments nuls important à l'analyse (pour ψ) et une bonne régularité à la synthèse (pour $\widetilde{\psi}$) corresponde à la meilleure configuration possible. Malheureusement, aucune étude théorique ne permet aujourd'hui de déterminer le nombre de moments nuls ni le degré de régularité optimaux pour un système donné [246]. Dans un article récent [239], Unser et Blu font une excellente étude sur les propriétés mathématiques des filtres "9-7".

Quelques résultats expérimentaux comparatifs illustrants les performances des filtres "9-7" sont donnés figure 2.3. La transformée en ondelettes 2D est obtenue par un produit tensoriel d'espaces d'approximation 1D, mis en œuvre par un filtrage séparable de filtres 1D. Ces résultats ont été obtenus avec le système de compression/décompression EBWIC présenté au chapitre 3 et dont les performances sont similaires à celles de JPEG-2000. Les filtres "9-7", "9-3"et "5-7" sont issus de [2], le filtre 5-3 est issu de [204] et le filtre "9-15" est issu de [158].

2.2.3 Le quinconce

Ces travaux ont été réalisés durant la thèse d'Annabelle Gouze (2002) [186] que j'ai co-encadrée en collaboration avec le Professeur Michel Barlaud à l'Uni-

donné dans [163].

[2] En traitement d'image, cette propriété est utilisée pour détecter les zones d'irrégularités comme les contours et autres zones de discontinuités [169]. Elle s'adapte au domaine de la compression, puisque les zones régulières seront mises à zéro par la transformée en ondelettes. Seules les zones présentant un certain degré de singularité seront représentées par des coefficients non nuls.

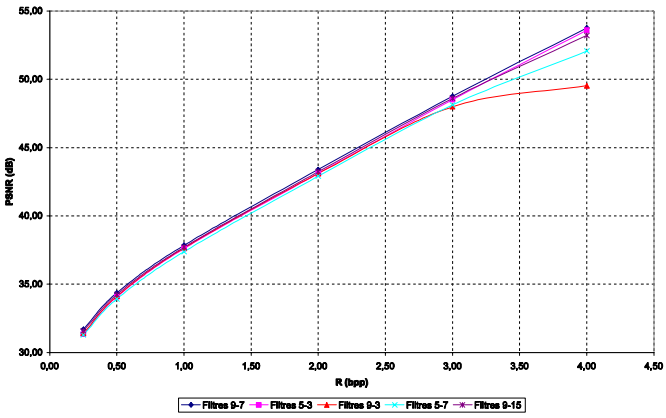


FIG. 2.3 – Rapport signal-à-bruit pic en fonction du débit pour l'image GOLD issue de la base de donnée JPEG-2000 (8 bpp−720 × 576 pixels). 3 niveaux de décomposition pour la transformée en ondelettes. Comparaison des performances pour différentes paires de filtres associés à des bases d'ondelettes.

versité de Nice-Sophia Antipolis. Une partie de nos travaux a été soumise dans la revue Transactions on Image Processing en mai 2003 [16].

2.2.3.1 L'analyse multirésolution avec un facteur $\sqrt{2}$

La transformée en ondelettes bidimensionnelles non-séparables se base à la fois sur le concept d'analyse multirésolution avec un facteur de résolution $\sqrt{2}$, introduit par Feauveau en 1990 [183], [184], et sur la définition des bancs de filtres bidimensionnels non-séparables, énoncée par Vetterli en 1984 [243] et Adelson et Simoncelli en 1990 [159]. L'analyse multirésolution génère des espaces emboîtés, avec un facteur de résolution $\sqrt{2}$ entre deux espaces consécutifs. Elle est deux fois plus fine que dans le cas séparable et n'est plus définie par un produit tensoriel d'espaces d'approximation 1D comme dans le cas dyadique séparable, mais elle est engendrée par les bases de fonctions d'échelle suivantes :

$$\Phi_{m,n}(x,y) = 2^{-m}\Phi\left(L^{-2m}(x,y) - n\right) \quad \text{avec } n = (n_x, n_y) \in \mathbb{Z}^2 \quad \text{et } j \in \frac{1}{2}\mathbb{Z} \tag{2.13}$$

où L est une transformation linéaire vérifiant $L(x,y) = (x+y, x-y)$ et $L \circ L = 2Id$. En outre, Meyer a montré qu'elle n'utilise qu'une seule ondelette au lieu de trois, et permet de décomposer une image en deux canaux au lieu de quatre

pour le cas séparable. L'ondelette bidimensionnelle est définie par la relation :

$$\Psi_{m,n}(x,y) = 2^{-m}\Psi\left(L^{-2m}(x,y) - n\right) \quad \text{avec } n = (n_x, n_y) \in \mathbb{Z}^2 \quad \text{et } j \in \frac{1}{2}\mathbb{Z} \tag{2.14}$$

Les filtres associés sont non-séparables, de forme diamant, et sont moins anisotropes que les filtres séparables. Le support fréquentiel idéal est aussi de forme diamant [247]. En conséquence, les filtres quinconces décorrèlent l'information, principalement sur les axes diagonaux. Au niveau pratique, le traitement des images par une transformée quinconce offre plusieurs intérêts. Notamment, la fonction de transfert de modulation de certains instruments optiques a un support fréquentiel proche du support quinconce [418], [202]. L'utilisation de la transformée non-séparable présente deux problèmes majeurs :

- Premièrement, la construction de filtres avec de bonnes propriétés de reconstruction parfaite, de biorthogonalité et de régularité n'est pas évidente. Plusieurs études ont été faites concernant la définition de bancs de filtres bidimensionnels non-séparables. La construction de Siohan se base sur la définition de filtres demi-bandes en dimension deux [230]. Celle de Moreau de Saint Martin [218] utilise les bases de Gröbner [172]. Les constructions d'Ansari et Guillemot [161], de Kim et Ansari [197], nos travaux sur l'extension des filtres "9-7" transverses au cas quinconce transverse [3], et ceux de Kovačević et Vetterli [198], [416] utilisent des transformées d'extension 1D-2D, telles que la transformée rotationnelle ou la transformée de McClellan.
- Deuxièmement, les transformées non-séparables ont pour inconvénient d'être coûteuses en opérations. Cependant une solution permettant de réduire cette complexité est donnée par une implémentation de la transformée en *schéma lifting* [235] (cf. figure 2.4). Ce schéma correspond à factoriser la matrice polyphase du banc de filtres. La définition d'un filtrage non-séparable moins coûteux passe donc par l'élaboration d'une implémentation en schéma lifting quinconce. Bien que de nombreux travaux est été menés dans le cas 1D, il n'existe que peu d'articles traitant l'extension 2D non-séparable du lifting. Parmi ces travaux, nous pouvons citer ceux de Kovačević et Sweldens qui ont défini des opérateurs lifting quinconce par interpolation à l'aide de l'algorithme de Neville [199]. Une version lifting quinconce du filtre (2,2) a été développée par Uytterhoeven et Bultheel dans [241] en utilisant une subdivision de l'échantillonnage et un moyennage des coefficients. Ansari, Kim et Devovic [162] définissent un triplet de filtres demi-bandes de dimension un, et les disposent dans une structure pouvant être assimilée à un schéma lifting à trois pas. Ils définissent un filtrage bidimensionnel par l'application d'une transformée rotationnelle à chaque filtre demi-bande. Bien que la réalisation d'Ansari soit assez proche de la nôtre, aucune de ces méthodes ne propose l'implé-

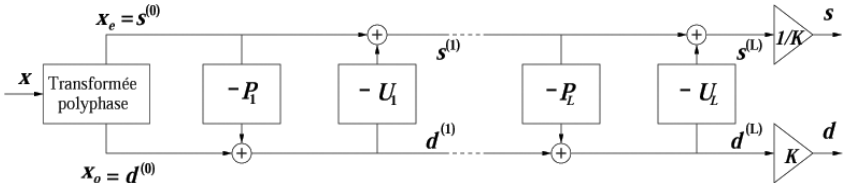


FIG. 2.4 – Schéma lifting 1D à L étages. s et d représentent respectivement le signal basses fréquences et le signal hautes fréquences. K correspond au gain des filtres.

mentation de bancs de filtres bidimensionnels non-séparables en schéma lifting à partir de la transformée de McClellan.

2.2.3.2 Schéma lifting quinconce : solution proposée

Problématique Daubechies [175] a développé une méthode de factorisation gérant le problème de création du schéma lifting pour des filtres de dimension un à partir des filtres transverses. Intuitivement, l'idée serait de généraliser la méthode de Daubechies pour des filtres à variables multiples et ensuite d'appliquer la factorisation dans le cas bidimensionnel à un banc de filtres quinconce, afin d'en obtenir une implémentation lifting. Si la formulation du problème est triviale, la réalisation est loin d'être aussi aisée. En effet, suite à l'application de la transformée en z, les filtres d'ondelettes 2D $h(z_1, z_2)$ et $g(z_1, z_2)$ s'expriment comme des polynômes de Laurent à deux variables. La factorisation de la matrice polyphase met en œuvre la division des deux composantes polyphases de h soit, $h_e(z_1, z_2)$ pour les coefficients pairs et $h_o(z_1, z_2)$ pour les coefficients impairs. Une absence de théorèmes fondamentaux traitant des polynômes à variables multiples proscrit la division. Malgré la conjecture de Serres et le théorème de Quillen-Suslin [233], [234] qui énoncent l'existence de la factorisation d'une matrice de déterminant un en matrices élémentaires, il n'y a pas de méthode automatique qui permet de définir la factorisation en matrices triangulaires [222]. En conséquence, la méthode de Daubechies n'est pas généralisable. Pour contourner ce problème nous ne traitons pas directement les filtres 2D, mais les filtres 1D qui en sont l'origine [69]. Une restriction s'impose alors, les filtres 2D considérés doivent obligatoirement être construits à partir d'une transformation des filtres 1D sur des filtres 2D [186].

McClellan pour le lifting La méthode que nous avons proposé utilise la transformation de McClellan [215] appliquée sur les pas lifting monodimensionnels. Cette transformation a pour principe de conserver les propriétés de biorthogonalité et de reconstruction parfaite. De plus elle préserve la compa-

cité du support et les zéros à la fréquence d'aliasing. Un point faible cependant concerne le degré de régularité qui n'est pas conservé par McClellan. Nous avons montré que tout couple de filtres 2D $(h[n_x, n_y], \widetilde{h}[n_x, n_y])_{(n_x,n_y)\in\mathbb{Z}^2}$ biorthogonaux, dont la transformée en z donne un polynôme de $\mathbb{R}\left[z_x + z_x^{-1} + z_y + z_y^{-1}\right]$ (domaine de définition de McClellan inverse), peut être "factorisé" en schéma lifting au moyen de la méthode que nous avons proposé. De plus, sous certaines conditions, notre méthode peut être généralisée pour des filtres multidimensionnels à phase nulle définis dans $l^2(\mathbb{R}^n)$ avec $n \in \mathbb{N}\backslash\{0\}$. Ces travaux sont présentés en détail dans la thèse de Gouze [186].

La solution de "factorisation" que nous avons proposée consiste à appliquer la transformée de McClellan aux différents opérateurs lifting 1D. Cependant, le domaine de définition de McClellan se restreint aux filtres à phase nulle. Or, les opérateurs lifting 1D obtenus par factorisation d'une matrice polyphase, ne le sont pas. Aussi, la transformée de McClellan ne peut être appliquée directement aux opérateurs $p^{(l)}$ et $u^{(l)}$ ($1 \leq l \leq L$, $2L$ étant le nombre de pas lifting). Nous avons donc introduit la proposition 1, démontrée dans la thèse de Gouze [186], qui nous permet de contourner ce nouveau problème et de générer des filtres vérifiant la condition de phase nulle.

Proposition 1 *Soient h et $\widetilde{h}$ des filtres passe-bas à phase nulle et définis dans $\ell^2(\mathbb{R})$, alors il existe une factorisation de la matrice polyphase*

$$\widetilde{\mathbf{P}}(z) = \begin{pmatrix} \widetilde{h_e}(z) & -h_o\left(z^{-1}\right) \\ \widetilde{h_o}(z) & h_e\left(z^{-1}\right) \end{pmatrix}$$

en schéma lifting, les opérateurs lifting $p^{(l)}$ et $u^{(l)}$ ($1 \leq l \leq L$, $2L$ étant le nombre de pas lifting) vérifiant

$$\widetilde{\mathbf{P}}(z) = \prod_{l=0}^{L-1} \begin{pmatrix} 1 & u^{(L-l)}(z) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ p^{(L-l)}(z) & 1 \end{pmatrix} \times \begin{pmatrix} K & 0 \\ 0 & 1/K \end{pmatrix},$$

telle que les $2L$ polynômes de Laurent $z^{-1}p^{(l)}(z^2)$ et $zu^{(l)}(z^2)$ ($\forall l \in \{1, \ldots, L\}$) soient à phase nulle.

Les opérateurs $z^{-1}p^{(l)}(z^2)$ et $zu^{(l)}(z^2)$ $\forall l \in \{1, \ldots, L\}$, sont à phase nulle et appartiennent au domaine de définition de la transformée de McClellan.

2.2.3.3 Algorithme proposé

La résolution du problème, lié à la restriction du domaine de définition de la transformée de McClellan, rend possible l'extension du schéma lifting au cas bidimensionnel non-séparable. La procédure d'extension peut débuter de deux manières : soit en partant d'un banc de filtres 2D à phase nulle, à symétrie octogonale et dont la transformée en z retourne un polynôme en $(z_x +$

$z_x^{-1} + z_y + z_y^{-1}$), soit directement à partir d'un banc de filtres 1D biorthogonal $(h, g, \widetilde{h}, \widetilde{g})$ où h et $\widetilde{h}$ sont des filtres symétriques en zéro. Suite à l'initialisation des données, le schéma lifting quinconce est défini par les trois étapes suivantes. La décomposition en lifting n'est pas unique, néanmoins nous retenons comme point de départ les opérateurs lifting $p^{(l)}$ et $u^{(l)}$ (issus de la factorisation) solutions de la proposition 1.

1. *Sur-échantillonnages des opérateurs :* Les différents opérateurs lifting $p^{(l)}$ et $u^{(l)}$ sont sur-échantillonnés et décalés d'une valeur 1, pour les pas de prédiction, et 0, pour les pas de mise à jour. Le sur-échantillonnage des opérateurs lifting se traduit par le remplacement des polynômes $p^{(l)}(z)$ et $u^{(l)}(z)$ par les polynômes $z^{-1}p^{(l)}(z^2)$ et $zu^{(l)}(z^2)$. A l'aide de la proposition 1 , nous pouvons affirmer que ces filtres sont à phase nulle. Les opérateurs lifting, définis par factorisation, subissent un changement d'échelle et un sur-échantillonnage d'un facteur 2. Un décalage d'indices de valeur 1 est appliqué aux opérateurs de prédiction et un décalage de -1 aux opérateurs de mise à jour. Suite à ces opérations les opérateurs $-p^{(l)}(z)$ et $-u^{(l)}(z)$ deviennent $-z^{-1}p^{(l)}\left(z^2\right)$ et $-zu^{(l)}\left(z^2\right)$.
2. *Application de la transformée de McClellan aux opérateurs* $z^{-1}p^{(l)}\left(z^2\right)$ *et* $zu^{(l)}\left(z^2\right)$: L'application de la transformée de McClellan sur $z^{-1}p^{(l)}(z^2)$ et $zu^{(l)}(z^2)$ retourne des polynômes de Laurent bidimensionnels correspondant à des opérateurs lifting. Ces opérateurs constituent les opérateurs d'un schéma lifting quinconce, sur-échantillonnés et décalés de ± 1.
3. *Sous-échantillonnage d'un facteur* 2 *des opérateurs bidimensionnels :* Les opérateurs lifting 2D, obtenus lors de l'étape précédente, ne correspondent pas réellement aux opérateurs lifting quinconce désirés. Le sur - échantillonnage mis en œuvre au point 2 subsiste. Pour remédier à ce problème, nous procédons à un sous-échantillonnage sur les opérateurs 2D et nous obtenons les opérateurs lifting quinconces.

2.2.3.4 Exemples et résultats

Sweldens et Schröder [236] ont développé des opérateurs lifting directement à partir des méthodes d'interpolation de Deslauriers et Dubuc [180]. Les exemples des opérateurs lifting (2,2), (4,2) et (6,2) sont donnés dans le tableau 2.2. Leur construction ne se base pas sur la factorisation de bancs de filtres biorthogonaux. Cependant, les opérateurs de prédiction et de mise à jour vérifient que $p^{(l)}[2n+1]$ et $u^{(l)}[2n-1]$ sont des opérateurs symétriques en zéro. En conséquence, ils peuvent être étendus en schéma lifting quinconce de la même façon que tout autre opérateur. Nous avons proposé une extension de ces opérateurs au cas lifting quinconce dans [186]. Le tableau 2.3 donne les coefficients de ces pas lifting quinconces correspondants. Les performances comparatives des filtres "9-7" quinconces transverses et des opérateurs (4,2) et (6,2) lifting

d	c	b	a	a	b	c	d

filtres lifting	(2,2)	(4,2)	(6,2)
$p^{(1)}$	$a = \frac{1}{2}$	$a = \frac{9}{2^4}$	$a = \frac{75}{2^7}$
		$b = -\frac{1}{2^4}$	$b = -\frac{25}{2^8}$
			$c = \frac{3}{2^8}$
			$d = \frac{3}{2^{13}}$
$u^{(1)}$	$a = -\frac{1}{4}$	$a = -\frac{1}{4}$	$a = -\frac{1}{4}$

TAB. 2.2 – Opérateurs lifting 1D construits à partir des méthodes d'interpolation de Deslaurier et Dubuc.

f	e	d	d	e	f
e	c	b	b	c	e
d	b	a	a	b	d
d	b	a	a	b	d
e	c	b	b	c	e
f	e	d	d	e	f

filtres lifting	(2,2)	(4,2)	(6,2)
$p^{(1)}$	$a = \frac{1}{4}$	$a = \frac{39}{2^7}$	$a = \frac{675}{2^{11}}$
		$b = -\frac{3}{2^7}$	$b = -\frac{165}{2^{12}}$
		$c = -\frac{1}{2^7}$	$c = -\frac{85}{2^{13}}$
			$d = \frac{15}{2^{12}}$
			$e = \frac{15}{2^{13}}$
			$f = \frac{3}{2^{13}}$
$u^{(1)}$	$a = -\frac{1}{8}$	$a = -\frac{1}{8}$	$a = -\frac{1}{8}$

TAB. 2.3 – Opérateurs lifting 2D quinconces que nous avons construits à partir des factorisations 1D.

quinconces sont données sur la figure 2.5 pour une image satellite NICE (numérisée sur 10 bpp) fournie par le CNES Toulouse (cf. figure 2.6). Cette image est une image simulée échantillonnée en quinconce. Ces résultats ont été obtenus avec le système de compression/décompression EBWIC présenté au chapitre 3 et dont les performances sont similaires à celles de JPEG-2000.

L'implémentation en schéma lifting permet de réduire la consommation de mémoire lors de l'exécution. La réduction est rendue possible par une actualisation des coefficients sur le même emplacement mémoire. De plus, la transformée est à reconstruction parfaite et son inverse se déduit très facilement du schéma lifting direct : comme pour n'importe quelle représentation lifting, elle est définie par une simple inversion des opérations. La réalisation d'une version entière du schéma lifting est toujours possible dans le cadre d'une transformée en ondelettes quinconce. Enfin, le nombre d'opérations arithmétiques requises par le lifting est réduite en comparaison à celui d'une transformée en ondelettes quinconce. Le gain est mis en évidence dans la thèse de Gouze [186]. Le schéma lifting quinconce offre une implémentation plus avantageuse de la transformée en ondelettes bidimensionnelle non-séparable. Il demeure toutefois limité aux filtres à phase linéaire obtenus par la transformation de McClellan.

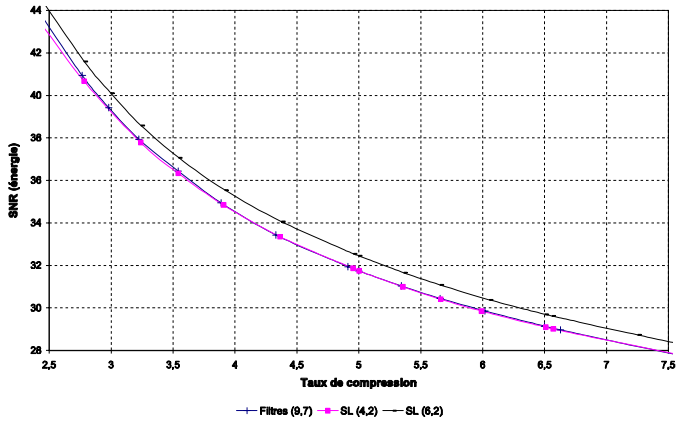


FIG. 2.5 – Rapport Signal-à-bruit en énergie en fonction du taux de compression pour l'image satellite NICE. Ces courbes comparent les filtres quinconces "9-7" transverses avec les opérateurs lifting (4,2) et (6,2) quinconces obtenus à partir de la transformée de McClellan.

2.2.3.5 Optimisation des filtres lifting

Nous avons mené ces travaux en collaboration avec le Professeur B. Macq de l'Université de Louvain-la-Neuve en Belgique [76], [16]. Notre objectif était de construire des filtres adaptés aux données sources afin de structurer au mieux le système de compression. En effet, la plupart des codeurs tiennent compte des propriétés statistiques des images. En conséquence, leur efficacité dépend du type de signaux à coder. L'étape transformée est capitale, puisqu'elle permet une restructuration des données au sein de la chaîne de compression, de façon à maximiser les performances du codeur et obtenir ainsi de plus hauts taux de compression pour une qualité donnée. Notre étude s'est articulée en deux parties. Dans un premier temps, la méthode visait à générer une sous-bande haute fréquence dont les caractéristiques statistiques s'adaptent au codage non-conservatif. Les pas lifting caractérisant les hautes fréquences sont définis par les opérateurs de prédiction. Nous avons proposé un critère afin d'optimiser le prédicteur et garantir ainsi un signal haute fréquence ayant de bonnes propriétés, en vue d'améliorer les performances du codeur exploité. Dans un second temps, la simulation de pertes dans l'image de coefficients d'ondelettes nous a conduit à la définition d'un nouveau critère optimisant le pas de mise à jour. Ce pas lifting a pour rôle d'affiner l'approximation basse résolution de l'image originale en y ajoutant l'information nécessaire. Pour un schéma de compression non-conservatif, la mise à jour peut être définie de façon à minimiser la distorsion entre l'image originale et l'image reconstruite, suite à la simulation

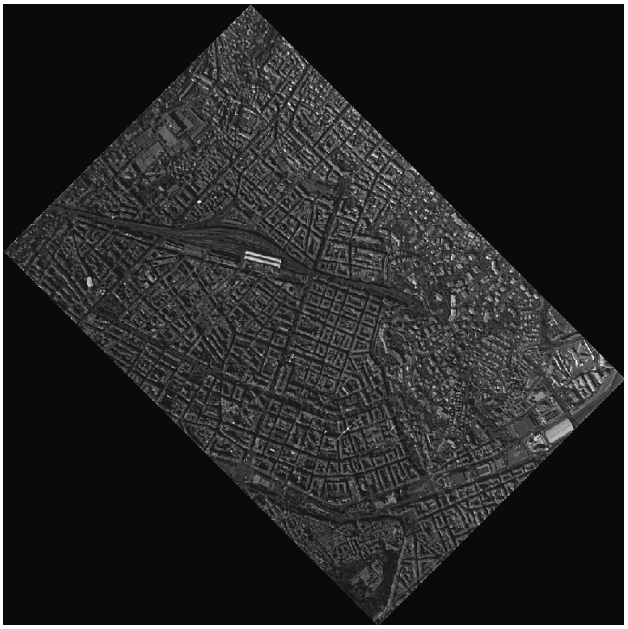


FIG. 2.6 – Image satellite quinconce de la ville de NICE numérisée sur 10 bpp. Cette image a été fournie par le CNES Toulouse.

de pertes dans les hautes fréquentes. La résolution de ce problème repose sur la méthode du filtrage de Wiener et s'inspire des travaux de [179] et [209]. Je ne présente pas ici la théorie de nos travaux qui peut être trouvée dans notre publication [16] et dans la thèse de Gouze [186].

2.2.4 La transformée "au fil de l'eau"

Ces travaux ont été initiés pendant le stage ingénieur de Pierre Charbonnier en 1995 et optimisés et étendus au cas de la vidéo durant la thèse de Christophe Parisot (2003) [299], stage et thèse que j'ai co-encadrés en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Une partie de nos travaux a été publiée dans la revue EURASIP Journal on Applied Signal Processing en janvier 2003 : "3D scan based wavelet transform and quality control for video coding" [12]. Cet article est donné en annexe B du document.

Le problème du "fil de l'eau" se pose lorsqu'il s'agit de comprimer de très grands volumes de données avec un minimum de ressources mémoire. Dans nos travaux, nous nous sommes intéressés à la compression dans le domaine

ondelettes de très grandes images au fur et à mesure de leur acquisition.

La première solution pour réduire les besoins mémoire consiste à découper le signal image en plusieurs blocs et à les comprimer indépendamment les uns des autres. Cette technique revient à considérer le signal à comprimer comme étant la concaténation de sous-signaux de taille mémoire admissible pour l'application. Les coefficients d'ondelettes des bords de chaque bloc sont calculés par symétrie en utilisant les valeurs des pixels qui sont dans le bloc. Le calcul des coefficients des bords de chaque bloc ne dépend donc pas des valeurs des pixels des blocs adjacents mais de pixels "imaginaires". En compression d'images par ondelettes, les effets visuels produits sur les bords de chaque bloc comprimé à l'aide d'une telle technique sont similaires à ceux obtenus avec JPEG [194]. Dès que le débit de compression est faible, on commence à voir apparaître des problèmes à la jointure entre les blocs qui ont été codés séparément. Notons que la possibilité de comprimer une image en plusieurs blocs indépendants (appelés tuiles) fait partie des fonctionnalités de JPEG-2000 [282].

Les méthodes à base de transformées calculées sur des blocs considérés comme indépendants ne sont pas satisfaisantes. En 1989, Malvar et Staelin proposent alors la transformée orthogonale à recouvrement (LOT) [212]. Cette transformée présente la particularité de décomposer un signal de longueur N sur une base de fonctions qui débordent sur les blocs adjacents. Les supports de ces fonctions sont supérieurs à N coefficients. La LOT offre les mêmes avantages que les méthodes de convolution par blocs avec gestion de recouvrements. En particulier, elle permet de supprimer les effets de bords qui étaient précédemment concentrés aux contours des blocs comprimés. En revanche, le support des fonctions de base grandit avec la taille du bloc. Donc, plus N est grand, plus on étale les erreurs de compression produites sur les discontinuités filtrées (phénomène de Gibbs, ringing).

La solution que nous avons développée a été suscitée par le Centre National d'Etudes Spatiales (CNES) de Toulouse, dans le cadre d'une application satellitaire. Nous avons été amenés à développer un algorithme de compression d'images adapté à un traitement bord de faible complexité et coût mémoire (cf. chapitre 3). Dans ce contexte, la solution que nous avons proposée est de calculer la véritable transformée en ondelettes de l'image au fur et à mesure de son acquisition en comprimant les coefficients d'ondelettes calculés dès que possible de façon à libérer la mémoire pour les calculs suivants. Ces travaux ont donnés lieu en 1995 à un algorithme de transformée en ondelettes $2D$ au "fil de l'eau" [140]. Ce rapport est resté non publié jusqu'en 2000 [72]. Historiquement, des travaux similaires liés à l'implémentation "au fil de l'eau" d'une transformée en ondelettes $1D$ avaient été publiés peu de temps auparavant par Vishwanath [250] en 1994. L'idée était de calculer la transformée en ondelettes au fur et à mesure de l'acquisition des échantillons du signal d'entrée à la façon d'une fenêtre glissante et de ne préserver les échantillons du signal d'entrée en mémoire que tant qu'ils sont nécessaires au calcul de la transformée. Plus

tard, d'autres approches ont été proposées en $2D$ par Ordentlich, Taubman, Weinberger, Seroussi et Marcellin en 1999 [220] et par Chrisafis et Ortega en 2000 [170].

La transformée en ondelettes d'un signal est obtenue par une convolution du signal d'entrée par un ou plusieurs filtres de longueur finie. Le fait que ces filtres aient un support fini permet de calculer les coefficients d'ondelettes à l'aide d'un nombre fini d'échantillons du signal. Les méthodes de calcul de la transformée en ondelettes au "fil de l'eau" utilisent cette propriété pour calculer les coefficients haute et basse fréquences de la transformée avec une quantité de mémoire minimale. Nous avons proposé des algorithmes de calcul de la transformée au "fil de l'eau" pour une transformée mise en œuvre par filtrage transverse ou par shéma lifting. Une étude complète et détaillée de ces travaux est présentée dans la thèse de Parisot [299] et dans l'article "3D scan based wavelet transform and quality control for video coding" [12] donné en annexe B du document.

2.3 Les ondelettes 2D+t pour les vidéos

2.3.1 La scalabilité

Nous sommes en train de travailler sur ce problème et je co-encadre actuellement la thèse de Marco Cagnazzo ainsi que celle de Thomas André en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis.

La compression des vidéos n'est pas un domaine de recherche nouveau. Il existe déjà de nombreuses normes (MPEG-1, MPEG-2, MPEG-4 et bientôt H264 et MPEG-4 partie 10). Cet état de fait présente un danger : penser que le problème de la compression de vidéos est résolu et cesser toute activité dans ce domaine. Ce mouvement de pensée était déjà présent à la fin des années 80 et au début des années 90. Et pourtant, nous avons vu apparaître depuis les standards MPEG-1 et MPEG-2 avec les succès commerciaux que nous leurs connaissons aujourd'hui.

Actuellement, le codage *scalable* de vidéos est un grand challenge pour les applications multimedias. La scalabilité permet de distribuer des vidéos sur des réseaux dont les capacités de débits, mémoires et calculs locaux sont hétérogènes. C'est le cas de la diffusion du cinéma ou de vidéos obtenues à partir de caméras hautes définitions et dont les services sont diffusés à des utilisateurs différents. On suppose dans ce cas que le codeur peut supporter une grande complexité alors qu'on va privilégier une faible complexité du coté des différents récepteurs. Cependant, il n'y a pas encore de solution scalable efficace, permettant d'adapter au mieux la transmission des signaux aux diverses ressources disponibles (réseaux ou terminaux). Les solutions de codage scalable

dites à grain fin existantes [190], essentielles pour une adaptation dynamique du contenu, sont loin d'être satisfaisantes. D'autre part, les méthodes les plus récentes tels que MPEG-4 [190] et H264-JVT [191] souffrent d'artefacts liés à l'utilisation de transformations par blocs ou encore ne permettent pas une allocation "optimale" des ressources binaires entre les vecteurs mouvement et le résiduel de compensation du mouvement. Tout ces problèmes deviennent sensibles dans les bas et très bas débits où l'information temporelle peut prendre une part très importante des ressources binaires disponibles.

Une solution a ces problèmes peut être trouvée grâce aux ondelettes, en combinant une transformation spatiale et une transformation temporelle [196]. Cependant, utiliser une transformée en ondelettes tridimensionnelle ne suffit pas pour prendre en compte le mouvement dans la séquence et rapidement les recherches se sont orientées vers des transformées compensées en mouvement 2D+t [219]. La transformée en ondelettes compensée en mouvement figure aujourd'hui parmi les méthodes les plus compétitives en terme de performances débit-distorsion et de scalabilité [214], [248], [102], [208], [229], [237]. C'est dans ce sens que nous avons commencé à entreprendre des travaux et que mon projet de recherches s'oriente (cf. chapitre 6).

2.3.2 La transformée temporelle au fil de l'eau

Ces travaux ont été réalisés durant la thèse de Christophe Parisot (2003) [299] que j'ai co-encadré en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Ils ont été publiés dans la revue EURASIP Journal on Applied Signal Processing en janvier 2003 : "3D scan based wavelet transform and quality control for video coding" [12]. Cet article est donné en annexe B du document.

Un des avantages de la compression basée sur une transformée en ondelettes 2D+t est de pouvoir fournir naturellement une séquence qui peut être décodée à plusieurs échelles différentes (aussi bien en résolution spatiale que temporelle). Malheureusement, le calcul de la transformée en ondelettes temporelle ne peut pas être réalisé sur la totalité de la séquence en une seule fois compte tenu des besoins mémoire que cela engendrerait. Une solution simple consiste à découper la séquence vidéo en des groupes d'images (par exemple 16 ou 32 images) et de traiter ces groupes indépendamment les uns des autres. Cependant, cette solution fait apparaître des effets de bords dans le sens du temps [252] de la même façon que la compression des images utilisant une transformée par blocs en produisait dans le cas 2D. Dans le cas des séquences vidéo, ce problème se traduit par un mouvement saccadé entre la dernière image d'un bloc d'images temporel et la première image du bloc temporel suivant. Les saccades sont dues au fait qu'une grande partie des coefficients d'ondelettes haute résolution temporelle est perdue lors de la compression. Ceci peut aussi rendre visible un effet

de pompage qui passait inaperçu sur la séquence d'origine.

Le principe de cette transformée temporelle se rapproche de celui introduit pour les images 2D. Une étude complète et détaillée de nos travaux est présentée dans la thèse de Parisot [299] et dans l'article "3D scan based wavelet transform and quality control for video coding" [12] donné en annexe B du document.

2.4 Les ondelettes pour les maillages 3D

Ces travaux ont été réalisés durant la thèse de Frédéric Payan commencée en octobre 2000 [300] et que j'encadre actuellement dans le cadre d'un contrat Région PACA / Entreprise avec la société Opteway. Les travaux présentés ici ont été soumis dans la revue IEEE Signal Processing Letters en septembre 2003 [17].

2.4.1 Maillages et surfaces

Les maillages constituent un outil puissant pour modéliser des objets 3D complexes grâce à leur double nature géométrique et combinatoire (les positions des sommets et la connectivité). Bien que de nombreuses alternatives existent pour la modélisation de formes surfaciques, les maillages sont aujourd'hui omniprésents et des efforts considérables sont développés pour le "traitement numérique de la géométrie" opérant essentiellement sur des maillages triangulaires. Il est important ici de distinguer les maillages des surfaces. Dans le premier cas, un maillage modélisé avec soin et décrivant la géométrie surfacique d'intérêt doit être considéré sous sa forme originale, il s'agit alors de décomposer le maillage sur une base d'ondelettes par simplification, on parle aussi de démaillage. Dans le second cas, seule la géométrie surfacique est considérée pour les applications visées. Si cette géométrie est initialement décrite par un maillage, ce dernier est perçu comme une instance de la géométrie. Il subsiste alors un degré de liberté dans la manière de mailler : on s'autorise ainsi à remailler la géométrie sans introduire de distorsion géométrique de manière à obtenir dans le nouveau maillage des propriétés de régularité et d'échantillonnage exploitables pour la représentation multirésolution, le traitement et la compression. On déduit ainsi deux scénarios de modélisation et de compression progressive par décomposition en ondelettes géométriques, l'un procède par démaillage, l'autre par remaillage. La décomposition en ondelettes géométriques est commune aux deux scénarios.

Bien que les méthodes multirésolutions ont fait leurs preuves pour la compression des images 2D [282], leur utilisation sur des maillages 3D est relativement récente. Ceci vient du fait que la conception de filtres sur des subdivisions irrégulières est relativement compliquée [332] et que la théorie des ondelettes sur

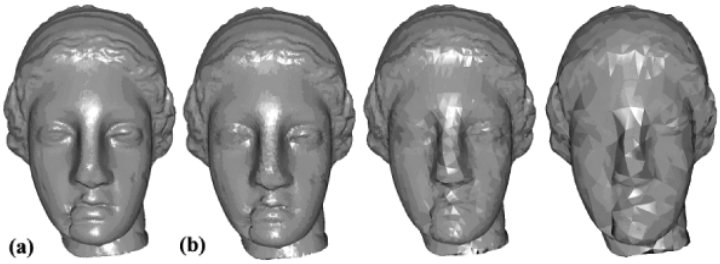


FIG. 2.7 – Image VENUS : exemple de maillages multirésolutions. (a) : maillage original, (b) : différentes résolutions d'approximation. L'information perdue entre deux niveaux de résolution est contenue dans les sous-images de coefficients d'ondelettes.

des maillages à subdivision régulière ou encore semi-régulière n'a été introduite que récemment [291]. En effet, Lounsbery a introduit en 1994 un schéma de compression progressive de surfaces en utilisant une transformée en ondelettes 3D sur des maillages surfaciques triangulaires obtenus par subdivision régulière d'un maillage de base irrégulier [291], [292]. Schröder, Sweldens et Kovacevic ont développé par la suite différentes techniques permettant de comprimer la géométrie des maillages et de réduire au minimum l'information de connectivité [228], [199]. Pour des applications où le remaillage n'est pas toléré, une généralisation de l'approche de Lounsbery à des maillages à subdivision irrégulière a été proposée dans [242]. Les approches multirésolutions obtenues sont scalables en résolution, en précision, en qualité ou encore en complexité (cf. figure 2.7). La représentation sur une base d'ondelettes géométriques d'un maillage surfacique à graphe de connectivité quelconque (irrégulier ou régulier), représentant la surface d'un objet 3D, s'appuie sur un problème d'inversion de subdivisions et apporte une méthode de compression réversible exacte du graphe de connectivité ainsi que de la géométrie des données. La transmission hiérarchique, progressive en qualité et débit binaire, est alors possible. Une décomposition géométrique en ondelettes efficace requiert de représenter au préalable le signal géométrique sous une forme compatible avec une structure régulière et un échantillonnage uniforme. La régularité parfaite étant impossible à obtenir pour des topologies arbitraires, le maillage est alors converti en un maillage semi-régulier (régulier par morceaux) [419],[188], [254] où la théorie des ondelettes sur M-canaux démontre sa supériorité.

Nous proposons dans ce paragraphe une étude générale sur l'erreur de reconstruction engendrée par la quantification des coefficients d'ondelettes dans le cadre d'une transformée en ondelettes avec un banc de filtres sur M-canaux.

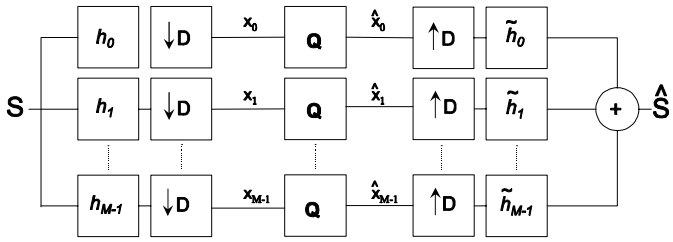


FIG. 2.8 – **Bancs de filtres**. Principe général d'un codeur ondelettes sur M-canaux. Le signal source est filtré en M sous-signaux par des filtres d'analyse h_i suivis d'un sous échantillonnage $\downarrow D$. Les sous-signaux quantifiés sont alors filtrés par les filtres de synthèse $\widetilde{h}_i$ suivi d'un sur-échantillonnage $\uparrow D$ puis additionnés pour reconstruire le signal.

Notre objectif est de pouvoir exprimer l'Erreur Quadratique Moyenne (EQM) de reconstruction pour n'importe quel sous-échantillonnage en fonction des coefficients de la matrice polyphase du banc de filtres. Dans cette optique, nous avons étendu les travaux d'Usevitch [325] à des filtres d-dimensionnels et des sous-échantillonnages autres que le sous-échantillonnage dyadique.

2.4.2 Le codage sur M-canaux

2.4.2.1 Principe

La figures 2.8 et 2.9 montrent le principe d'un codeur par transformée en ondelettes sur M-bandes. Un signal source s est transformé sur un ensemble de M sous-signaux $\{s_i, i = 0, ..., M-1\}$ au moyen d'une transformée sur M-bandes fréquentielles $\{h_i, i = 0, ..., M-1\}$ et d'un sous-échantillonnage. Les sous-signaux sont ensuite quantifiés et l'erreur de quantification ε_i entre le sous-signal s_i et sa valeur quantifiée $\widehat{s}_i$ est donnée par :

$$\epsilon_i = (s_i - \hat{s_i}) \tag{2.15}$$

Un sur-échantillonnage suivi d'un filtrage par les filtres de synthèse $\widetilde{h}_i$ donne le signal source quantifié $\widehat{s}$.

2.4.2.2 Notations

Soit un signal source s échantillonné sur un réseau $\mathcal{K}$ tel que :

$$s = \{s(\mathbf{k}) \in \mathbb{R} \mid \mathbf{k} \in \mathcal{K}\} \tag{2.16}$$

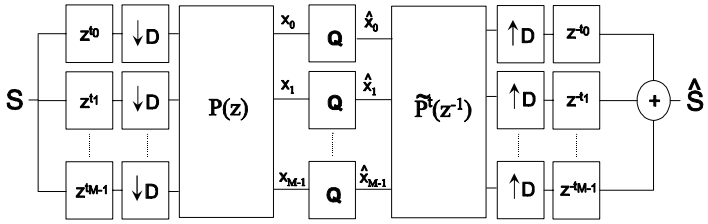


FIG. 2.9 – Représentation polyphase du banc de filtre donné à la figure 2.8.

où $\mathcal{K}$ est tel que $\mathcal{K} = \Gamma\mathbb{Z}^d$, avec Γ une matrice $d \times d$ inversible qui permet d'obtenir un échantillonnage sur des réseaux autres que le réseau $\mathbb{Z}^d$. Par exemple, dans le cas d'un échantillonnage sur le réseau triangulaire donné figure 2.10, cette matrice est telle que :

$$\Gamma = \begin{bmatrix} 1 & 1/2 \\ 0 & \sqrt{3}/2 \end{bmatrix}. \tag{2.17}$$

L'introduction de cette matrice permet de donner les coordonnées des points physiquements voisins d'un point filtré. Pour développer les calculs et par soucis de simplicité, nous supposerons que cette matrice est égale à l'identité, ce qui n'altère en rien le résultat obtenu [199].

Un sous réseau de $\mathcal{K} = \mathbb{Z}^d$ peut être obtenu par $D\mathbb{Z}^d$ où D est une matrice $d \times d$ de coefficients entiers. Le déterminant de D est un entier que l'on note $M \in \mathbb{Z}$. On peut alors écrire le réseau $\mathbb{Z}^d$ comme une somme de sous-réseaux

$$\mathbb{Z}^d = \bigcup_{j=0}^{M-1} \left(D\mathbb{Z}^d + \mathbf{t}_j \right), \tag{2.18}$$

avec $\mathbf{t}_j \in \mathbb{Z}^d$, $\mathbf{t}_0 = (0, 0, ..., 0)$ et $D^{-1}\mathbf{t}_j$ appartient à l'hypercube unité. Ainsi, un banc de M filtres $\{\widetilde{h}_i\}$ sur un réseau $\mathcal{K}$, peut s'exprimer en terme de la matrice polyphase par :

$$\widetilde{H}_i(\mathbf{z}) = \sum_{j=0}^{M-1} \mathbf{z}^{-\mathbf{t}_j}\, \widetilde{P}^t_{i,j}(\mathbf{z}^D) \quad \text{pour } i \in \{0, ..., M-1\} \tag{2.19}$$

où $\widetilde{P}^t_{i,j}$ est l'élément (i, j) de la matrice polyphase, $\mathbf{z}^{-\mathbf{t}_j}$ un décalage qui dépend du sous-réseau donné par :

$$\mathbf{z}^{-\mathbf{t}_j} = \prod_{n=1}^{d} z_n^{-\mathbf{t}_j(n)}, \tag{2.20}$$

et

$$\mathbf{z}^D = \{\mathbf{z}^{\mathbf{d}_1}, \mathbf{z}^{\mathbf{d}_2}, ..., \mathbf{z}^{\mathbf{d}_d}\}.$$

Le vecteur $\mathbf{d}_j$ est le $j^{ème}$ vecteur colonne de la matrice D tel que $\mathbf{z}^{\mathbf{d}_i}$ est donné par la relation suivante :

$$\mathbf{z}^{\mathbf{d}_j} = \prod_{n=1}^{d} z_n^{\mathbf{d}j(n)}. \tag{2.21}$$

2.4.3 Erreur Quadratique Moyenne du signal reconstruit

2.4.3.1 Cas d'un niveau de décomposition

Soit $r_\varepsilon(\tau)$ la fonction d'autocorrélation de l'erreur de reconstruction $\varepsilon = \{\varepsilon(\mathbf{k}) \in \mathbb{R} \mid \mathbf{k} \in \mathcal{K}\}$ introduite par la quantification du signal transformé. L'EQM entre le signal reconstruit et le signal d'origine sur N_s échantillons est donnée par la relation :

$$\sigma_\varepsilon^2 = \frac{1}{N_s}\left[r_\varepsilon(\mathbf{0})\right]. \tag{2.22}$$

avec $\mathbf{0}$ le vecteur nul en dimension d.

Le problème posé consiste à trouver une expression de l'EQM en fonction de l'erreur de quantification introduite dans chaque sous-bande fréquentielle et de la connaissance de l'ensemble des filtres de synthèse $\{\widetilde{h}_i\}$ du banc de filtres. Pour cela, nous allons développer l'expression de la fonction de corrélation $r_\varepsilon(\tau)$. La transformée en z de cette fonction est donnée par la relation suivante

$$R_\varepsilon(\mathbf{z}) = \varepsilon(\mathbf{z})\,\varepsilon(\mathbf{z}^{-1}), \tag{2.23}$$

où la transformée en z du bruit de quantification peut s'écrire en fonction du bruit introduit dans chaque sous-signal s_i :

$$\varepsilon(\mathbf{z}) = \sum_{i=0}^{M-1} \widetilde{H}_i(\mathbf{z})\,\varepsilon_i(\mathbf{z}). \tag{2.24}$$

Le signal $\varepsilon_i(\mathbf{z})$ correspond à la transformée en z de l'erreur de quantification $\varepsilon_i = \{\varepsilon_i(\mathbf{k}) \in \mathbb{R} \mid \mathbf{k} \in (D\mathcal{K} + \mathbf{t}_i)\}$ relative au $i^{ème}$ sous-signal. En supposant que les erreurs $\varepsilon_i(\mathbf{k})$ sont mutuellement décorrélées [193], [185], c'est-à-dire que $\varepsilon_i(\mathbf{z})\varepsilon_j(\mathbf{z}^{-1}) = \delta_{i,j} R_{\varepsilon_i}(\mathbf{z})$ (avec $\delta_{i,j}$ le symbole de Kronecker), alors les équations (2.23) et (2.24) permettent d'exprimer $R_\varepsilon(\mathbf{z})$ par :

$$R_\varepsilon(\mathbf{z}) = \sum_{i=0}^{M-1} R_{\widetilde{H}_i}(\mathbf{z})\,R_{\varepsilon_i}(\mathbf{z}). \tag{2.25}$$

La transformée en z inverse de l'équation (2.25) nous donne pour $\tau = \mathbf{0}$:

$$r_{\varepsilon}(\mathbf{0}) = \sum_{i=0}^{M-1} r_{\widetilde{h}_i}(\mathbf{0}) r_{\varepsilon_i}(\mathbf{0}). \tag{2.26}$$

Il s'agit donc maintenant de trouver l'expression de $r_{\widetilde{h}_i}(\mathbf{0})$ et $r_{\varepsilon_i}(\mathbf{0})$. La valeur $r_{\widetilde{h}_i}(\mathbf{0})$ correspond à l'énergie du filtre $\widetilde{h}_i$. En introduisant la notation polyphase, cette énergie est donnée par [17], [300] :

$$r_{\widetilde{h}_i}(\mathbf{0}) = \sum_{j=0}^{M-1} \sum_{\mathbf{k}} \left| \widetilde{p}_{i,j}^{t}(\mathbf{k}) \right|^2 \tag{2.27}$$

où les $\widetilde{p}_{i,j}^{t}(\mathbf{k})$ correspondent aux coefficients du polynôme $\widetilde{P}_{i,j}^{t}(\mathbf{z})$ (élément (i,j) de la matrice polyphase). D'autre part, en supposant le signal stationnaire et ergodique, l'énergie $r_{\varepsilon_i}(\mathbf{0})$ de l'erreur de quantification ε_i est donnée par la relation :

$$r_{\varepsilon_i}(\mathbf{0}) = \sum_{\mathbf{k} \in (D\mathcal{K} + \mathbf{t}_i)} \varepsilon_i(\mathbf{k})^2 = N_{s_i} \sigma_{\varepsilon_i}^2, \tag{2.28}$$

où $\sigma_{\varepsilon_i}^2$ est la variance du bruit de quantification estimée par l'EQM sur le sous-signal s_i et N_{s_i} le nombre d'échantillons dans le sous-signal s_i. En regroupant les équations (2.27) et (2.28) dans (2.26) et (2.22) il est facile d'obtenir la relation

$$\sigma_{\varepsilon}^2 = \sum_{i=0}^{M-1} \left[\frac{N_{s_i}}{N_s} \sigma_{\varepsilon_i}^2 \sum_{j=0}^{M-1} \sum_{\mathbf{k}} \left| \widetilde{p}_{i,j}^{t}(\mathbf{k}) \right|^2 \right] \tag{2.29}$$

qui donne l'EQM du signal reconstruit, dont on peut simplifier l'écriture par :

$$\sigma_{\varepsilon}^2 = \sum_{i=0}^{M-1} w_i \sigma_{\varepsilon_i}^2 \quad \text{avec} \quad w_i = \frac{N_{s_i}}{N_s} \sum_{j=0}^{M-1} \sum_{\mathbf{k}} \left| \widetilde{p}_{i,j}^{t}(\mathbf{k}) \right|^2. \tag{2.30}$$

Cette relation permet de calculer l'EQM de reconstruction introduite par la quantification d'un signal échantillonné sur une grille $\mathcal{K}$ donnée. Cette EQM dépend des pondérations w_i qui elles-mêmes dépendent des coefficients des filtres et du sous-échantillonnage introduit par le banc de filtres.

Nous pouvons remarquer que dans le cas orthogonal, c'est-à-dire lorsque $r_{\widetilde{h}_i}(\tau) = 1$ pour $\tau = \mathbf{0}$ et 0 sinon, alors l'équation (2.30) se résume à

$$\sigma_{\varepsilon}^2 = \sum_{i=0}^{M-1} \frac{N_{s_i}}{N_s} \sigma_{\varepsilon_i}^2. \tag{2.31}$$

où $\frac{N_{s_i}}{N_s} = \frac{1}{2}$ si le sous-échantillonnage est dyadique. De plus, pour $d = 1$ et $D = 2$, nous retrouvons les résultats donnés par Usewitch dans son article [325].

2.4.3.2 Cas d'une décomposition multi-niveaux

Il est intéressant d'estimer cette pondération lorsque plusieurs niveaux de décomposition sont effectués. Soit $\varepsilon_{i,m}$ l'erreur de quantification faite sur le canal i à l'indice de résolution m. Alors, de façon similaire au cas d'un niveau de décomposition, il est possible d'écrire à partir de l'équation (2.24) que l'EQM sur $\#m$ niveaux de décomposition est donnée par :

$$\sigma_{\varepsilon}^2 = \pi_{0,\#m}^{*}\sigma_{\varepsilon_{0,\#m}}^2 + \sum_{m=1}^{\#m}\sum_{i=1}^{M-1}\pi_{i,m}^{*}\sigma_{\varepsilon_{i,m}}^2 \tag{2.32}$$

avec

$$\begin{cases} \pi_{0,\#m}^{*} = \left(\frac{N_s}{N_{s_0}}\right)^{\#m-1}(w_0)^{\#m} \\ \pi_{i,m}^{*} = \left(\frac{N_s}{N_{s_0}}w_0\right)^{m-1} w_i. \end{cases} \tag{2.33}$$

La démonstration de ce cas de figure est donnée dans notre article [17]. Nous avons donc introduit une relation qui permet de déterminer l'EQM d'un signal transformé en ondelettes sur $\#m$ niveaux à partir de la puissance du bruit de quantification dans chaque sous-bande et de la matrice polyphase du banc de filtres. Nous avons développé cette relation dans le cas de sous-échantillonnages quelconques de façon à pouvoir estimer l'EQM dans le cadre de la compression de maillages 3D multirésolutions, c'est-à-dire lorsque la grille d'échantillonnage est une grille triangulaire.

2.4.4 Lifting et échantillonnage sur une grille triangulaire

2.4.4.1 Le lifting sur 4 canaux

Nous nous plaçons dans le cadre des maillages 3D obtenus par triangulation. Dans ce cas, le schéma lifting se base sur des échantillonnages triangulaires tels que le montre la figure 2.10 ($d = 2$ et $\mathcal{K} = \Gamma\mathbb{Z}^2$ avec Γ donné par (2.17)). Comme nous l'avons précisé précédemment, ce réseau peut être transposé dans un repère orthogonal et donc un sous réseau de $\mathcal{K}$ est obtenu pour $D = 2I$ ($M = 4$). Ainsi, la grille d'échantillonnage est constituée de 4 sous-signaux décalés par $\mathbf{t}_0 = (0,0)$, $\mathbf{t}_1 = (1,0)$, $\mathbf{t}_2 = (0,1)$ et $\mathbf{t}_3 = (1,1)$. Un schéma lifting adapté à cet échantillonnage est présenté sur la figure 2.11. Il comporte 4-canaux avec p_i et u_i les opérateurs de prédiction et de mise à jour associés au sous-signal i. La matrice polyphase associée à ce schéma lifting est une matrice 4×4 donnée par [199] :

$$\widetilde{P}^t = \begin{pmatrix} 1 & p_0 & p_1 & p_2 \\ -u_0 & 1-u_0p_0 & -u_0p_1 & -u_0p_2 \\ -u_1 & -u_1p_0 & 1-u_1p_1 & -u_1p_2 \\ -u_2 & -u_2p_0 & -u_2p_1 & 1-u_2p_2 \end{pmatrix} \tag{2.34}$$

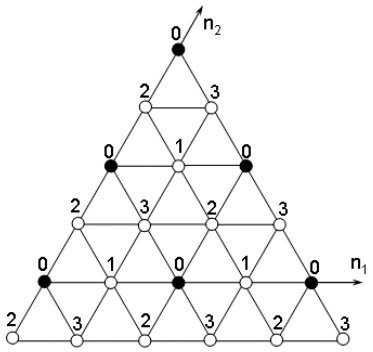


FIG. 2.10 – Grille d'échantillonnage d'un maillage triangulaire. Les points noirs correspondent à $\mathbf{t}_0 = (0,0)$ et les points blancs à $\mathbf{t}_1 = (1,0)$, $\mathbf{t}_2 = (0,1)$ et $\mathbf{t}_3 = (1,1)$.

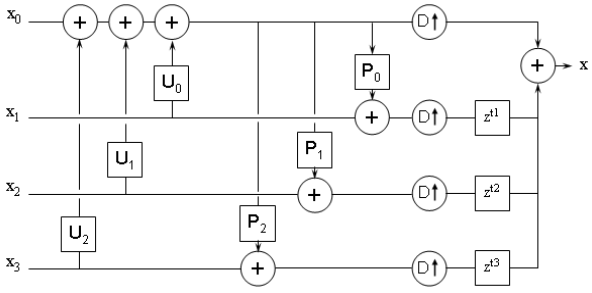


FIG. 2.11 – Schéma lifting de synthèse sur 4 canaux. Les opérateurs p_i et u_i sont respectivement le prédicteur et la mise à jour pour le sous-signal x_i.

2.4.4.2 Les pondérations du filtre Lifting de Butterfly

Il existe plusieurs méthodes pour concevoir des transformées en ondelettes sur des maillages triangulaires semi-réguliers [228], [292]. Elles sont généralement basées sur le schéma d'interpolation de Butterfly donné dans [182], [253]. Nous ne détaillerons pas ici la construction du filtre de Butterfly qui se trouve dans [292] et dans [199] pour la version lifting. Notons simplement que les transformées en z des opérateurs p_i et u_i sont données par les relations (2.35) et (2.36) suivantes :

$$\left\{\begin{array}{l} p_1(z_1,z_2) = \frac{1}{2}\,(z_1{}^1 + \frac{1}{z_1{}^1}) + \frac{1}{8}\,(\frac{z_2{}^2}{z_1{}^1} + \frac{z_1{}^1}{z_2{}^2}) - \frac{1}{16}\,(z_1{}^1 z_2{}^2 + \frac{z_1{}^3}{z_2{}^2} + \frac{1}{z_1{}^1 z_2{}^2} + \frac{z_2{}^2}{z_1{}^3}) \\ p_2(z_1,z_2) = \frac{1}{2}\,(\frac{1}{z_2{}^1} + z_2{}^1) + \frac{1}{8}\,(\frac{z_1{}^2}{z_2{}^1} + \frac{z_2{}^1}{z_1{}^2}) - \frac{1}{16}\,(z_2{}^1 z_1{}^2 + \frac{z_2{}^3}{z_1{}^2} + \frac{1}{z_2{}^1 z_1{}^2} + \frac{z_1{}^2}{z_2{}^3}) \\ p_3(z_1,z_2) = \frac{1}{2}\,(\frac{z_2{}^1}{z_1{}^1} + \frac{z_1{}^1}{z_2{}^1}) + \frac{1}{8}\,(z_1{}^1 z_2{}^1 + \frac{1}{z_1{}^1 z_2{}^1}) - \frac{1}{16}\,(\frac{z_1{}^3}{z_2{}^1} + \frac{z_2{}^3}{z_1{}^1} + \frac{z_1{}^1}{z_2{}^3} + \frac{z_2{}^1}{z_1{}^3}) \end{array}\right. \tag{2.35}$$

et

$$\left\{\begin{array}{lcl} u_1(z_1,z_2) &=& \frac{1}{8}\,\frac{1}{z_1{}^1} + \frac{1}{8}\,z_1 \\ u_2(z_1,z_2) &=& \frac{1}{8}\,\frac{1}{z_2{}^1} + \frac{1}{8}\,z_2 \\ u_3(z_1,z_2) &=& \frac{1}{8}\,\frac{1}{z_1{}^1 z_2{}^1} + \frac{1}{8}\,z_1 z_2 \end{array}\right. \tag{2.36}$$

Par identification avec la matrice polyphase (2.34) il est possible de calculer les pondérations données équation (2.30). Le calcul de ces pondérations est détaillé dans notre article [17]. Elles sont données par :

$$\left\{\begin{array}{lcll} w_0 &=& \frac{N_{s_0}}{N_s}\frac{169}{256} &\simeq \frac{N_{s_0}}{N_s} 0.66015625 \\ w_1 &=& \frac{N_{s_1}}{N_s}\frac{1727}{2048} &\simeq \frac{N_{s_1}}{N_s} 0.843261715 \\ w_2 &=& \frac{N_{s_2}}{N_s}\frac{1727}{2048} &\simeq \frac{N_{s_2}}{N_s} 0.843261715 \\ w_3 &=& \frac{N_{s_3}}{N_s}\frac{1727}{2048} &\simeq \frac{N_{s_3}}{N_s} 0.843261715 \end{array}\right. \tag{2.37}$$

La prise en compte de ces pondérations dans un codeur de maillage 3D est très importante. La figure 2.12 présente des résultats de compression sur l'objet 3D VENUS (cf. figure 2.13) au moyen du codeur 3D multiéchelles développé dans le chapitre 3, avec les pondérations π_i^* optimales ou égales à un sur cinq niveaux de résolution. Les gains obtenus en terme de Rapport Signal-à-Bruit peuvent atteindre plus de 2 dB selon les objets 3D et le débit considérés, par rapport à un codeur qui utilise des pondérations égales à un.

2.5 Conclusion-Synthèse

Les travaux que j'ai mené sur l'insertion de la transformée en ondelettes dans un schéma de compression d'images ont permis la construction des filtres dits "9-7" qui fournissent à l'heure actuelle les meilleurs résultats en compression d'image. Ces filtres ont fait l'objet d'une implantation sur circuit intégré commercialisé par Analog Device sous le nom de ADV601 ainsi que d'une implantation sur DSP par Texas Instrument. Ils sont de plus retenus par toutes les propositions de pointe pour la future norme de compression d'images fixes JPEG-2000 et constituent un des filtres de référence pour cette nouvelle norme. Nous avons d'autre part développé une version "au fil de l'eau" de la transformée en ondelettes permettant de réduire au minimum le coût mémoire nécessaire pour effectuer la transformée. Cette méthode introduite dans le cas des images 2D pour des filtres transverses ou *lifting* (séparables ou quinconces) a été étendue par la suite aux vidéos dans un cadre 2D+t.

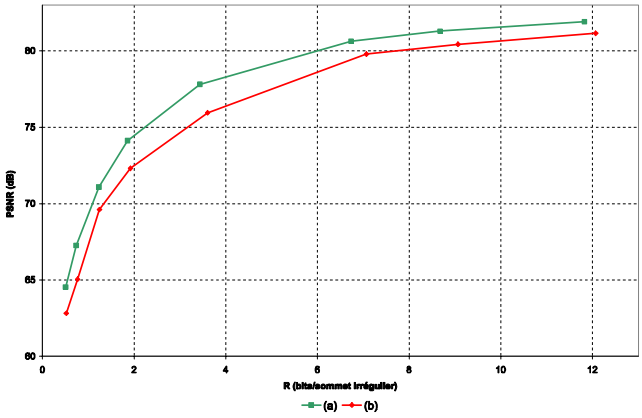


FIG. 2.12 – Rapport Signal-à-Bruit Pic en fonction du débit pour l'objet VENUS. (a) : EQM avec les pondérations π_i^* optimales (b) : pondérations π_i^* égales à 1.

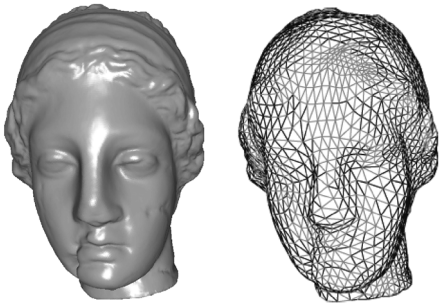


FIG. 2.13 – Objet 3D VENUS et son maillage semi-régulier. Le maillage irrégulier d'origine comporte 50002 sommets et 100000 triangles.

2.6 Références

[159] E.H Adelson and E. Simoncelli, "Non-separable extensions of quadrature mirror filters to multiple dimensions," *Proc. of the IEEE*, vol. 78, avril 1990.

[160] P. Alliez, D. Cohen-Steiner, O. Devillers, Bruno Levy, and Mathieu Desbrun, "Anisotropic polygonal remeshing," *ACM Transactions on Graphics. Special issue for SIGGRAPH conference*, 2003, To appear.

[161] R. Ansari and C. Guillemot, "Exact reconstruction filter banks using diamond fir filters," *Proc. Bilcon 1990, Elsevier Press*, pp. 1412–1424, juillet 1990.

[162] R. Ansari, C. W. Kim, and M. Dedovic, "Sructure and design of two-channel filter banks derived from a triplet of halfband filters," *IEEE Transactions on Circuits and Systems II*, pp. 1487–1496, decembre 1999.

[163] N. Baaziz and C. Labit, "Transformations pyramidales d'images numérique," *Rapport Interne IRISA*, , no. 526, mars 1990.

[164] M. Barlaud, *Wavelet and Image Communication*, Jan Biemond. Elsevier, 1994.

[165] G. Battle, "A block spin construction of wavelets. part i : Lemarié functions," *Comm. Math. Phys.*, vol. 110, pp. 601–615, 1987.

[166] P.J. Burt and E.H Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[167] A.R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Applied and Computational Harmonic Analysis*, vol. 5, no. 3, pp. 332–369, 1998.

[168] E. J. Candès and D. L. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges, curves and surfaces," *L. L. Schumaker et al. (eds), Vanderbilt University Press, Nashville, TN*, 1999.

[169] J.M. Chassery and J. Waku, "Spécification d'une ondelette pour la représentation en multirésolution d'un contour discret," *Traitement du Signal*, vol. 10, no. 3, pp. 231–240, 1993.

[170] C. Chrysafis and A. Ortega, "Line based reduced memory wavelet image compression," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 378–389, 2000.

[171] A. Cohen, I. Daubechies, and J. Feauveau, "Bi-orthogonal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 45, pp. 485–560, 1992.

[172] J. Little D. Cox and D. O'Shea, *Ideal,Varieties, and Algorithms : An Introduction to computation Algebraic Geometry and Commutative Algebra*, undergraduate Texts in mathematics, Springer, 1992.

[173] A. Croisier, D. Esteban, and C. Galand, "Perfect channel splitting by use of interpolation-decimation-tree decomposition techniques," in *Internationale Conference on Info. Sciences and Systems*, Grèce, août 1976, pp. 443–446.

[174] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *Journal of math. and phys.*, vol. 27, pp. 1271–1283, 1986.

[175] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, pp. 909–996, 1988.

[176] I. Daubechies, "The wavelet transform, time-frequency localization and signals analysis," *IEEE Transactions on Information Theory*, vol. 36, pp. 961–1005, 1990.

[177] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Series in Appl. Math., Vol. 61 Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.

[178] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.

[179] P. Delsarte, B. Macq, and D. T. M. Slock, "Signal-adapted multiresolution transform for image coding," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 897–904, 1992.

[180] G. Deslauriers and S. Dubuc, "Symmetric iterative interpolation processes," *Constr. Approx.*, vol. 5, no. 1, pp. 49–68, 1989.

[181] M.N. Do and M. Vetterli, "Orthonormal finite ridgelet transform for image compression," *International Conference on Image Processing, Vancouver, BC, Canada*, 2000.

[182] N. Dyn, D. Levin, and J. Gregory, "A butterfly subdivision scheme for surface interpolation with tension control," *ACM Transactions on Graphics*, vol. 9, no. 2, pp. 160–169, 1990.

[183] J.C. Feauveau, *Analyse Multirésolution par Ondelettes non Orthogonales et Bancs de Filtres Numériques*, Ph.D. thesis, Université Paris Sud, 1990.

[184] J.C. Feauveau, "Analyse multirésolution pour les images avec un facteur de résolution $\sqrt{2}$," *Traitement du signal*, vol. 7, no. 2, 1990.

[185] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[186] A. Gouze, *Schéma Lifting Quinconce pour la Compression d'Images*, Ph.D. thesis, Université de Nice - Sophia Antipolis, France, décembre 2002.

[187] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.*, vol. 15, no. 4, pp. 723–736, 1984.

[188] I. Guskov, K. Vidimce, W. Sweldens, and P. Schroder, "Normal meshes," in *Computer Graphics Proceedings, SIGGRAPH*, 2000, pp. 95–102.

[189] A. Haar, "Zur theorie der orthogonalen funktionen-systeme," *Math. Ann.*, vol. 69, pp. 331–371, 1910.

[190] *ISO MPEG4 Standard*.

[191] Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, *Joint Committee Draft, JVT-C167*, May 2002.

[192] A.E. Jacquin, "Fractal image coding : A review," *Proceedings of IEEE*, vol. 81, no. 10, pp. 1451–1465, 1993.

[193] N. Jayant and P. Noll, *Digital Coding of Waveforms*, Englewood Cliffs, New Jersey : Prentice Hall, 1984.

[194] ISO and CCITT, "Digital compression and coding of continuous-tone still images," ISO/IEC 10918-1,2,3 Information Technology.

[195] ISO/IEC 15444-1 :2000, "Information technology – JPEG 2000 image coding system – part 1 : Core coding system," .

[196] G. Karlsson and M. Vetterli, "Three-dimensionnal subband coding of video," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, New York, Etat-Unis, avril 1988, vol. 2, pp. 1100–1103.

[197] C. W. Kim and R. Ansari, "Subband decomposition procedure for quincunx sampling grids," in *SPIE Conference on Visual Communication and Image Processing*, novembre 1991, pp. 112–123.

[198] J. Kovacevic and M. Vetterli, "Non-separable multidimensional perfect reconstruction filter banks and wavelets bases for $\mathbb{R}^n$," *IEEE Transactions on Information theory*, vol. 38, no. 2, pp. 533–555, mars 1992.

[199] J. Kovacevic and W. Sweldens, "Wavelet families of increasing order in arbitrary dimensions," *IEEE Transactions on Image Processing*, vol. 9, no. 3, 1999.

[200] J. Kovacevic and M. Vetterli, "Nonseparable two- and three-dimensional wavelets," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1269–1273, mai 1995.

[201] C. Latry and B. Rougé, "Spot5 thr mode," in *Visual Communication and Image Processing*, Etats-Unis, 1998, SPIE.

[202] C. Latry and B. Rougé, "Optimized sampling for ccd instruments : The supermode scheme," in *IGARSS*, Etats-Unis, 2000.

[203] A.W.F Lee, W. Sweldens, P. Schröder, P. Cowsar, and D. Dobkin, "Maps : Multiresolution adaptive parametrization of surfaces," *SIGGRAPH*, 1998.

[204] D. Le Gall and A. Tabatabai, "Sub-band coding of digital images using symmetric short kernel filters and arithemetic coding techniques," in *Proc. Int. Conf. Acoust. Speech and Signal Processing*, Vancouver, avril 1988, pp. 761–764.

[205] P.G. Lemarié, "Une nouvelle base d'ondelettes de $l^2(r^n)$," *Journal de Math. pures et Appl.*, vol. 67, pp. 227–238, 1988.

[206] M. Lounsbery, T. DeRose, and J. Warren, "Multiresolution analysis for surfaces of arbitrary topological type," *ACM Transactions on Graphics 16,1*, vol. 99, 1997.

[207] J.M. Lounsbery, *Multiresolution Analysis for Surfaces of Arbitrary Topological Type*, Ph.D. thesis, Seattle, WA, Etats-Unis, 1994.

[208] L. Luo, J. Li, S. Li, Z. Zhuang, and Y.-Q. Zhang, "Motion compensated lifting wavelet and its application in video coding," in *IEEE International Conference on Multimedia and Expo*, Tokyo, août 2001, pp. 481–484.

[209] B. Macq and J. Mertès, "Optimization of linear multiresolution transforms for scene adaptive coding," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3568–3572, 1993.

[210] S. Mallat, "A theory for multiresolution signal decomposition : The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, juillet 1989.

[211] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1997.

[212] H.S. Malvar and D.H. Staelin, "The LOT : Transform coding without blocking effects," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 4, pp. 553–559, avril 1989.

[213] X. Marichal and B. Macq, "Asymetric motion estimation/compensation," in *IEEE International Conference on Image Processing, Lausanne, Suisse*, septembre 1996.

[214] G. Marquant, S. Pateux, and C. Labit, "Mesh-based scalable video coding with rate-distortion optimization," in *SPIE Visual Communication and Image Processing, Perth, Australie*, juin 2000, vol. 4067, pp. 967–976.

[215] J. McClellan, "The design of two-dimensional filters by transformations," in *Seventh Ann. Princeton Conference on ISS*, Princeton, NJ, 1973, pp. 247–251.

[216] Y. Meyer, "Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs," 1985-1986.

[217] Y. Meyer, *Ondelettes et Opérateurs, I : Ondelettes, II : Opérateurs de Calderón-Zygmund, III : (with R. Coifman), Opérateurs Multilinéaires)*, Hermann, Paris, 1990.

[218] F. Moreau de Saint-Martin, P. Siohan, and A. Cohen, "Biorthogonal filterbanks and energy preservation property in image compression," *IEEE Transactions on Image Processing*, vol. 8, no. 2, 1999.

[219] J.R. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 559–571, septembre 1994.

[220] E. Ordentlich, D. Taubman, M. Weinberger, G. Seroussi, and M.W. Marcellin, "Memory efficient scalable line-based image coding," in *Proceedings of Data Compression Conference*, Snowbird, Etats-Unis, 1999.

[221] C. Parisot, *Allocations Basées Modèles et Transformée en Ondelettes au Fil de l'Eau pour le Codage d'Images et de Vidéos*, Ph.D. thesis, Université de Nice - Sophia Antipolis, 2003.

[222] H. Park, T. Kalker, and M. Vetterli, "Gröbner bases and multidimensional fir multirate systems," *Journal of multidimensional systems and signal processing*, vol. 8, pp. 11–30, 1997.

[223] F. Payan, *Compression Multirésolution de Maillages Géométriques 3D*, Ph.D. thesis, Université de Nice - Sophia Antipolis, soutenance prévue en décembre 2003.

[224] E. Pennec and S. Mallat, "Image compression with geometrical wavelets," *International Conference on Image Processing, Vancouver, BC, Canada*, vol. 1, pp. 661–664, 2000.

[225] O. Rioul, "Simple regularity criteria for subdivision schemes," *SIAM Jouranal Math. Anal.*, vol. 23, no. 6, pp. 1544–1576, novembre 1992.

[226] J. Romberg, M. Wakin, and R. Baraniuk, "Approximation and compression of piecewise smooth images using a wavelet/wedgelet geometric model," *International Conference on Image Processing, Barcelone, Espagne*, 2003.

[227] P. Salembier and H. Sanson, "Robust motion estimation using connected operators," in *IEEE International Conference on Image Processing, Santa Barbara, Etats-Unis*, octobre 1997.

[228] P. Schröder and W. Sweldens, "Spherical wavelets : Efficiently representing functions on the sphere," *Proceedings of SIGGRAPH 95*, pp. 161–172, 1995.

[229] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," in *Proceedings of IEEE International Conference on Image Processing*, Thessalonique, Grèce, octobre 2001, pp. 1029–1032.

[230] P. Siohan and V. Ouvrard, "Design of two-dimensional non-separable qmf banks," in *Procedings of the IEEE ICASSP conference*, San Francisco, Etats-Unis, 1992, vol. 4, pp. 645–648.

[231] M.J. Smith and K.M. Barnwell, "Exact reconstruction for tree-structured subband coders," *IEEE Transactions ASSP*, vol. 34, pp. 434–441, 1986.

[232] J.O. Stromberg, *A Modified Franklin System and Higher Order Spline System on* $\mathbf{R}^n$ *as Unconditional Bases for Hardy Spaces*, vol. II, Conf. in honor of A. Zygmund, W. Beckner et al., ed., Wadsworth math. series, 1982.

[233] A. Suslin, "Projective modules over a polynomial ring are free," *Dokl. Akad. Nauk. SSSR*, vol. 229, pp. 1063–1066, 1976, (traduction anglaise du russe : Soviet. Math. Dokl., Vol. 17, 1976, p. 1160- 1164).

[234] A. Suslin, "On the structure of the special linear group over polynomial rings (russian)," *Izv. AN. Nauk SSSR*, vol. 41, pp. 235–252, 1977, Traduction : Math. USSR, Izv. 11, 221-238 (1977).

[235] W. Sweldens, "The lifting scheme : A new philosophy in biorthogonal wavelet constructions," in *Wavelet Applications in Signal and Image Processing III*, A. F. Laine and M.Unser, Eds., pp. 68–79. Proc. SPIE 2569, 1995.

[236] W. Sweldens and P. Schröder, "Building your own wavelets at home," in *Wavelets in Computer Graphics*, pp. 15–87. ACM SIGGRAPH Course Notes, 1996.

[237] D. Taubman and A. Secker, "Highly scalable video compression with scalable motion coding," in *IEEE International Conference on Image Processing, Barcelone, Espagne*, septembre 2003.

[238] P. Tchamitchian, *Biorthogonalité et Théorie des Opérateurs*, vol. 3, Rev. Math. Iberoamer., 1987.

[239] M. Unser and T. Blu, "Mathematical properties of the jpeg-2000 wavelet filters," *Transactions on Image Processing*, vol. 12, no. 9, pp. 1080–1090, septembre 2003.

[240] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proceedings of Data Compression Conference*, Snowbird, USA, mars 1996, pp. 387–395.

[241] G. Uytterhoeven and A. Bultheel, "The red-black wavelet transform," Tech. Rep., Department of Computer Science, K.U. Leuven, 1997.

[242] S. Valette, Y.S. Kim, H.Y. Jung, I. Magnin, and R. Prost, "A multiresolution wavelet scheme for irregularly subdivided 3d triangular mesh," *IEEE International Conference on Image Processing, Kobe, Japon*, vol. 1, pp. 171–174, octobre 1999.

[243] M. Vetterli, "Multidimensionnal subband coding : Some theory and algorithms," *Signal Processing*, 1984.

[244] M. Vetterli, "Filter banks allowing perfect reconstruction," *Signal Processing*, vol. 10, pp. 219–244, 1986.

[245] M. Vetterli and C. Herley, "Wavelets and filter banks : Relationships and new results," in *IEEE ICASSP*, Albuquerque, Etats-Unis, avril 1990, pp. 1723–1726.

[246] M. Vetterli and C. Herley, "Wavelets and filter banks : Theory and design," *IEEE Transactions on Acoustic Speech Signal Processing*, vol. 40, no. 9, pp. 2207–2232, 1992.

[247] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[248] J. Viéron, C. Guillemot, and S. Pateux, "Motion compensated 2d+t wavelet analysis for low rate fgs video compression," in *Proceedings of Tyrrhenian International Workshop on Digital Communications*, Capri, Italie, septembre 2002.

[249] J.D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Transactions on Image Processing*, vol. 4, no. 8, août 1995.

[250] M. Vishwanath, "The recursive pyramid algorithm for the discrete wavelet transform," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 673–676, mars 1994.

[251] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk, "Geometric methods for wavelet-based image compression," in *International Symposium on Optical Science and Technology*, San Diego, CA, August 2003.

[252] J. Xu, S. Li, Y.-Q. Zhang, and Z. Xiong, "A wavelet video coder using three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," in *Proceedings of IEEE Pacific-Rim Conference on Multimedia*, Sydney, Australia, décembre 2000.

[253] D. Zorin, P. Schroder, and W. Sweldens, "Interpolating subdivision for meshes with arbitrary topology," in *Proceedings of SIGGRAPH*, 1996, pp. 189–192.

Chapitre 3

Le compromis débit-distorsion

Dans ce chapitre je développe les activités de recherche que j'ai effectuées ou que j'effectue encore dans le domaine lié à l'optimisation du compromis débit-distorsion pour des applications de compression d'images (2D ou 3D) et de vidéos. Le plan de ce chapitre est le suivant. Tout d'abord j'introduis dans le paragraphe 3.1 le problème général de l'allocation des ressources binaires et développe dans le paragraphe 3.2 la solution d'allocation de la "qualité" ou de débits que nous avons développée, basée sur l'utilisation de modèles pour l'estimation de la distorsion et du débit. Le paragraphe 3.3 concerne les travaux que nous avons menés dans le cadre de la transmission d'images et de vidéos sur des canaux bruités de troisième génération et principalement la solution d'allocation des ressources pour un codage source/canal basée sur les descriptions multiples. Enfin, dans le paragraphe 3.4 je présente les premiers travaux que nous avons effectués dans le domaine de la compression de maillage 3D multirésolution.

3.1 Problème de l'allocation des ressources binaires

La base du codage de source repose sur la théorie dite du "débit-distorsion" introduite par Shannon à la fin des années 40 [311], [312]. Ces travaux traitent de la minimisation de la distorsion liée au codage de la source sous la contrainte d'un débit canal ou de son problème dual, la minimisation du débit canal sous la contrainte d'une distorsion source. Un système de compression est constitué par un ensemble fini de quantificateurs admissibles, tous caractérisés par leur fonction débit-distorsion $D(R)$ [274]. Le problème de l'allocation de débits qui se pose alors, consiste à distribuer de façon efficace les ressources binaires disponibles au niveau de la source parmi un ensemble fini de N signaux (ou de sous-bandes) sources. C'est un problème (P) multidimensionnel, désormais

classique en compression des signaux, qui se formule de la façon générale suivante :

$$(P)\begin{cases} \min\limits_{R_1,R_2,...,R_N} D\left(R_1,R_2,...,R_N\right) \\ \text{sous la contrainte } a_1R_1+a_2R_2+...+a_NR_N \leq R_{cible}. \end{cases} \tag{3.1}$$

La principale difficulté du problème réside dans l'estimation de la fonction débit-distorsion. Historiquement, certains résultats apparaissent dès 1898 dans les travaux de Sheppard [314]. Cependant, les premiers travaux importants sur le sujet datent de 1948. Ils sont basés sur deux approches complémentaires : la *théorie de l'information* initiée par Shannon dès 1948 [311] et complétée en 1959 dans [312], et la théorie *haute résolution* introduite par Olivier, Pierce et Shannon dans [296] ainsi que par les travaux de Bennett [258] la même année et de Panter et Dite en 1951 [298]. En 1966, les travaux de Zador [339] introduisent la quantification vectorielle comme étant une limite fondamentale des performances de quantification. Parallèlement à ces travaux théoriques asymptotiques, il existe quelques travaux non asymptotiques. Les plus anciens sont ceux de Clavier, Panter et Grieg qui en 1947 [267],[268] font une analyse exacte de l'erreur de quantification pour des sources sinusoïdales quantifiées uniformément, ou encore ceux de Bennett en 1948 [258] qui donnent une formulation de la densité spectrale de puissance d'un processus aléatoire Gaussien quantifié uniformément. Cependant, les *conditions d'optimalité* d'un quantificateur, résultat fondamental en théorie non asymptotique, ont été introduites par Steinhaus en 1956 [318] et Lloyd en 1957 [290]. Ces travaux ont été par la suite popularisé par Max en 1960 [293]. La théorie haute résolution fournie des équations qui permettent de mesurer les performances d'un quantifcateur. La plupart des articles dans la littérature se sont concentré sur l'estimation de la distorsion pour des quantificateurs à débit fixe. Néanmoins, la théorie asymptotique peut aussi être utilisée pour estimer l'entropie [275]. Un historique détaillé sur les théories asymptotiques et non asymptotiques est donné dans le papier de Gray et Neuhoff [280].

Les méthodes d'allocation de débits qui nécessitent une connaissance de la relation débit-distorsion, peuvent être classées en deux grandes catégories [299] :

- **Le basé signal**. Cette catégorie comprend les méthodes d'allocation basées sur la mesure simultanée du débit et de la distorsion réels. On retrouve dans cette catégorie toutes les méthodes de la famille EZW [313] et SPIHT [307] ainsi que JPEG-2000 [282].
- **Le basé modèle**. Dans ce cas, la procédure d'allocation est faite de façon indépendante à la procédure de quantification et d'encodage. Elle se base sur une modélisation des fonctions débit-distorsion. Les modèles peuvent

être asymptotiques ou encore non-asymptotiques. **C'est dans ce cadre que se situent nos travaux de recherche**.

L'allocation de débit a été étudié depuis de nombreuses années et des solutions ont été données entre autres par les travaux de Segall en 1976 [308], Shoham et Gersho en 1988 [317], Chou, Lookabaugh et Gray (BFOS) en 1986 [265], Westerink, Biemond and Boekee en 1988 [334], Ramchandran et Vetterli en 1993 [303] et Kasner, Marcellin et Hunt en 1999 [285].

Notre contribution dans ce domaine a consisté à introduire des modèles exacts pour le débit et la distorsion dans le cas de la quantification scalaire uniforme et à développer un algorithme d'allocation rapide et de faible complexité, concurrent en terme de performances avec le standard JPEG-2000 pour les images fixes.

3.2 Solution proposée basée sur des modèles

Ces travaux on été réalisés durant la thèse de Christophe Parisot (2003) [299] que j'ai co-encadré en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Une partie de nos travaux a été publiée dans la revue EURASIP Journal on Applied Signal Processing en janvier 2003 : "3D scan based wavelet transform and quality control for video coding" [12]. Cet article est donné en annexe B du document. Une partie de nos travaux fait l'objet d'un dépôt de brevet conjoint CNRS-CNES en cours [23].

3.2.1 Le basé modèle

3.2.1.1 Idée

Les méthodes classiques de détermination des pas de quantification pour les différentes sous-bandes d'une transformée en ondelettes requièrent un codage exhaustif pour calculer les courbes débit-distorsion et retenir la meilleure solution [317], [282]. A l'inverse de ces méthodes, nous avons proposé un algorithme d'allocation de débits (ou de pas de quantifications) basé sur des modèles théoriques de distorsion et de débit [70]. Comme nous allons le voir aux paragraphes 3.2.1.2 et 3.2.1.3, ces modèles continus sont basés sur la connaissance de la distribution statistique des données sources. Ils permettent d'obtenir un ensemble de solutions continu.

3.2.1.2 Modèle non asymptotique pour la distorsion

De façon générale, la distorsion granulaire d'un quantificateur (puissance du bruit de quantification) est définie par la relation :

$$\sigma_Q^2 = E\left[d\left(S,\hat{s}\right)\right] = \sum_{m=-\infty}^{+\infty} \int_{x\in P^m} d\left(x,\hat{s}_m\right) p_S\left(x\right) dx, \tag{3.2}$$

où d est la mesure de distorsion entre un symbole source x et sa quantification $\hat{s}_m$, P^m est la classe de quantification à laquelle appartient le symbole source représentée par son centroïde $\hat{s}_m$ et $p_S\left(x\right)$ la densité de probabilité du signal source S.

L'objectif de nos recherches sur la distorsion a consisté à développer une expression analytique permettant de calculer exactement la formule de distorsion (3.2) pour des quantificateurs scalaires uniformes à pas de quantification q et à zone morte de largeur z dont le décodage utilise le centroïde comme représentant de l'intervalle [259], [260]. Dans ce cas, la classe P^m correspond simplement à l'intervalle de quantification $\left[-\frac{z}{2},\frac{z}{2}\right[$ pour $m=0$ et à $\left[mq-\frac{q}{2},mq+\frac{q}{2}\right[$ sinon. Nous avons montré, dans le cas où d est la distance Euclidienne au carré et que la densité de probabilité $p_S\left(x\right)$ est symétrique paire, que l'Erreur Quadratique Moyenne (EQM) de quantification pouvais s'écrire sous la forme [299], [12] :

$$\sigma_Q^2 = \sigma^2 - 2\sum_{m=1}^{+\infty} \frac{\left(\int_{\frac{z}{2}+(m-1)q}^{\frac{z}{2}+mq} x p_{\sigma,\alpha}(x)dx\right)^2}{\int_{\frac{z}{2}+(m-1)q}^{\frac{z}{2}+mq} p_{\sigma,\alpha}(x)dx}, \tag{3.3}$$

avec σ^2 la variance de la source. Cette formule est très intéressante car elle ne dépend que de la connaissance de la densité de probabilité du signal source et fournie une valeur exacte de l'EQM.

Dans le cadre de la compression de signaux sources issus d'une transformation en ondelettes (qui est le cadre principal de nos travaux) le modèle "naturel" pour la densité de probabilité $p_S(x)$ est donné par la Gaussienne généralisée [2]. Pour ce type de distribution, nous avons démontré [299] la proposition 2 suivante qui permet une simplification dans le calcul de l'EQM donné par l'équation (3.3).

Proposition 2 *Lorsque la densité de probabilité d'un signal source est une distribution Gaussienne généralisée d'écart type σ et de paramètre de forme α, notée $p_{\sigma,\alpha}\left(x\right)$, il existe une famille de fonctions $f_{n,m}$ vérifiant*

$$\int_{-\frac{z}{2}}^{+\frac{z}{2}} x^n p_{\sigma,\alpha}(x)dx = \sigma^n f_{n,0}\left(\alpha,\frac{z}{\sigma}\right)$$

$$\int_{\frac{z}{2}+(m-1)q}^{\frac{z}{2}+mq} x^n p_{\sigma,\alpha}(x)dx = \sigma^n f_{n,m}\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)$$

avec

$$f_{n,0}\left(\alpha,\tfrac{z}{\sigma}\right)=\int_{-\frac{1}{2}\frac{z}{\sigma}}^{+\frac{1}{2}\frac{z}{\sigma}}x^n p_{1,\alpha}(x)dx \quad et \quad f_{n,m}\left(\alpha,\tfrac{z}{\sigma},\tfrac{q}{\sigma}\right)=\int_{\frac{1}{2}\frac{z}{\sigma}+(m-1)\frac{q}{\sigma}}^{\frac{1}{2}\frac{z}{\sigma}+m\frac{q}{\sigma}}x^n p_{1,\alpha}(x)dx$$

Preuve. Soit

$$p_{\sigma,\alpha}(x)=\frac{A(\alpha)}{\sigma}e^{-|B(\alpha)\frac{x}{\sigma}|^\alpha}$$

une densité de probabilité Gaussienne généralisée d'écart type σ et de paramètre de forme α avec $B(\alpha)=\sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}$ et $A(\alpha)=\frac{\alpha B(\alpha)}{2\Gamma(1/\alpha)}$.

Calculons l'intégrale suivante :

$$I=\int_{x_1}^{x_2}x^n p_{\sigma,\alpha}(x)dx=\int_{x_1}^{x_2}x^n\frac{A(\alpha)}{\sigma}e^{-|B(\alpha)\frac{x}{\sigma}|^\alpha}dx$$

En utilisant le changement de variable $X=\frac{x}{\sigma}$, nous obtenons

$$\begin{aligned}I&=\int_{\frac{x_1}{\sigma}}^{\frac{x_2}{\sigma}}\sigma^n X^n\frac{A(\alpha)}{\sigma}e^{-|B(\alpha)X|^\alpha}\sigma dX\\&=\sigma^n\int_{\frac{x_1}{\sigma}}^{\frac{x_2}{\sigma}}X^n A(\alpha)e^{-|B(\alpha)X|^\alpha}dX\\&=\sigma^n\int_{\frac{x_1}{\sigma}}^{\frac{x_2}{\sigma}}X^n p_{1,\alpha}(X)dX\end{aligned}$$

Nous avons donc

$$\int_{-\frac{z}{2}}^{+\frac{z}{2}}x^n p_{\sigma,\alpha}(x)dx=\sigma^n\int_{-\frac{1}{2}\frac{z}{\sigma}}^{+\frac{1}{2}\frac{z}{\sigma}}X^n p_{1,\alpha}(X)dX \tag{3.4}$$

et

$$\int_{\frac{z}{2}+(m-1)q}^{\frac{z}{2}+mq}x^n p_{\sigma,\alpha}(x)dx=\sigma^n\int_{\frac{1}{2}\frac{z}{\sigma}+(m-1)\frac{q}{\sigma}}^{\frac{1}{2}\frac{z}{\sigma}+m\frac{q}{\sigma}}X^n p_{1,\alpha}(X)dX \tag{3.5}$$

De (3.4) et (3.5) nous déduisons qu'il existe une famille de fonctions $f_{n,m}$ vérifiant

$$\int_{-\frac{z}{2}}^{+\frac{z}{2}}x^n p_{\sigma,\alpha}(x)dx=\sigma^n f_{n,0}\left(\alpha,\frac{z}{\sigma}\right)$$

$$\int_{\frac{z}{2}+(m-1)q}^{\frac{z}{2}+mq}x^n p_{\sigma,\alpha}(x)dx=\sigma^n f_{n,m}\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)$$

avec

$$f_{n,0}\left(\alpha,\tfrac{z}{\sigma}\right)=\int_{-\frac{1}{2}\frac{z}{\sigma}}^{+\frac{1}{2}\frac{z}{\sigma}}x^n p_{1,\alpha}(x)dx \quad \text{et} \quad f_{n,m}\left(\alpha,\tfrac{z}{\sigma},\tfrac{q}{\sigma}\right)=\int_{\frac{1}{2}\frac{z}{\sigma}+(m-1)\frac{q}{\sigma}}^{\frac{1}{2}\frac{z}{\sigma}+m\frac{q}{\sigma}}x^n p_{1,\alpha}(x)dx$$

■

En introduisant les fonctions $f_{n,m}$ données proposition 2 nous pouvons simplifier l'écriture de l'EQM donnée équation (3.3), sous la forme :

$$\sigma_Q^2=\sigma^2\left[1-2\sum_{m=1}^{+\infty}\frac{f_{1,m}\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)^2}{f_{0,m}\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)}\right]. \tag{3.6}$$

Ainsi, pour une source de distribution Gaussienne généralisée, l'EQM de quantification est égale à la variance σ^2 de la source multipliée par une fonction qui ne dépend que de $\alpha,\frac{z}{\sigma},\frac{q}{\sigma}$. Nous pouvons donc écrire :

$$\sigma_Q^2=\sigma^2 D\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right), \tag{3.7}$$

avec

$$D\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)=1-2\sum_{m=1}^{+\infty}\frac{f_{1,m}\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)^2}{f_{0,m}\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)}. \tag{3.8}$$

D est de classe C^∞ pour $z>0$ et $q>0$ [299]. Une étude complète est donnée dans [299]. L'allure de la fonction $D\left(\alpha,\frac{z}{\sigma},\frac{q}{\sigma}\right)$ pour $z=q$ et pour différentes valeurs de α est donnée sur la figure 3.1.

3.2.1.3 Modèle non asymptotique pour le débit

Le codage des coefficients quantifiés étant fait au moyen d'un codeur entropique, il semble naturel d'approximer le débit de sortie de l'encodeur par l'entropie du signal quantifié donné par :

$$H=-\sum_{m=-\infty}^{+\infty}\Pr(m)\log_2\Pr(m), \tag{3.9}$$

où, $\Pr(m)=\Pr\left(S\in P^m\right)$ est la probabilité que l'échantillon source S appartienne à la classe P^m du quantificateur. Dans le cas d'un quantificateur scalaire uniforme à zone morte z l'expression de $\Pr(m)$ est donnée par :

$$\begin{cases}\Pr(0)=\int_{-\frac{z}{2}}^{\frac{z}{2}}p_S(x)dx \quad \text{pour } m=0\\ \Pr(m)=\int_{\frac{z}{2}+(m-1)q}^{\frac{z}{2}+mq}p_S(x)dx \quad \text{sinon.}\end{cases} \tag{3.10}$$

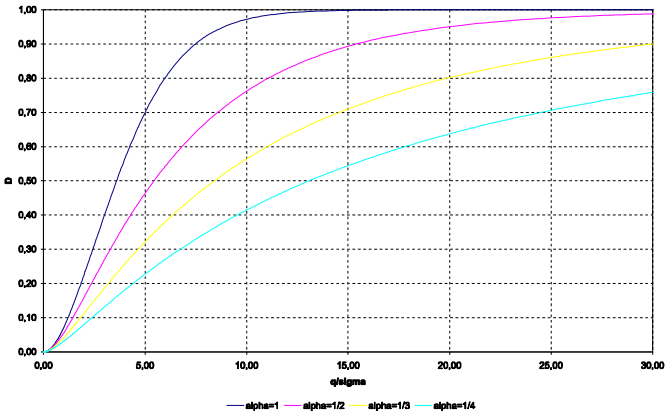


FIG. 3.1 – Distorsion D (EQM normalisée) pour un quantificateur scalaire uniforme sans zone morte ($z = q$) en fonction du pas de quantification q et pour différentes valeurs du paramètre α de la distribution Gaussienne généralisée.

Dans le cas où la densité de probabilité du signal source est une Gaussienne généralisée et en utilisant la proposition 2, il est évident de déduire que :

$$\begin{cases} \Pr(0) = f_{0,0}\left(\alpha, \frac{z}{\sigma}\right) & \text{pour } m = 0 \\ \Pr(m) = f_{0,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right) & \text{sinon.} \end{cases} \tag{3.11}$$

Il nous est alors possible d'exprimer l'entropie par la relation suivante [299],

$$H = -f_{0,0}\left(\alpha, \frac{z}{\sigma}\right) \log_2 f_{0,0}\left(\alpha, \frac{z}{\sigma}\right) - 2\sum_{m=1}^{+\infty} f_{0,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right) \log_2 f_{0,m}\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right). \tag{3.12}$$

Cette relation est intéressante car elle permet de montrer que l'entropie est une fonction qui ne dépend que de α, $\frac{z}{\sigma}$ et $\frac{q}{\sigma}$. Le débit R de sortie du quantificateur peut alors être approximé par :

$$R\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right) \simeq H\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right). \tag{3.13}$$

R est de classe C^∞ pour $z > 0$ et $q > 0$ [299]. L'allure de la fonction $R\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right)$ pour $z = q$ et pour différentes valeurs de α est donnée sur la figure 3.2.

Enfin, la courde débit-distorsion qui résulte des modélisations de $D\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right)$ et $R\left(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma}\right)$ est une fonction convexe donnée sur la figure 3.3.

Cette première approche, bien que simple, produit des résultats intéressant en terme d'allocation de débits comme nous allons le voir dans les paragraphes

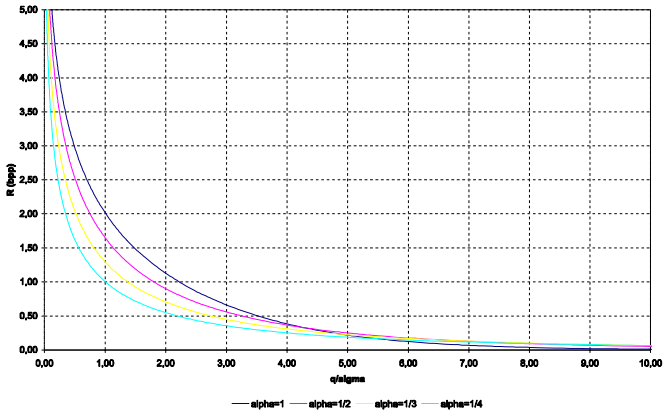


FIG. 3.2 – Débit R pour un quantificateur scalaire uniforme sans zone morte ($z = q$) en fonction du pas de quantification q et pour différentes valeurs du paramètre α de la distribution Gaussienne généralisée.

suivants. Cependant, dans le soucis d'améliorer les performances d'allocation et compte tenu que l'encodage réel des coefficients quantifiés se fait en général au moyen de codeurs entropiques contextuels, il serait intéressant d'intégrer un modèle d'entropie conditionnelle dans le critère. Ce travail constitue une suite logique de notre approche, mais il ne se fera pas sans difficultés, car dans ce cas il faut estimer et modéliser les probablités conditionnelles liées au signal source quantifié.

3.2.2 Solution proposée

3.2.2.1 Allocation de débits ou de "qualité" ?

En général, les codeurs d'images et de vidéos principalement appliqués à des problèmes de transmission "temps réel" sont basés sur une allocation des ressources binaires au niveau des N sous-bandes de coefficients issus de la transformée. Dans ce cas, la résolution du problème (P) d'allocation des débits peut se résoudre par une approche Lagrangienne. Il s'agit alors de minimiser par rapport à $\{R_N\}$ la fonctionnelle suivante :

$$J_\lambda(\{R_N\}) = D(R_1, R_2, ..., R_N) + \lambda [R_T(R_1, R_2, ..., R_N) - R_{cible}]. \quad (3.14)$$

où R_T désigne une fonction qui permet de calculer le débit total.

Cependant, certaines applications telles que l'archivage ou la transmission différée d'images ou de vidéos de très haute qualité nécessite au contraire un contrôle de la qualité image. Dans cette optique, la fonctionnelle à minimiser

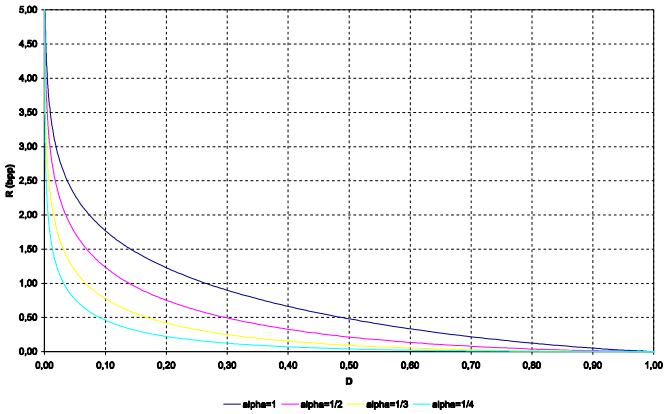


FIG. 3.3 – Fonction débit-distorsion (EQM normalisée) pour plusieurs paramètres α de la distribution Gaussienne généralisée dans le cas d'un quantificateur scalaire uniforme sans zone morte ($z = q$).

en $\{D_N\}$ n'est plus donnée par (3.14), mais par son dual :

$$J_\lambda\left(\{D_N\}\right) = R\left(D_1, D_2, ..., D_N\right) + \lambda\left[D_T\left(D_1, D_2, ..., D_N\right) - D_{cible}\right]. \quad (3.15)$$

où D_T désigne une fonction qui permet de calculer la distorsion totale.

Nous avons étudié ces deux solutions duales et proposé une méthode d'allocation des ressources "débit" ainsi qu'une méthode d'allocation des ressources "qualité" dans les différentes sous-bandes issues de la transformée en ondelettes. Dans la suite du document, nous allons focaliser sur la résolution de l'allocation de "qualité". Les travaux concernant la solution duale "débit" sont similaires à ceux présentés sur la "qualité" et peuvent être trouvés dans [299].

3.2.2.2 Problème d'allocation sous-contrainte "qualité"

Le problème d'allocation de qualité consiste à allouer les ressources binaires disponibles lorsque plusieurs sources différentes doivent être codées, de façon à minimiser le débit total de sortie sous la contrainte d'une distorsion de reconstruction cible D_{cible} qui définit la "qualité". La mesure de "qualité" peut par exemple être donnée par l'EQM. Dans le cas de l'analyse multirésolution cela consiste à distribuer les ressources binaires aux différentes sous-bandes de coefficients d'ondelettes et à la sous-bande de basse fréquence. Dans notre approche, la distribution des ressources binaires est implicite. En effet, notre objectif est plutôt de déterminer les paramètres du quantificateur à utiliser dans chaque sous-bande et non pas directement le débit qui leur est associé. Ceci se

résume simplement à l'estimation du pas de quantification q_k (ou largeur de la classe P^m pour $m \in \mathbb{Z}$) à utiliser pour chaque quantificateur associé à chaque sous-bande k.

Dans la suite des calculs, nous allons supposer que la zone morte z est égale au pas de quantification q et donc que les quantificateurs utilisés sont simplement uniformes avec décodage par le centroïde ; une étude détaillée incluant la prise en compte de la zone morte est donnée dans [299], [79]. Moyennant cette hypothèse et en utilisant les expressions (3.7) et (3.13) de D et de R en fonction de q, l'écriture de la fonctionnelle (3.15) dans le cadre de l'analyse multirésolution devient[1] :

$$J_\lambda\left(\{q_k\}\right) = \sum_{k=1}^{N} a_k R\left(\alpha_k, \frac{q_k}{\sigma_k}\right) + \lambda\left[\sum_{k=1}^{N} \Delta_k \pi_k \sigma_k^2 D\left(\alpha_k, \frac{q_k}{\sigma_k}\right) - D_{cible}\right], \quad (3.16)$$

où a_k correspond au poids de la sous-bande k dans le débit total (rapport entre la taille de la sous-bande k et la taille totale du signal source), $\{\pi_k\}$ est l'ensemble des coefficients correcteurs qui prennent en compte la non-orthogonalité des filtres [325], [96], [91] (ils sont obtenus à partir des relations établies au Chapitre 2 dans le cas de M canaux) et $\{\Delta_k\}$ est l'ensemble des pondérations qui permettent d'introduire une mesure de qualité autre que l'EQM, par exemple pour prendre en compte des aspects psychovisuels. Enfin, D_{cible} est la qualité cible que l'on cherche à atteindre.

La fonctionnelle J_λ est de classe C^∞ pour $q_k > 0, \forall k \in [1, ..., N]$ et son expression par rapport à l'ensemble $\{q_k\}$ est extrêmement complexe. De ce fait, l'étude analytique en vue de prouver l'existence et l'unicité d'un minimum reste un problème ouvert. Cependant, pour résoudre le problème, nous allons supposer l'existence et l'unicité de la solution de minimisation et nous proposons de dériver J_λ par rapport à ses inconnues pour chercher son point stationnaire.

Posons $\widetilde{q}_k = \frac{q_k}{\sigma_k}$. La dérivée du critère (3.16) par rapport à $\widetilde{q}_k$ et λ est alors donnée par le système de $card\,(E) + 1$ équations à $card\,(E) + 1$ inconnues suivant, avec E l'ensemble des sous-bandes tel que $E = \left\{i \in [1, ..., N]\,/\sigma_i^2 \neq 0\right\} \cap \left\{j \in [1, ..., N]\,/\Delta_j \neq 0\right\}$:

$$\begin{cases} \frac{\partial J_\lambda(\{\widetilde{q}_k\})}{\partial \widetilde{q}_k} = a_k \frac{\partial R(\alpha_k, \widetilde{q}_k)}{\partial \widetilde{q}_k} + \lambda \Delta_k \pi_k \sigma_k^2 \frac{\partial D(\alpha_k, \widetilde{q}_k)}{\partial \widetilde{q}_k} = 0 & \forall k \in E \\ \frac{\partial J_\lambda(\{\widetilde{q}_k\})}{\partial \lambda} = \sum_{k=1}^{N} \Delta_k \pi_k \sigma_k^2 D\left(\alpha_k, \widetilde{q}_k\right) - D_{cible} = 0. \end{cases} \quad (3.17)$$

En supposant que $\frac{\partial R(\alpha_k, \widetilde{q}_k)}{\partial \widetilde{q}_k} \neq 0$ pour tout $k \in [1, ..., N]$, alors le système à

[1] Par soucis de simplification dans l'écriture, nous remplaçons $R\left(\alpha, \frac{q}{\sigma}, \frac{q}{\sigma}\right)$ par $R\left(\alpha, \frac{q}{\sigma}\right)$ et $D\left(\alpha, \frac{q}{\sigma}, \frac{q}{\sigma}\right)$ par $D\left(\alpha, \frac{q}{\sigma}\right)$.

résoudre devient :

$$\begin{cases} h_\alpha\left(\widetilde{q}_k\right) = \frac{\frac{\partial D\left(\alpha_k,\widetilde{q}_k\right)}{\partial \widetilde{q}_k}}{\frac{\partial R\left(\alpha_k,\widetilde{q}_k\right)}{\partial \widetilde{q}_k}} = -\frac{a_k}{\lambda \Delta_k \pi_k \sigma_k^2} & \forall k \in E \\ \sum_{k=1}^{N} \Delta_k \pi_k \sigma_k^2 D\left(\alpha_k, \widetilde{q}_k\right) = D_{cible}. \end{cases} \tag{3.18}$$

L'expression de $h_\alpha\left(\widetilde{q}\right)$ est une fonction des $f_{n,m}$. Elle est donnée par la relation :

$$h_\alpha\left(\widetilde{q}\right) = \frac{\sum_{m=1}^{+\infty} \frac{2\frac{df_{1,m}}{d\widetilde{q}}(\alpha,\widetilde{q},\widetilde{q}) f_{1,m}(\alpha,\widetilde{q},\widetilde{q}) f_{0,m}(\alpha,\widetilde{q},\widetilde{q}) - f_{1,m}(\alpha,\widetilde{q},\widetilde{q})^2 \frac{df_{0,m}}{d\widetilde{q}}(\alpha,\widetilde{q},\widetilde{q})}{f_{0,m}(\alpha,\widetilde{q},\widetilde{q})^2}}{\frac{p_{1,\alpha}(\widetilde{q}/2)}{2}\left[\ln f_{0,0}\left(\alpha,\widetilde{q}\right)+1\right] + \sum_{m=1}^{+\infty} \frac{df_{0,m}}{d\widetilde{q}}(\alpha,\widetilde{q},\widetilde{q})\left[\ln f_{0,m}\left(\alpha,\widetilde{q},\widetilde{q}\right)+1\right]} \ln 2, \tag{3.19}$$

avec

$$\frac{df_{n,m}}{d\widetilde{q}}(\alpha,\widetilde{q},\widetilde{q}) = \left[\left(m+\frac{1}{2}\right)^{n+1} p_{1,\alpha}\left(m\widetilde{q}+\frac{\widetilde{q}}{2}\right) - \left(m-\frac{1}{2}\right)^{n+1} p_{1,\alpha}\left(m\widetilde{q}-\frac{\widetilde{q}}{2}\right)\right]\widetilde{q}^n \tag{3.20}$$

La fonction $h_\alpha\left(\widetilde{q}\right)$ est représentée sur la figure 3.4. De plus, connaissant l'évolution de D en fonction de $\widetilde{q}$ (cf. formule (3.8)), il est possible d'avoir un lien direct entre D et h_α, c'est-à-dire entre D et λ en introduisant la courbe paramétrique (de paramètre $\widetilde{q}$) suivante (voir figure 3.5) :

$$\left[\ln D\left(\alpha,\widetilde{q}\right);\ln\left(-h_\alpha\left(\widetilde{q}\right)\right)\right]. \tag{3.21}$$

De façon évidente, on peut remarquer que cette relation paramétrique est équivalente à la relation

$$\left[\ln D\left(\alpha,\widetilde{q}\right);\ln\frac{a_k}{\lambda\Delta_k\pi_k\sigma_k^2}\right] \tag{3.22}$$

ce qui nous permet d'avoir un lien direct entre l'EQM et le paramètre de Lagrange λ. Ce lien est très utile pour la résolution de notre problème d'optimisation, comme nous le verrons au paragraphe 3.2.2.3.

3.2.2.3 Algorithme proposé

Nous avons introduit cet algorithme d'allocation dynamique des ressources binaires dans le cadre du codage par transformée en ondelettes "au fil de l'eau". Il permet soit un contrôle du débit binaire soit un contrôle de la "qualité" image en terme d'EQM ou de rapport signal-à-bruit. Ici, nous ne présentons que la version qui permet l'allocation de "qualité", le lecteur intéressé par l'allocation de débits pourra se référer à [299].

La solution du problème d'allocation est donnée par la résolution du système d'équations non linéaires (3.18). Une façon classique pour résoudre un tel

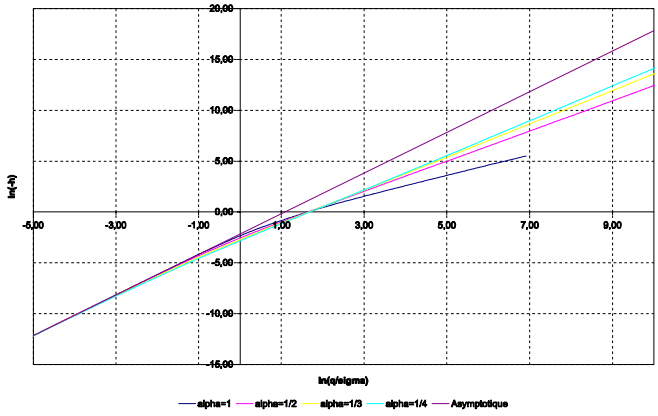


FIG. 3.4 – Evolution de h_α en fonction de $\ln \widetilde{q}$ pour différentes valeurs du paramètre α de la distribution Gaussienne généralisée.

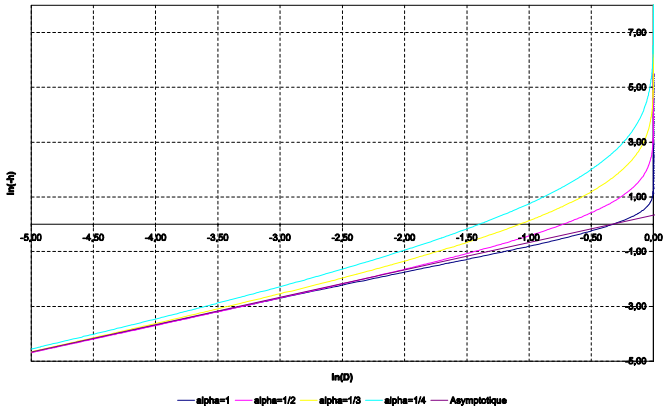


FIG. 3.5 – Evolution de $\ln(-h_\alpha)$ en fonction de $\ln D$ pour différentes valeurs du paramètre α de la distribution Gaussienne généralisée.

système serait d'utiliser une méthode de type Newton-Raphson. Cependant, compte-tenu de la proposition 2 et du fait que les expressions de D et R sont indépendantes de la variance σ^2 pour un signal de distribution Gaussienne généralisées, il est possible de pré-calculer ces fonctions de façon à limiter le coût calcul lors de l'allocation. Les fonctions pré-calculées sont paramétrées par α et peuvent être stockées une fois pour toute puisqu'elle sont indépendantes de σ^2. Nous avons donc développé une solution de ce type qui nous à permis de définir l'algorithme suivant de faible complexité et permettant une allocation rapide connu sous le nom EBWIC (pour "**E**fficient **B**it allocation **W**avelet **I**mage **C**oder").

ALGORITHME

1. *Pour toute sous-bande k n'appartenant pas à l'ensemble E, faire $q_k \leftarrow +\infty$ de façon à obtenir un débit nul pour la sous-bande k. Pour toutes les autres sous-bandes, continuer à l'étape 2.*
2. *Choisir une valeur initiale pour le paramètre λ.*
3. *Pour chaque sous-bande k de l'ensemble E, calculer $\ln\left(\frac{a_k}{\lambda\Delta_k\pi_k\sigma_k^2}\right)$ et en déduire la distorsion normalisée correspondante D_k à partir de la relation paramétrique 3.22.*
4. *Calculer la quantité $\left|\sum_{k\in E}\Delta_k\pi_k\sigma_k^2 D_k - D_{cible}\right|$.*
   (a) **Si** *celle-ci est inférieure à un seuil donné,* **alors** *la contrainte est vérifiée et le λ courant est alors optimal.*
   (b) **Sinon**, *calculer un nouveau λ et retourner à l'étape 3.*
5. *Pour chaque sous-bande k de l'ensemble E, utiliser la valeur $\ln\left(\frac{a_k}{\lambda\Delta_k\pi_k\sigma_k^2}\right) = \ln(-h_{\alpha_k})$ calculée au cours de l'étape 3 pour en déduire $\widetilde{q}_k$ à partir de la relation (3.19)* [2]*. q_k est alors le pas de quantification optimal pour la sous-bande k.*

Lorsque la compression est réalisée au fil de l'eau, une bonne estimation initiale de λ est la valeur de λ optimale trouvée lors de l'allocation du groupe de coefficients précédents. Dans les autres cas, nous proposons de prendre la valeur de λ optimale théorique sous hypothèse asymptotique [299], c'est-à-dire :

$$\lambda = \frac{\sum_{k\in E} a_k}{2D_{cible}\ln 2} \tag{3.23}$$

[2] L'inversion de la fonction $h_\alpha(\widetilde{q})$ est une opération difficile. La solution que nous avons proposé est de pré-calculer et de tabuler cette fonction pour différents paramètres α et d'effectuer l'inversion par simple lecture de tables.

3.2.2.4 Complexité de l'algorithme

Une étude précise nous a permis de montrer que la complexité de la méthode d'allocation proposées était de l'ordre de 10^{-2} opérations arithmétiques par pixel pour une image monochrome de taille 512×512 pixels, découpée en 19 sous-bandes fréquentielles. Le coût principal des méthodes d'allocation basées modèles proposées réside dans l'estimation des paramètres des distributions Gaussiennes généralisées de chaque sous-bande. Les paramètres σ et α sont calculés à partir des moments d'ordre deux et quatre des coefficients dans chaque sous-bande. Si ces moments sont calculés simultanément, on peut voir que leurs estimations peuvent être réalisées en 4 opérations par coefficient d'ondelette (2 additions et 2 multiplications).

Nous pouvons donc approximer la complexité de la méthode d'allocation de "qualité" (ou de débit) à 4 opérations arithmétiques simples par pixel de l'image [299].

Dans le cadre d'une étude menée contractuellement avec le CNES, nous avons montré qu'il était possible d'adapter la version précédente de notre algorithme d'allocation de débit pour l'intégrer sur un DSP 21020 en moins de 400 cycles sans perte de performance pour le codage en temps réel des images de la future génération des satellites d'observation de la terre PLEIADE. Cette version de la méthode d'allocation de débit fait d'ailleurs actuellement l'objet d'un dépôt de brevet conjoint CNRS-CNES [23].

3.2.3 Le basé modèles et JPEG-2000

Les résultats présentés sur les figures 3.6, 3.7, 3.8 et 3.9 comparent les performances de notre algorithme d'allocation EBWIC avec ceux obtenus par JPEG-2000 pour des images de test (LENA et GOLD) issues de la base de données JPEG-2000. Nous avons testé les allocations de "qualité" et de débit en les intégrant dans un algorithme de compression par ondelettes d'images fixes. La transformée en ondelettes sélectionnée est une décomposition dyadique standard sur trois niveaux, utilisant le banc de filtres biorthogonal "9-7" [2]. Le codage entropique des coefficients d'ondelettes quantifiés est réalisé par le codeur arithmétique contextuel par plans de bits de la norme JPEG-2000 [282], [83]. Les résultats numériques que nous proposons permettent de mesurer l'efficacité de notre algorithme d'optimisation des quantificateurs scalaires avec zone morte et décodage par le centroïde par rapport à la procédure de recherche exhaustive des quantificateurs de JPEG-2000 [283]. Les résultats présentés ont été obtenus en générant un train binaire qui a été ensuite décodé. Les problèmes d'allocation de débit et de qualité sont duaux. Aussi, quand le codage n'est pas réalisé au fil de l'eau, les performances des deux types d'allocation sont identiques en terme de couples débit-distorsion. Les résultats qui suivent

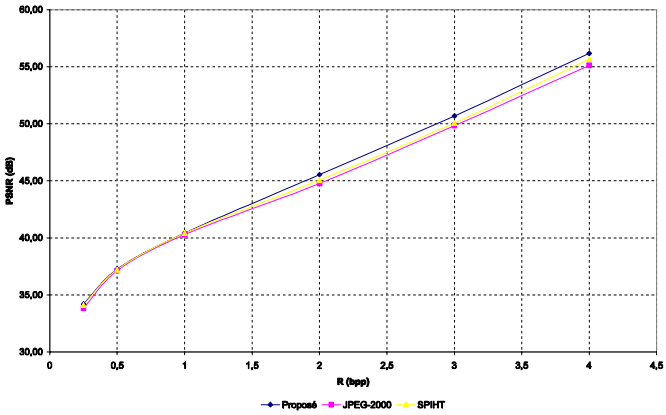


FIG. 3.6 – Rapport Signal-à-Bruit pic en fonction du débit pour l'image LENA (8 bpp−512 × 512 pixels). 3 niveaux de décomposition pour la transformée en ondelettes.

montrent donc les performances d'EBWIC comparées à celles de JPEG-2000 [283] et de SPIHT [307].

Les performances de compression de JPEG-2000 [282] sont proches de celles fournies par EBWIC, et toutes les deux dépassent celles de JPEG en particulier pour les forts taux de compression. Cependant JPEG-2000 présente une complexité algorithmique environ cinq fois plus grande que celle de JPEG. Cette complexité est un frein à son adoption dans le cas des systèmes embarqués et/ou mobiles. En effet, JPEG-2000 est basé sur la mesure simultanée du débit réel et de la distorsion réelle de quantification des coefficients issus de la transformée en ondelettes. De ce fait, pour effectuer l'allocation des débits, il effectue des codages et décodages successifs au moyen d'un codeur par plans de bits [322], [323]. C'est dans le soucis d'optimisation de la complexité algorithmique de la norme JPEG-2000 que nous participons au projet RNRT EIRE (http ://www.telecom.gouv.fr/rnrt). L'objectif du projet EIRE est de proposer des améliorations au schéma de base de JPEG-2000, de réduire la complexité algorithmique, d'améliorer la qualité tant à l'encodage qu'au décodage, et d'assurer une utilisation optimale de JPEG-2000, y compris dans un contexte d'interopérabilité avec d'autres schémas de codage. Dans ce contexte, nous travaillons sur l'intégration de la méthode d'allocation EBWIC basée sur des modèles non aymptotiques dans un codeur JPEG-2000 (partie I du standard) afin de réduire sa complexité [101]. Ces travaux permettront de fournir la meilleure qualité de service dans un contexte limité en ressources (bande passante, capacité de traitement...), et de démontrer la possibilité d'implémenter

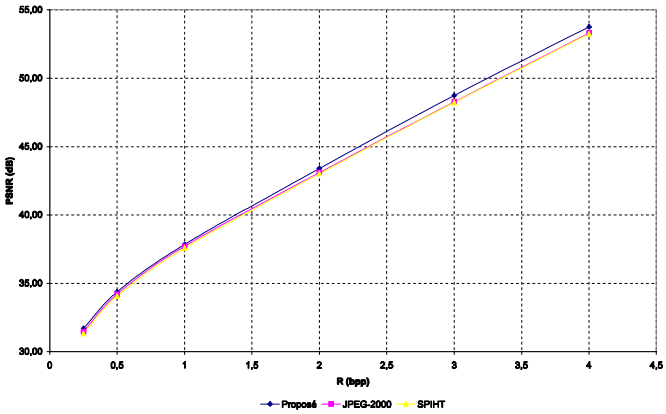


FIG. 3.7 – Rapport Signal-à-Bruit pic en fonction du débit pour l'image GOLD (8 bpp−720 × 576 pixels). 3 niveaux de décomposition pour la transformée en ondelettes.

efficacement JPEG-2000. Les applications visées concerneront principalement l'imagerie de type scientifique (satellite et médical) et la télésurveilance (dans un contexte de transmission vidéo sur IP).

3.2.4 La solution satellitaire

Les images acquises par des systèmes embarqués à bord de satellites d'observation de la Terre ou encore de sondes spatiales représentent dans la plupart des cas de très gros volumes de données (de l'ordre de 24 000 pixels par ligne et de plusieurs centaines de lignes dans une fauchée). Ces images sont transmises sur Terre ou encore stockées à bord durant les périodes de "non visibilité". Cependant, du aux limitations intrinsèques du satellite en terme de masse, puissance de consommation ou encore de coûts, il est nécessaire de limiter au maximum la capacité de stockage à bord (mémoire) et le débit de transmission nécessaire pour accomplir la mission. Ainsi, la mémoire de masse présente à bord de celui-ci sont telles qu'il n'est pas envisageable de stocker entièrement une image acquise. Les premiers algorithmes utilisés dans l'espace étaient basés sur le DPCM puis la DCT. Une évaluation de la qualité image a montré cependant que pour des applications utilisant la DCT adaptative, telle que celle utilisée par SPOT 5 [289], le taux de compression maximum acceptable était de l'ordre de 3 pour des missions d'observation à haute résolution et de l'ordre de 15 pour des missions scientifiques. Au dela de ces taux de compression, l'apparition des effects de blocs dans les zones uniformes et la perte de détails liés au bruit de compression ne sont plus acceptables pour un usage scientifique. Il était donc

FIG. 3.8 – Comparaison visuelle de différentes méthodes de compression pour un débit de 0,25 bpp. (a) Extrait de l'image LENA originale ; (b) EBWIC (PSNR=34,19 dB) ; (c) SPIHT (PSNR=34,11 dB) ; (d) JPEG-2000 (PSNR=33,81 dB). Les images ont subi un renforcement des contours sous le logiciel PHOTOSHOP.

FIG. 3.9 – Comparaison visuelle de différentes méthodes de compression pour un débit de 0,25 bpp. (a) Extrait de l'image GOLD originale ; (b) EBWIC (PSNR=31,70 dB) ; (c) SPIHT (PSNR=31,33 dB) ; (d) JPEG-2000 (PSNR=31,47 dB). Les images ont subi un renforcement des contours sous le logiciel PHOTOSHOP.

nécessaire de dépasser ces limites afin de préparer les futures générations de satellites d'observation de la Terre et de missions scientifiques [288], [100], [72]

Dans le cadre d'une application satellitaire suscitée par le Centre National d'Etudes Spatiales (CNES) de Toulouse, nous avons été amenés à développer un algorithme de compression d'images adapté à un traitement bord de faible complexité et coût mémoire. Ces travaux ont donné lieu en 1995 à un algorithme de transformée en ondelettes 2D au "fil de l'eau" [140] (cf. chapitre 2). Par la suite, l'algorithme que nous avons développé à l'occasion de divers contrats avec le CNES [119], [124], [147], [151], [153], [154], [155], permet d'effectuer le traitement des données (compression) au fur et à mesure que le satellite acquiert des lignes images et ce pour des coûts mémoire et calcul relativement faibles. La méthode est détaillée dans la thèse de Parisot [299] et pour le cas des vidéos dans [78] et dans l'article "3D Scan Based Wavelet Transform and Quality Control for Video Coding" [12] joint en annexe B, mais le principe de la méthode reste identique dans le cas des images 2D.

La méthode d'allocation de débits a été adaptée aux contraintes bord liées à la compression d'images satellites et nous avons développé une méthode d'allocation de débits dite "dynamique". Cet algorithme a aussi été validé dans le cas du contrôle de la "qualité" [85]. La chaîne de décomposition multirésolution satellitaire développée pour le CNES consiste tout d'abord à acquérir des blocs de lignes image "au fil de l'eau" et à les décomposer successivement d'une part en une ou plusieurs sous-bandes spectrales, d'autre part en une sous-bande basse fréquence destinée à être redécomposée à son tour jusqu'à atteindre le nombre de sous-bandes désirées (nominalement 10 sous-bandes en mode dyadique et 6 en mode quinconce). Chacune de ces sous-bandes est ensuite quantifiée scalairement, puis codée par un codeur entropique adapté, et enfin encapsulée dans un format binaire destiné au transfert par télémesure. Pour obtenir de façon optimale le taux de compression souhaité, les débits de chaque sous-bande sont alloués dynamiquement. L'allocation dynamique est effectuée par un algorithme de prédiction prenant en compte l'énergie contenue dans la sous-bande ainsi que la modélisation de la distribution des coefficients d'ondelettes par une gaussienne généralisée. Les débordements éventuels sont alors régulés par une boucle d'asservissement afin de garantir le débit de consigne. De plus, nous avons proposé une optimisation de cet algorithme d'allocation dynamique afin de permettre son implantation en "temps réel" sur un DSP 21020 dans le cadre de l'application satellitaire d'observation de la Terre PLEIADE. Ces travaux font l'objet d'un dépôt de brevet conjointement par le CNRS et le CNES [23], et l'algorithme qui résulte de nos travaux a été retenu pour être embarqué dans la future génération de satellites PLEIADE.

Des résultats expérimentaux comparatifs entre EBWIC "au fil de l'eau" et JPEG-2000 tuilé sont donnés sur les figures 3.10 et 3.11 pour l'image satellitaire de NIMES de taille 512×512 pixels codée sur 8 bits par pixel. Cette image nous a été fournie par le CNES Toulouse.

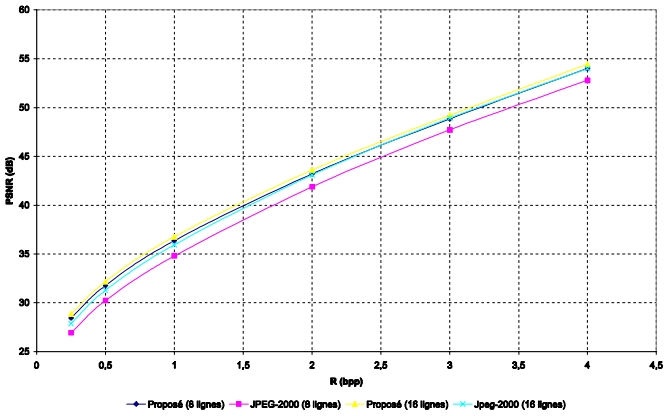


FIG. 3.10 – Rapport Signal-à-Bruit pic en fonction du débit pour l'image NIMES (8 bpp−512 × 512 pixels). 3 niveaux de décomposition pour la transformée en ondelettes. Compression au "fil de l'eau" avec la méthode proposée EBWIC sur des zones de traitement de tailles 8 lignes et 16 lignes image. Comparaison avec JPEG-2000 en mode tuilé avec des tuiles de tailles 8 et 16 lignes image.

3.3 Prise en compte de l'hétérogénéité des canaux

Ces travaux on été réalisés durant la thèse de Manuela Pereira (2004) [301] que j'ai co-encadré en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Une partie de nos travaux va paraître dans la revue EURASIP Journal on Applied Signal Processing : "Multiple description image and video coding for wireless channels" en 2003 [15].

3.3.1 La transmission sur canaux bruités

3.3.1.1 Problème des canaux bruités

La communication mobile et multimédia a connu un vif succès ces dernières années, faisant des canaux radio-mobiles et Internet des médias de transport privilégiés. Cependant, le canal radio-mobile présente une faible bande passante et des caractéristiques variant dans le temps, ce qui rend la conception de système de compression/décompression très difficile. De plus, les données transmises sont corrompue par le bruit du canal qui peut apparaître sous la forme de pertes aléatoires de bits, de pertes en rafales ou encore de pertes de paquets de bits dans le cas de l'Internet. Chaque type de perte engendre évidemment une perte de qualité de l'image ou de la vidéo restituée au décodeur

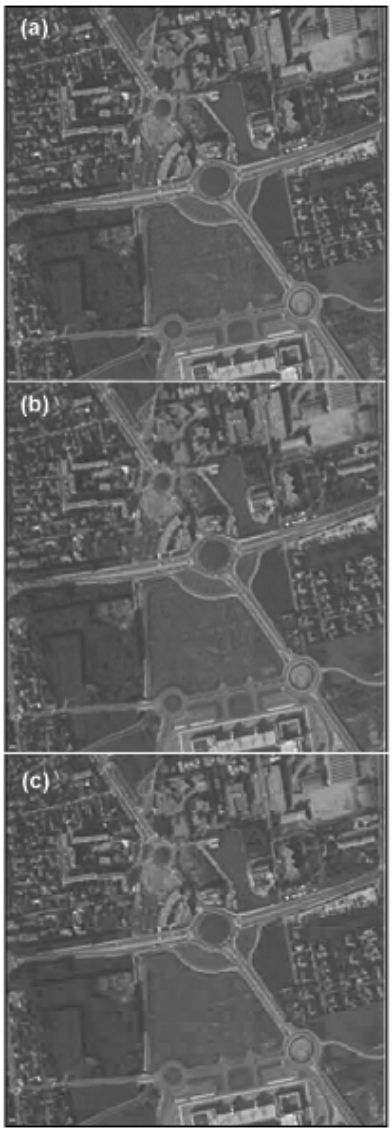


FIG. 3.11 – Comparaison visuelle de différentes méthodes de compression "au fil de l'eau" pour un débit de 1 bpp. (a) Extrait de l'image NIMES originale ; (b) EBWIC en mode régulé avec des blocs de lignes de taille 8 lignes image (PSNR=36,37 dB) ; (c) JPEG-2000 en mode tuilé avec des tuiles de tailles 8 lignes image (PSNR=34,80 dB). Les images ont subi un renforcement des contours sous le logiciel PHOTOSHOP.

et cette perte de qualité est d'autant plus importante que le taux de compression est élevé (bas débits de transmission). Il est donc nécessaire d'envisager des méthodes de codage robuste au bruit canal pour palier à ces problèmes de transmission.

Une méthode classique pour lutter contre les erreurs de transmission est l'utilisation de codes correcteurs d'erreurs ou "Foward Error Correction" (FEC) en Anglais. L'objectif ici n'est pas de faire un état de l'art sur les techniques de codage canal. Citons simplement, parmi les nombreux travaux qui existent sur la transmission robuste des images, les travaux récents suivants [319], [304], [263], [264], [315], [273], [316]. De façon générale, les FEC nécessitent l'ajout de données redondantes au signal comprimé qui permettent au décodeur de corriger des erreurs jusqu'à un certain niveau de bruit canal. Ce rajout de redondance augmente de manière évidente le nombre de bits dans le train binaire comprimé et diminue donc le taux de compression. D'autre part, un code correcteur doit être construit de façon à prendre en compte le pire des cas supposé, c'est-à-dire le bruit maximal supposé du canal, réduisant ainsi l'efficacité de la méthode lorsque le canal est hautement variable.

Historiquement la théorie de l'information (et notamment Shannon) a établi que, sous certaines hypothèses, l'optimisation des schémas de codage source/canal combinés était atteinte en "séparant" le codage source et le codage canal et cette approche a été jusqu'aujourd'hui appliquée aux systèmes de transmission sur canaux bruités. Cependant les systèmes sous contraintes de faible complexité et de temps réel utilisés pour les besoins de vidéoconférence, vidéo sur Internet, etc., ne vérifient pas ces hypothèses de séparation. La structure actuelle des systèmes de communication est telle que les fonctions de codage source et de codage canal ne devraient plus être conçue séparemment mais globalement, en tenant compte des caractéristiques précises de la source et du canal. On parle alors de codage conjoint source/canal dont les avantages sont déjà reconnus.

Un système de codage pour les communications radio-mobiles ou sur Internet devrait à la fois optimiser le codage de la source et le codage canal, devrait être adaptatif et robuste au bruit de transmission et enfin utiliser de façon "optimale" les ressources limitées du canal. Dans cette optique, une méthode de codage source/canal conjointe connue sous le nom de codage par descriptions multiples ("Multiple Description Coding" que nous appellerons ici MDC) a déjà fait ses preuves. En effet, ce type de méthode est robuste au bruit du canal au prix d'un faible surcoût de transmission. Le codage par descriptions multiples suppose l'existence de plusieurs canaux de transmission entre l'émetteur et le récepteur et que les canaux peuvent présenter de longues périodes de pertes de données. De plus, les erreurs qui se présentent sur les différents canaux sont supposées indépendantes, de sorte que la probabilité que tous les canaux subissent des pertes simultanément reste faible. Dans le cas MDC, les canaux doivent en général être physiquement distincts comme par exemple c'est le cas pour les canaux sans fils multi-trajets ou les réseaux de transmission par paquets tels

qu'Internet. Toutefois, même s'il n'existe qu'un seul trajet physique entre la source et le destinataire, il peut être divisé en plusieurs sous-canaux virtuels en utilisant par exemple des méthodes de multiplexage temporel et rendre ainsi possible l'utilisation d'un codeur MDC [284], [257], [338], [286].

Le grand nombre de travaux relatifs au MDC pour des canaux de type Internet comparativement au petit nombre d'articles de référence dédiés au MDC pour les canaux radio-mobiles s'explique par le fait que les performances du MDC dans le cas où le canal est faiblement bruité sont très inférieures à celle obtenues au moyen d'un codage utilisant une seule description ("Single Description Coding" ou SDC). En effet, il n'est pas simple de contrôler et d'adapter au bruit canal le niveau de redondance d'un codeur MDC. Nous nous sommes intéressé à ce problème et notre contribution a été de concevoir un système de codage par descriptions multiples combiné à la transformée en ondelettes et dont le niveau de redondance entre les descripteurs s'adapte automatiquement en fonction du type de canal et du niveau de bruit.

3.3.1.2 Les systèmes à descriptions multiples

L'idée principale des descriptions multiples est la suivante. Supposons que l'on désire envoyer une description d'un processus stochastique source sur un canal de transmission bruité. Comme le canal est bruité il y a une chance de perdre la description ! Cependant, si au lieu de transmettre une seule description on en transmet deux, chacune possédant suffisamment d'information descriptive pour être à elle seule représentative de la source, alors on augmente les chances qu'une des deux arrivera à destination. De plus, si toutes les deux descriptions arrivent à destination, alors on veut que l'information descriptive conjointe issue des deux descriptions soit la meilleure possible. Ce problème a été introduit en 1979 par Gersho, Witsenhausen, Wolf, Wyner, Ziv et Ozarow [272]. Des travaux de base qui traitent de ce problème peuvent être trouvés dans les articles de Witsenhausen [336], Wolf, Wyner et Ziv [337], Ozarow [297], Witsenhausen et Wyner [335] et, Gamal et Cover [272]. En fait, la difficulté pour concevoir un tel système est que chaque description individuelle doit être "proche" du processus source et nécessairement dépendante de l'autre description. Ce problème peut être généralisé au cas de N descriptions avec $N \geq 2$

Dans le problème de codage par descriptions multiples réduit au cas simple de deux descriptions, une source est décrite par deux descripteurs avec respectivement les débits R_1 et R_2. Ces deux descriptions prises individuellement permettent respectivement un décodage avec une distorsion D_1 et D_2, que l'on appelle distorsions latérales. Les deux descriptions ensembles fournissent une distorsion D_0, ou distorsion centrale, telle que $D_0 \leq D_1$ et $D_0 \leq D_2$. Il existe trois types d'approches disctinctes de MDC pour les sytèmes de codage source/canal :

- MDC au niveau du quantificateur. Historiquement, Vaishampayan est le premier à avoir proposé des systèmes de codage source/canal basés sur le MDC pour la transmission d'images. La technique de MDC qu'il proposa est conçue pour fournir deux descriptions au moyen de quantificateurs scalaires ou encore vectoriels en ajoutant de la redondance au niveau des index des vecteurs de quantification [326], [327], [328], [309], [310], [331] ;
- MDC au niveau des coefficients de la transformation. Cette approche a été initiée par Wang, Orchard et Reibman [333]. Des codeurs MDC codent les $N \times N$ coefficients d'une transformation par blocs qui a été conçue de façon à introduire un certain degré de corrélation contrôlable entre les coefficients transformés [277], [276]. La redondance est placée ici sur les coefficients de la transformée eux-mêmes ;
- MDC en utilisant une transformation "sur-complète" ("overcomplete"). Cette approche est due à Goyal, Kovacevic et Vetterli [277], [278]. Des codeurs MDC sont construits de façon séparée pour coder les $N \times K$ ($K \geq N$) coefficients d'une transformée redondante sur-complète.

De nombreux travaux ont été menés par la suite sur le codage source/canal d'images fixes par MDC. Notons principalement les travaux de Jiang et Ortega [281], Rogers et Cosman [305], Mohr, Riskin et Ladner [295]. Il existe aussi dans la littérature quelques travaux concernant le codage de vidéos par MDC. Notons les travaux de Vaishampayan [329], [330], les travaux de Apostopoulos et Wee [255], et ceux de Reibman, Jafarkhani, Wang, Orchard et Puri [302]. Un état de l'art complet sur les techniques de codage source/canal par MDC se trouve dans la thèse de Pereira [301].

3.3.2 Approche proposée

3.3.2.1 Position du problème

Le système MDC que nous avons proposé a été développé pour le cas de deux descriptions et nous avons fait l'hypothèse de canaux transmission de capacités égales. Dans un tel schéma, le codeur doit produire deux trains binaires d'importances égales qui sont transmis vers trois décodeurs sur un canal bruité. Le décodeur *central* reçoit l'information envoyée sur les deux canaux alors que les décodeurs *latéraux* reçoivent chacun l'information envoyée sur le canal qui leur a été associé. La quantité de redondance est distribuée entre les deux descriptions en prenant en considération un modèle du canal ainsi que son état (ou taux d'erreur bit que nous appellerons BER). Un schéma de principe est donné sur la figure 3.12. Pour un débit latéral R_l donné et une distorsion latérale D_l donnée, la construction des deux descriptions est soumise à trois conditions :

1. **Condition 1** : le décodeur central doit reconstruire le signal source original avec une distorsion centrale D_0 à partir des deux descriptions.

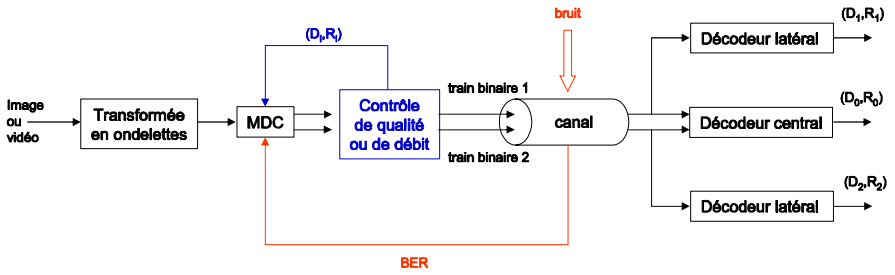


FIG. 3.12 – Schéma de principe général du codeur/décodeur MDC proposé dans le cas de deux descriptions.

2. **Condition 2** : le codeur MDC est équilibré. Il doit générer deux descriptions avec chacune un débit R_l tel que $\mathbf{R}_1 = \mathbf{R}_2 = R_l$.
3. **Condition 3** : chaque descripteur latéral doit permettre la reconstruction du signal source avec une distorsion latérale $\mathbf{D}_1 \leq D_l$ et $\mathbf{D}_2 \leq D_l$.

Le problème posé est de construire un codeur qui minimise la distorsion centrale D_0 lorsque les conditions 2 et 3 sont vérifiées. Ainsi, pour un système basé sur une décomposition en ondelettes sur N sous-bandes, minimiser D_0 en vérifiant la condition 2 se ramène à chercher les ensembles de débits $\{R_{k,1}\}$ et $\{R_{k,2}\}$ qui minimisent la distorsion centrale D_0, avec $R_{k,j}$ le débit latéral dans la sous-bande $k \in \{1, ..., N\}$ pour le descripteur $j \in \{1, 2\}$. C'est donc un problème d'allocation de débits que nous avons à résoudre et qui est posé de la façon suivante :

$$(P) \begin{cases} \min\limits_{\{R_{k,1}\},\{R_{k,2}\}} D_0\left(\{R_{k,1}\},\{R_{k,2}\}\right) \\ \text{sous les contraintes } \mathbf{R}_1 = \sum_{k=1}^{N} a_k R_{k,1} \leq R_l \text{ et } \mathbf{R}_2 = \sum_{k=1}^{N} a_k R_{k,2} \leq R_l \end{cases} \tag{3.24}$$

avec a_k le poids de la sous-bande k (cf. paragraphe 3.2.2.2). Comme nous l'avons proposé au paragraphe 3.2.2, ce problème peut se résoudre par une approche Lagrangienne et le problème (P), en terme de pas de quantification q, peut se mettre sous la forme de la fonctionnelle (3.25) suivante :

$$J_\lambda\left(\{q_{k,1}\},\{q_{k,2}\}\right) = D_0 + \sum_{j=1}^{2} \lambda_j \left(\mathbf{R}_j \leq R_l\right) \tag{3.25}$$

D'autre part, vérifier la condition 3 nous amène à formaliser le problème de la minimisation de cette fonctionnelle par le problème (P') :

$$(P') \begin{cases} \min\limits_{\{q_{k,1}\},\{q_{k,2}\}} J_\lambda\left(\{q_{k,1}\},\{q_{k,2}\}\right) \\ \text{avec les pénalités } \mathbf{D}_1 \leq D_l \text{ et } \mathbf{D}_2 \leq D_l \ . \end{cases} \tag{3.26}$$

Ainsi, nous avons proposé de ramener le problème d'allocation des débits dans un cadre de descriptions multiples, à la minimisation en q de la fonctionnelle suivante :

$$J_\lambda\left(\{q_{k,1}\},\{q_{k,2}\}\right) = D_0 + \sum_{j=1}^{2} \lambda_j \left(\mathbf{R}_j \leq R_l\right) + \sum_{j=1}^{2} \mu_j(\mathbf{D}_j \leq D_l). \tag{3.27}$$

Pour une source de distribution Gaussienne généralisée, la distorsion centrale D_0 peut s'écrire sous la forme suivante [70], [86] :

$$D_0 = \sum_{k=1}^{N} \Delta_k \pi_k \sigma_{k,0}^2 D_{k,0}\left(\frac{q_{k,1}}{\sigma_{k,1}}, \frac{q_{k,2}}{\sigma_{k,2}}\right), \tag{3.28}$$

où $\sigma_{k,0}^2 D_{k,0}$ est la distorsion centrale dans la sous-bande k. Les distorsions latérales sont données par :

$$\mathbf{D}_j = \sum_{k=1}^{N} \Delta_k \pi_k \sigma_{k,j}^2 D_{k,j}\left(\frac{q_{k,j}}{\sigma_{k,j}}\right) \text{ pour } j \in \{1,2\}, \tag{3.29}$$

où $\sigma_{k,j}^2 D_{k,j}$ est la distorsion dans la sous-bande k pour le descripteur j. Comme précédemment, on désigne par $\{\Delta_k\}$ l'ensemble des pondérations qui permettent d'introduire une mesure de qualité autre que l'EQM (cf. paragraphe 3.2.2.2) et par $\{\pi_k\}$ l'ensemble des coefficients correcteurs pour la non orthogonalité des filtres (cf. paragraphe 3.2.2.2 et chapitre 2).

3.3.2.2 Modélisation de la distorsion centrale

La distorsion proposée La distorsion centrale est la distorsion de l'image décodée en utilisant les deux descriptions. Lorsque le décodeur reçoit les deux descriptions non corrompues par le bruit canal, chaque sous-bande de coefficients d'ondelettes apparait deux fois avec deux débits différents (pas de quantification différents). Le décodeur central peut alors choisir la sous-bande qui possède le plus grand débit (ou plus petit pas de quantification). L'autre sous-bande (ou sous-bande redondante) servira uniquement pour le décodeur latéral. Dans ce cas, la distorsion centrale s'exprime par la relation

$$\sum_{k=1}^{N} \min(D_{k,1}, D_{k,2}), \tag{3.30}$$

Dans le cas ou le canal est bruité, le décodeur central va prendre en compte les deux descriptions pour la reconstruction. Ainsi, dans le cas d'un canal fortement bruité (fort BER) et comme la redondance est une fonction croissante du BER, les deux sous-bandes sont considérées avec la même importance. Nous définissons donc la distorsion centrale par :

$$\sum_{k=1}^{N} \left(\min(D_{k,1}, D_{k,2}) + \max(D_{k,1}, D_{k,2})\right) = \sum_{i=1}^{N} \left(D_{k,1} + D_{k,2}\right). \tag{3.31}$$

De façon générale, en introduisant le paramètre r_N pour contrôler la redondance et dans le cas d'une distribution source Gaussienne généralisé nous avons proposé d'exprimer la distorsion centrale par la relation suivante [301] :

$$\begin{aligned} D_{k,0}\left(\frac{q_{k,1}}{\sigma_{k,1}}, \frac{q_{k,2}}{\sigma_{k,2}}\right) = \frac{1}{\sigma_{k,0}^2}\frac{1}{r_N+1}\Bigg[\min\left(\sigma_{k,1}^2 D_{k,1}\left(\frac{q_{k,1}}{\sigma_{k,1}}\right), \sigma_{k,2}^2 D_{k,2}\left(\frac{q_{k,2}}{\sigma_{k,2}}\right)\right) \\ + r_N \times \max\left(\sigma_{k,1}^2 D_{k,1}\left(\frac{q_{k,1}}{\sigma_{k,1}}\right), \sigma_{k,2}^2 D_{k,2}\left(\frac{q_{k,2}}{\sigma_{k,2}}\right)\right)\Bigg] \end{aligned} \tag{3.32}$$

L'écriture de l'équation (3.32) peut se simplifier par :

$$\begin{cases} \frac{\sigma_{k,1}^2}{\sigma_{k,0}^2}\frac{1}{r_N+1} D_{k,1}\left(\frac{q_{k,1}}{\sigma_{k,1}}\right) + \frac{\sigma_{k,2}^2}{\sigma_{k,0}^2}\frac{r_N}{r_N+1} D_{k,2}\left(\frac{q_{k,2}}{\sigma_{k,2}}\right), & \text{si } \sigma_{k,1}^2 D_{k,1} \leq \sigma_{k,2}^2 D_{k,2} \\ \frac{\sigma_{k,2}^2}{\sigma_{k,0}^2}\frac{1}{r_N+1} D_{k,2}\left(\frac{q_{k,2}}{\sigma_{k,2}}\right) + \frac{\sigma_{k,1}^2}{\sigma_{k,0}^2}\frac{r_N}{r_N+1} D_{k,1}\left(\frac{q_{k,1}}{\sigma_{k,1}}\right), & \text{sinon} \end{cases} \tag{3.33}$$

La quantité de redondance r_N, c'est-à-dire l'importance de la sous-bande redondante, dépend du BER canal. Nous proposons une stratégie dans le paragraphe suivant pour évaluer cette quantité.

Le paramètre de redondance En prenant en compte ce que nous avons dit dans le paragraphe précédent, il est facile de conlure que le paramètre r_N appartient au domaine $[0, 1]$. Ainsi, $r_N = 0$ lorsque le canal est sans bruit et $r_N = 1$ lorque le canal est très bruité. Cependant, tout le problème réside dans le choix de r_N pour des niveaux de bruit canal intermédiaires. Shannon a montré (théorème 10 de la référence [311]) que l'entropie conditionnelle $H_y(x)$ correspondait à la quantité de redondance dont le décodeur a besoin pour "corriger" le signal reçu. En utilisant ce résultat, nous avons proposé de calculer le paramètre de redondance par [301], [15] :

$$r_N = \frac{H_y(x)}{\max\limits_{p_S(x)} H(x)}, \tag{3.34}$$

où $p_S(x)$ est la distribution des symboles sources en entrée du canal. La difficulté pour évaluer ce paramètre est due au fait que $H_y(x)$ est inconnu au

niveau du codeur. Nous avons proposé de borner cette quantité en introduisant la proposition 3 suivante.

Proposition 3 *Soit un canal de capacité C telle que $C = \max_{p_S(x)}(H(x) - H_y(x))$, alors :*

$$\min_{p_S(x)} H_y(x) \leq \max_{p_S(x)} H(x) - C \leq \max_{p_S(x)} H_y(x)$$

La démonstration de cette proposition se trouve dans notre article [15] et dans la thèse de Pereira [301].

En utilisant le résultat de la proposition 3 il est possible d'approximer r_N par la relation suivante :

$$r_N \approx \frac{\max_{p_S(x)} H(x) - C}{\max_{p_S(x)} H(x)} \tag{3.35}$$

Notons que r_N vérifie $0 \leq r_N \leq 1$. Les valeurs de r_N que nous avons estimées pour les canaux *binaire symétrique*, *Gaussien* et *Rayleigh* peuvent être trouvées dans l'article [15] et dans la thèse de Pereira [301].

3.3.2.3 Ecriture des contraintes

Contrainte sur le débit La condition 2 donnée précédemment se traduit par une contrainte égalité "classique" sur les débits latéraux, donnée par

$$P_1(\mathbf{R}_j) = \sum_{k=1}^{N} a_k R_{k,j} \left(\frac{q_{k,j}}{\sigma_{k,j}} \right) - R_l \text{ pour } j \in \{1, 2\}. \tag{3.36}$$

Pénalité sur la distorsion La répartition de la redondance entre les différents descripteurs influe sur la fonction débit-distorsion de chaque descripteur. Ainsi, pour un même débit latéral, il est possible d'obtenir des distorsions latérales différentes en fonction de la répartition de la redondance qui a été faite. De façon à imposer à la distorsion latérale d'être inférieure à une distorsion maximum D_l (pénalité du problème (P') et condition 3), nous avons introduit une pénalité $P(x)$ qui s'exprime par la relation

$$P(x) = \left(\frac{|x| - x}{2} \right)^2, \tag{3.37}$$

entraînant $P(x) = 0$ si la pénalité est vérifiée, c'est-à-dire lorsque $x \geq 0$ (laissant donc le système libre), et $P(x) = x^2$ sinon (pénalisant le système). En

terme de distorsion latérale l'équation (3.37) s'écrit :

$$P_2\left(\mathbf{D}_j\right) = \left[\frac{|\mathbf{D}_j - D_l| + (\mathbf{D}_j - D_l)}{2}\right]^2, \text{ pour } j \in \{1,2\}, \tag{3.38}$$

permettant de vérifier $\sum_{k=1}^{N} \Delta_k \pi_k \sigma_{k,j}^2 D_{k,j}\left(\frac{q_{k,j}}{\sigma_{k,j}}\right) \leq D_l$ pour $j \in \{1,2\}$.

3.3.2.4 Le critère proposé

En considérant la distorsion centrale D_0 donnée par la relation (3.32), la contrainte P_1 et la pénalité P_2 données par les relations (3.36) et (3.38), la fonctionnelle à minimiser donnée à l'équation (3.27) devient :

$$J\left(\{q_{k,1}\},\{q_{k,2}\}\right) = \sum_{k=1}^{N} \Delta_k \pi_k \sigma_{k,0}^2 D_{k,0}\left(\frac{q_{k,1}}{\sigma_{k,1}}, \frac{q_{k,2}}{\sigma_{k,2}}\right)$$
$$+ \sum_{j=1}^{2} \lambda_j \left(\sum_{k=1}^{N} a_k R_{k,j}\left(\frac{q_{k,j}}{\sigma_{k,j}}\right) - R_l\right) + \sum_{j=1}^{2} \mu_j \left[\frac{|\mathbf{D}_j - D_l|}{2} + \frac{(\mathbf{D}_j - D_l)}{2}\right]^2.$$

La solution du problème de minimisation de cette fonctionnelle par rapport à $\{q_{k,1}, k = 1,...,N\}$ et $\{q_{k,2}, k = 1,...,N\}$ est obtenue pour μ_j ($j \in \{1,2\}$) fixés par

$$\begin{cases} \frac{\partial J_\lambda(\{q_{k,1}\},\{q_{k,2}\})}{\partial q_{k,1}} = 0 \\ \frac{\partial J_\lambda(\{q_{k,1}\},\{q_{k,2}\})}{\partial q_{k,2}} = 0 \\ \frac{\partial J_\lambda(\{q_{k,1}\},\{q_{k,2}\})}{\partial \lambda} = 0, \end{cases} \tag{3.39}$$

c'est-à-dire, par le système suivant de $2N+1$ équations à $2N+1$ inconnues :

$$\begin{cases} \frac{\partial D_{k,j}}{\partial R_{k,j}}\left(\frac{q_{k,j}}{\sigma_{k,j}}\right) = \frac{-\lambda_j a_k}{\Delta_k \sigma_{k,j}^2 \left(\frac{C_{k,j}}{1+r_N} + \mu_j E_j\right)} & \text{(a)} \\ \sum_{k=1}^{N} a_k R_{k,j}\left(\frac{q_{k,j}}{\sigma_{k,j}}\right) - R_l = 0. & \text{(b)} \end{cases} \tag{3.40}$$

Le paramètre $C_{k,j}$ est défini par :

$$C_{k,j} = \begin{cases} 1, & \text{si } \min(\sigma_{k,1}^2 D_{k,1}, \sigma_{k,2}^2 D_{k,2}) = \sigma_{k,j}^2 D_{k,j} \\ r_N, & \text{sinon.} \end{cases} \tag{3.41}$$

Il contrôle la répartition de la redondance entre les différents descripteurs. De plus, E_j est tel que

$$E_j = \begin{cases} 2\left(\mathbf{D}_j - D_l\right) & \text{si } \mathbf{D}_j > D_l \\ 0 & \text{sinon} \end{cases} \tag{3.42}$$

Le lecteur intéressé trouvera le développement complet du calcul dans la thèse de Pereira [301] et dans [15]. Nous avons proposé de résoudre ce système par l'algorithme EBWIC décrit au paragraphe 3.2.2.3 en utilisant les modèles de distorsion et de débit théoriques que nous avons introduit aux paragraphes 3.2.1.2 et 3.2.1.3.

3.3.2.5 Algorithme proposé

Le paramètre $C_{k,j}$ défini dans l'équation (3.41) dépend de $\sigma_{k,1}^2 D_{k,1}$ et de $\sigma_{k,2}^2 D_{k,2}$. Cependant, ces valeurs de distorsion ne sont pas connues tant que le système (3.40) n'a pas été résolu puisqu'elle dépendent des pas de quantification $q_{k,j}$. Nous avons donc proposé un algorithme itératif qui permet de modifier les valeurs de $C_{k,j}$ en fonction de l'évolution de $\sigma_{k,1}^2 D_{k,1}$ et de $\sigma_{k,2}^2 D_{k,2}$. De façon précise, si on définit par F l'ensemble de toutes les sous-bandes appartenant aux descripteurs 1 et 2, l'algorithme cherche quelle est la sous-bande i dans F qui présente la plus forte distorsion et affecte la valeur r_N au $C_{k,j}$ du descripteur correspondant et la valeur 1 au $C_{k,j}$ de l'autre descripteur. A chaque itération, les sous-bandes des deux descripteurs dont le paramètre $C_{k,j}$ a été affecté sont enlevées de l'ensemble F. Ainsi, seulement N itérations seront effectuées. Il est alors clair que la convergence vers une solution optimale dépend des conditions initiales de l'algorithme. Une étude complète est fournie dans la thèse de Pereira concernant ce problème [301]. Pour r_N donné, l'algorithme proposé est le suivant :

ALGORITHME

1. **Initialiser :** λ_1 et λ_2, μ_1 et μ_2, $\{C_{k,j}\}$, et $E_1 = E_2 = 0$, $t = 0$.
2. $t \leftarrow t + 1$. *Résoudre le système (3.40) au moyen de l'algorithme EBWIC donné au paragraphe 3.2.2.3.*
3. *Actualiser* $\{D_j\}$ *et* $\{C_{k,j}\}$ *pour tout* $j \in \{1, 2\}$ *et* $k \in \{1, .., N\}$ *à partir des distorsions* $D_{k,j}$ *obtenues à l'étape 2.*
   (a) **Si** *les nouvelles valeurs de* $\{C_{k,j}\}$ *à l'itération* t *sont différentes des anciennes valeurs à l'itération* $t-1$ **alors** *aller en 2* **sinon** *continuer.*
   (b) **Si** $D_j > D_l$ *pour* $j \in \{1, 2\}$ **alors** *calculer* $d_j = D_j - D_l$ *et* $\mu_j^{t+1} = \mu_j^t + \eta_t d_j$ *et aller à l'étape 2* **sinon** *continuer.*
4. *La solution est donnée par* $\{q_{k,j}, k = 1, ..., N$ et $j = 1, 2\}$.

Les pas de déplacement η peuvent être choisis *a priori.*

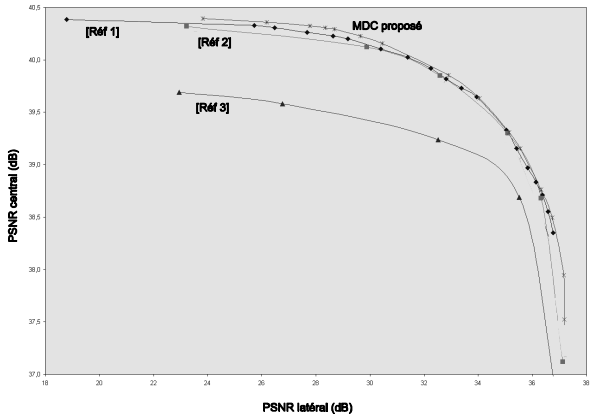


FIG. 3.13 – Comparaison des Rapports Signal-à-Bruit central et latéral pour l'image LENA codée à 1 bpp ($R_1 = R_2 = 0,5$ bpp pour chaque descripteur). Comparaison de notre méthode MDC (pour r_N variable) avec des codeurs MDC existants.

3.3.2.6 Complexité de l'algorithme

Il est difficile de donner une valeur précise de la complexité de cet algorithme qui dépend du nombre d'itérations effectué pour estimer les valeurs des paramètres $C_{k,j}$ ainsi que celles des paramètres μ_j qui permettent de vérifier la pénalité. Si l'on note t ce nombre d'itérations, alors la complexité de l'algorithme peut être approximée d'un premier abord par t fois la complexité d'EBWIC, soit $4t$ opérations par pixel. Une étude plus précise peut être trouvée dans la thèse de Pereira [301].

3.3.3 Résultats

3.3.3.1 Cas des images fixes

Une simulation est donnée figure 3.13 pour l'image LENA. Le codage est effectué au moyen d'un codeur arithmétique contextuel de type EBCOT et le débit total a été fixé à 1 bpp, soit $R_1 = R_2 = 0,5$ bpp. Nous avons tracé la courbe qui représente le Rapport Signal-à-bruit central en fonction du Rapport Signal-à-bruit latéral pour différentes valeurs du paramètre r_N entre 0 et 1. Nous avons comparé notre méthode avec les méthodes données dans [294] (Réf 1), [281] (Réf 2) et [309] (Réf 3). Nous pouvons remarquer qu'en terme de Rapport Signal-à-Bruit, notre méthode présente les meilleurs résultats.

3.3.3.2 Cas de la vidéo

Le codeur MDC est appliqué sur des sous-bandes issues d'une transformée en ondelettes tridimensionnelle (2D+t). Le filtrage spatial utilise les filtres "9-7", et le filtrage temporel les filtres lifting (2,2). Le traitement est effectué au "fil de l'eau" sans prédiction/compensation de mouvement. Toutes les simulations ont été effectuée dix fois, et le Rapport Signal-à-Bruit Pic (PSNR) moyen est évalué à partir de l'EQM moyenne obtenue en moyennant les EQM de chaque réalisation. Le codage est ici aussi effectué au moyen d'un codeur arithmétique contextuel de type EBCOT.

Nous présentons des résultats dans le cas du canal Internet (cf. figures 3.14 et 3.15) et du canal UMTS (cf. figures 3.16 et 3.17). Le canal Internet a été simulé au moyen d'un modèle de Markov à 2 états donné dans [262] en considérant un intervalle $T = 100$ ms entre l'envoi de deux paquets successifs et un taux de pertes de 5%. Pour le canal UMTS, nous avons utilisé un simulateur fourni par France Télécom R&D [271]. Les résultats obtenus avec le codeur MDC proposé sont comparés à ceux obtenus en utilisant une seule description (SDC) avec le même codec sur un canal "véhicule" [270] et un BER égal à 0,001. Des résultats comparatifs avec un codec SDC incluant un codage correcteur d'erreur basé sur les Turbo Codes [261] et un canal Gaussien sont donnés à la figure 3.18. Dans ce cas, les débits annoncés dans les simulations (200 Kb/s) prennent en compte le débit lié au codage canal.

3.4 Maillage 3D et distance surfacique

Ces travaux ont été réalisés durant la thèse de Frédéric Payan commencée en octobre 2000 [300] et que j'encadre actuellement dans le cadre d'un contrat Région PACA / Entreprise avec la société Opteway.

3.4.1 Le problème posé

Il existe typiquement deux approches pour représenter et comprimer les maillages surfaciques 3D : les représentations monorésolutions et les représentations multirésolutions. Dans le premier groupe de méthodes il est important de citer les travaux de Deering [269], Touma et Gotsman [324], Taubin et Rossignac [320] et Rossignac [306]. Les méthodes développées dans ce cadre monorésolution sont généralement non progressives et se basent sur la réduction de la représentation topologique de l'objet combinée avec une simple quantification scalaire de la géométrie. Dans le deuxième groupe de méthodes, on peut trouver principalement les travaux de Khodakovsky, Schröder et Sweldens sur l'agorithme PGC [287] qui est actuellement la référence en terme débit-distorsion et ceux de Lounsbery, DeRose et Warren [292]. Les méthodes de compression

FIG. 3.14 – **Codec SDC** -Vidéo FOREMAN comprimée à 200 Kbit/s et transmise sur un simulateur Internet avec un taux de pertes de 5%.

FIG. 3.15 – **Codec MDC proposé** -Vidéo FOREMAN comprimée à 200 Kbit/s et transmise sur un simulateur Internet avec un taux de pertes de 5%.

FIG. 3.16 – **Codec SDC** - Vidéo SILENT comprimée à 200 Kbit/s et transmise sur un simulateur UMTS (canal "véhicule") avec un BER égal à 0,001.

FIG. 3.17 – **Codec MDC proposé** - Vidéo SILENT comprimée à 200 Kbit/s et transmise sur un simulateur UMTS (canal "véhicule") avec un BER égal à 0,001.

FIG. 3.18 – Image silent comprimée à 200 Kbit/s et transmise sur un canal Gaussien avec un BER égal à 0,001. **MDC proposé** : deux colonnes de gauche. **SDC et Turbo Codes** : deux colonnes de droite.

géométriques multirésolutions existantes se basent sur la quantification et le codage des coefficients issus de la transformée en ondelettes sur des maillages 3D. Ces coefficients se répartissent dans différentes sous-images 3D comme il a été précisé dans le chapitre 2. Une approche basée sur la quantification vectorielle des sommets a aussi été récemment proposée dans [266]. Le lecteur intéressé peut trouver un état de l'art complet des méthodes de compression de maillages dans [321] et [254]. .

Dans le cadre multirésolution et pour assurer une compression efficace des différentes sous-bandes, les paramètres de quantification et de codage doivent être estimés de façon à optimiser le compromis débit-distorsion. Il est alors nécessaire d'effectuer une allocation des débits (ou de la qualité) au travers des différentes sous-bandes issues de la transformation. Cette optimisation est un point délicat et important qui établit les performances de la méthode de compression. Dans nos travaux, nous avons proposé une extension de l'algorithme EBWIC pour le codage de la géométrie des maillages représentée par les coefficients d'ondelettes. En effet, cette approche est réaliste lorsqu'il s'agit de traiter des données de tailles énormes et lorsque la complexité du système de compression/décompression doit être minimale. La mise en place d'un tel système d'allocation de ressources binaires a nécessité de résoudre différents problèmes que nous pouvons décomposer en deux axes principaux de recherche :

- Choix de la mesure de distorsion ou de qualité. Généralement, l'EQM est le critère optimisé pour la compression des données. Cependant, l'estimation de qualité d'une image 3D se fait au moyen du calcul d'une distance surface à surface (type distance de Hausdorff) entre l'objet comprimé et l'objet original et non pas par un calcul de l'EQM sommet à sommet [95]. Il est donc primordial d'intégrer une approximation de cette distance surface à surface dans le schéma d'allocation/compression ;
- Modélisation statistique des coefficients d'ondelettes géométriques 3D au moyen de distributions de type Gaussienne généralisée. On utilise ces modèles pour estimer théoriquement la distorsion de quantification et le débit, et on cherche à être le plus proche possible des valeurs expérimentales.

Le schéma général de la méthode proposée est donné sur la figure 3.19.

3.4.2 Information tangentielle ou information normale ?

Les coefficients d'ondelettes représentent l'information géométrique du maillage semi-régulier, alors que les coefficients basses fréquences correspondent à une représentation grossière et irrégulière de l'objet 3D original. De part la structure semi-régulière du maillage, l'information topologique se retrouve dans cette image basse fréquence et se résume à un nombre restreint de connexions qu'il

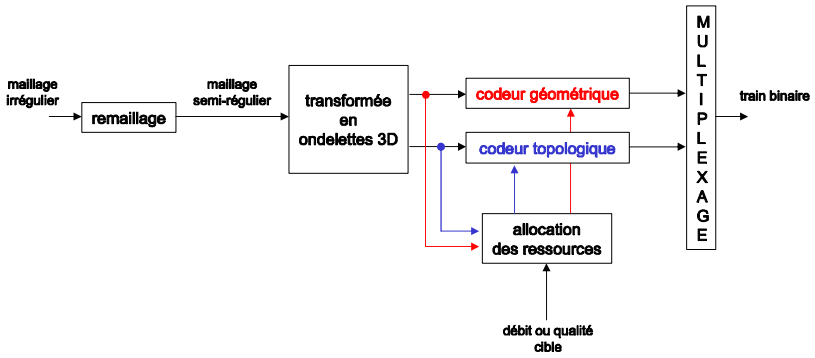


FIG. 3.19 – Schéma général de principe du codeur de maillages 3D proposé.

est possible de coder au moyen d'un codeur topologique classique tel que celui défini dans [324]. Les coefficients géométriques ou sommets sont des vecteurs 3D que nous noterons $C_m = (c_{1,m}, c_{2,m}, c_{3,m})$ à la résolution m. A chaque résolution, les coefficients C_m sont calculés dans un repère local généré à partir du plan tangent à la surface de l'objet au sommet considéré. Ainsi, la coordonnée $c_{3,m}$ et les coordonnées $c_{1,m}$ et $c_{2,m}$ dépendent respectivement du vecteur normal et du plan tangent associé à chaque sommet. Nous pouvons alors séparer l'information géométrique en deux sous-ensembles [287] :

- Le sous-ensemble tangentiel S_{T_m} d'indice de résolution m. Il contient les coefficients d'ondelettes d'indice de résolution m qui apportent une information *tangentielle*. On notera $s_{1,m} = (c_{1,m}, c_{2,m}) \in S_{T_m}$ un vecteur tangent ;
- Le sous-ensemble normal S_{N_m} d'indice de résolution m. Il contient les coefficients d'ondelettes d'indice de résolution m qui apportent une information *normale*. On notera $s_{2,m} = c_{3,m} \in S_{N_m}$ un coefficient normal.

Nous avons montré dans [81] que ces deux ensembles pouvaient être traités séparément et que la distribution de chacun pouvait être modélisée par une distribution Gausienne généralisée. D'autre part, il a été montré dans [287] que la distribution des coefficients d'ondelettes dans un repère local se répartie généralement le long de l'axe normal. Les coefficients normaux $s_{2,m} \in S_{N_m}$ ont donc une influence relativement importante lors de l'évaluation de la distorsion qui s'effectue par une mesure de la distance surface à surface. Nous avons donc introduit une pondération β dans le critère de distorsion lors de l'allocation des débits qui favorise la minimisation de l'EQM pour les coefficients normaux au détriment des vecteurs tangents lors de la quantification.

3.4.3 Critère proposé

Dans le cas d'un filtre à M bandes, le problème d'allocation de débits que nous avons à résoudre exprimé en fonction de la bande se pose de la façon suivante :

$$(P)\begin{cases} \min\limits_{\{\overline{q}_{i,m}\}} D\left(\{\overline{q}_{i,m}\}\right) \\ \text{sous la contrainte } \sum_{m=1}^{\#m}\sum_{i=0}^{M-1} a_{i,m} R_{i,m}\left(\overline{q}_{i,m}\right) \leq R_{cible} \end{cases} \tag{3.43}$$

où $\#m$ représente le nombre de niveaux de résolution, $a_{i,m}$ le poids de la sous-bande i à l'indice de résolution m et $\overline{q}_{i,m} = (q^i_{s_{1,m}}, q^i_{s_{2,m}})$ le vecteur de pas de quantification dans le canal i à l'indice de résolution m qui dépend du pas de quantification des coefficients tangents $s_{1,m}$ et de celui des coefficients normaux $s_{2,m}$. L'expression de la distorsion D est donnée par :

$$D = \pi^*_{0,\#m}\overline{D}\left(\overline{q}_{0,\#m}\right) + \sum_{m=1}^{\#m}\sum_{i=1}^{M-1} \pi^*_{i,m}\overline{D}\left(\overline{q}_{i,m}\right), \tag{3.44}$$

avec les pondérations π^* données au chapitre 2. La distorsion $\overline{D}$ s'exprime en fonction des sous-ensembles tangentiel et normal par :

$$\overline{D}\left(\overline{q}\right) = \sigma^2_{s_1} D\left(q_{s_1}\right) + \beta\sigma^2_{s_2} D\left(q_{s_2}\right), \text{ avec } \overline{q} = (q_{s_1}, q_{s_2}) \text{ et } \beta \geq 1 \tag{3.45}$$

En introduisant les opérateurs de Lagrange et en supposant l'existence d'une solution, ce problème de minimisation peut être résolu par l'algorithme EBWIC présenté au paragraphe 3.2.2.3. Nous ne détaillerons donc pas ici les calculs qui sont similaires à ceux du paragraphe 3.2.2.2 donnés dans le cas d'une allocation de "qualité" et qui peuvent être trouvés dans la thèse de Payan [300].

3.4.4 Résultats

Nous présentons sur les figures 3.20 et 3.21 des résultats comparatifs en terme de Rapport Signal-à-Bruit pic (PSNR) entre la méthode que nous avons proposé, l'algorithme PGC ("Progressive Geometry Compression") [287] et le codeur topologique donné dans [324]. Les deux codeurs multirésolutions utilisent un schéma lifting basé sur le filtre de Butterfly.

Pour chaque image, le PSNR est évalué entre le maillage original irrégulier et le maillage comprimé/décomprimé semi-régulier au moyen d'une distance surface à surface d [256]. Pour des objets 3D il est défini par la relation :

$$PSNR_{dB} = 20\log_{10}\frac{P}{d} \tag{3.46}$$

avec P la diagonale du cube 3D qui englobe l'objet. Le codeur entropique utilisé par notre méthode est basé sur le codeur de JPEG-2000 avec des contextes 3D

multirésolutions que nous avons introduit et présenté dans [95]. Le débit est évalué en bits par sommet irrégulier.

Les courbes présentées montrent la supériorité des méthodes multirésolutions géométriques par rapport aux méthodes monorésolutions topologiques. En terme de PSNR, le codeur que nous avons développé est aussi performant que le codeur PGC basé sur l'algorithme SPIHT [307]. Cependant, l'allocation étant effectuée à partir de modèles théoriques du débit et de la distorsion, nous bénéficions d'une faible complexité algorithmique de l'ordre de celle d'EBWIC [300]. Des comparaisons visuelles sont présentées à la figure 3.22.

3.5 Conclusion-Synthèse

Les méthodes classiques de détermination des pas de quantification pour les différentes sous-bandes d'une transformée en ondelettes requièrent un codage exhaustif pour calculer les courbes débit-distorsion et retenir la meilleure solution. A l'inverse de ces méthodes, nous avons proposé un algorithme d'allocation de débits (pas de quantifications) basé sur des modèles théoriques de distorsion et de débit. Cet algorithme d'allocation dynamique de ressources binaires, connu sous le nom d'EBWIC, a été introduit dans le cadre du codage par transformée en ondelettes "au fil de l'eau" et permet soit un contrôle du débit binaire, soit un contrôle de la "qualité" image en terme d'erreur quadratique moyenne ou de rapport signal-à-bruit. Cette méthode nous a servi comme brique de base pour construire un algorithme de codage source/canal par descriptions multiples efficace pour la transmission de données images ou vidéos sur des réseaux Internet ou encore de troisième génération (UMTS). De plus, nous avons montré qu'EBWIC était aussi efficace pour la compression de maillages 3D et pouvait concurrencer en terme de compromis débit-distorsion-complexité les meilleures méthodes de compression actuelles telles que le standard JPEG-2000.

3.6 Références

[254] P. Alliez and C. Gotsman, "Recent advances in compression of 3d meshes," *To appear in Proceedings of the Symposium on Multiresolution in Geometric Modeling*, 2003.

[255] J. Apostolopoulos, "Error-resilient video compression via multiple state streams," in *International Workshop on Very Low Bitrate Video Coding (VLVB)*, octobre 1999, pp. 168–171.

[256] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, "Mesh : Measuring errors between surfaces using the hausdorff distance," in *Proceedings of the*

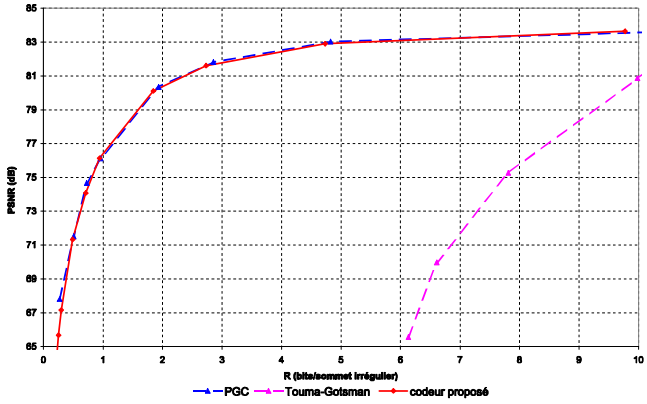


FIG. 3.20 – Comparaison de différentes méthodes de compression pour l'objet RABBIT.

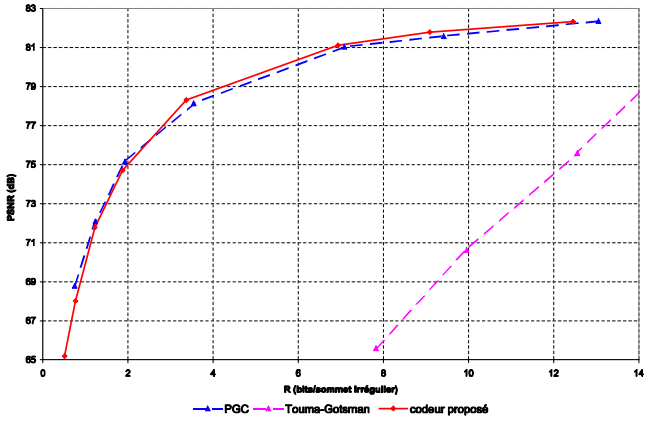


FIG. 3.21 – Comparaison de différentes méthodes de compression pour l'objet VENUS.

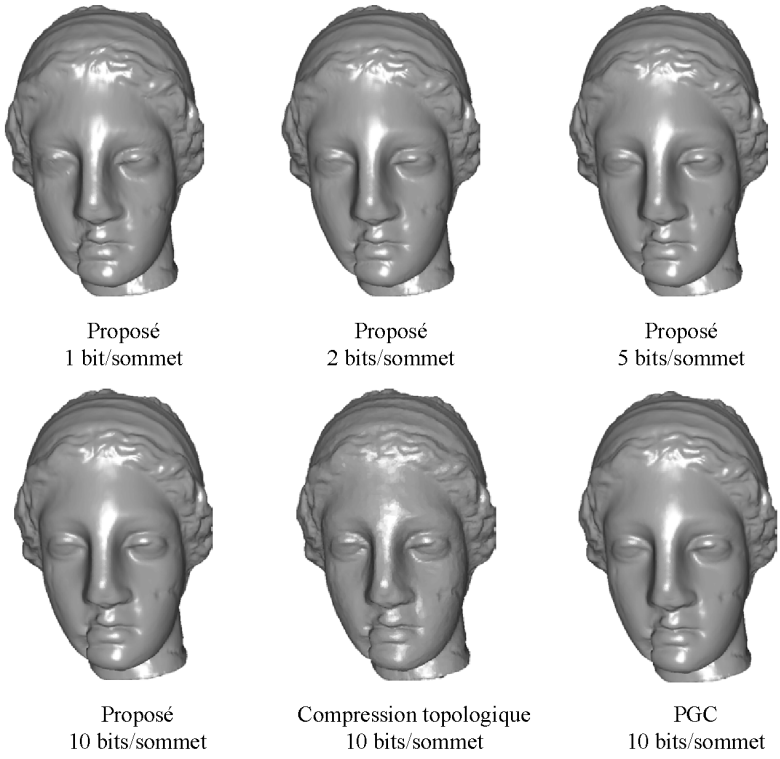


FIG. 3.22 – Objet VENUS comprimé par la méthode proposée. Comparaison avec l'algorithme PGC de Khodakovsky, Schröder et Sweldens et l'algorithme de compression topologique de Touma et Gotsman.

IEEE International Conference on Multimedia and Expo, 2002, vol. I, pp. 705 – 708, `http ://mesh.epfl.ch`.

[257] N. At and Y. Altunbasak, "Multiple description coding for wireless channels with multiple antennas," in *GLOBECOM*, San Antonio, TX, november 2001, pp. 2040–2044.

[258] W.R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, juillet 1948.

[259] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, N.J. : Prentice-Hall, 1971.

[260] T. Berger, "Optimum quantizers and permutation codes," *IEEE Transactions on Information Theory*, vol. 18, pp. 149–157, novembre 1972.

[261] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding : Turbo-codes," in *Proc. of IEEE ICC*, Geneve, Suisse, mai 1993, pp. 1064–1070.

[262] J. Bolot, "Characterizing end-to-end packet delay and loss in the internet," *Journal of High-Speed Networks*, 1993.

[263] Q. Chen and T. Fisher, "Robust quantization for image coding and noisy digital transmission," in *Data Compression Conference*, Snowbird, UT, 1996, pp. 3–12.

[264] Q. Chen and T. Fisher, "Image coding using robust quantization for noisy digital transmission," *IEEE Trans. on Image Processing*, 1996.

[265] P.A. Chou, T. Lookabaugh, and R.M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Transactions on Information Theory*, vol. 35, pp. 299–315, Mars 1986.

[266] P.H. Chou and T.H.Y. Meng, "Vertex data compression through vector quantization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 4, pp. 373–382, 2002.

[267] P.F. Panter A.G. Clavier and D.D. Grieg, "Distortion in a pulse count modulation system," *AIEE Transactions*, vol. 66, pp. 989–1005, 1947.

[268] P.F. Panter A.G. Clavier and D.D. Grieg, "Pcm distortion analysis," *Electrical Engineering*, pp. 1110–1122, novembre 1947.

[269] M. Deering, "Geometry compression," *Proceedings SIGGRAPH*, 1995.

[270] *ETSI (European Telecommunications Standards Institute, Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS*, tr 101 112 v3.2.0 edition, avril 1998.

[271] J. Farah Francis, "Etude de systèmes de traitement numérique de canal pour la réception radio-mobile umts/tdd," M.S. thesis, Institut Natonal Polytechnique de Grenoble, INPG, Grenoble, France, septembre 2002.

[272] A. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, pp. 851–857, novembre 1982.

[273] J. Garcia-Frias and J. Villasenor, "An analytical treatment of channel-induced distortion in entropy coded image subbands," in *Data Compression Conference*, Snowbird, UT, 1997.

[274] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[275] H. Gish and J.N. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 14, pp. 676–683, septembre 1968.

[276] V. Goyal and J. Kovacevic, "Optimal multiple description transform coding of gaussian vectors," in *International Conference on Image Processing (ICIP)*, Chicago, Etats-Unis, octobre 1998.

[277] V. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of image," in *International Conference on Image Processing (ICIP)*, Chicago, Etats-Unis, octobre 1998.

[278] V. Goyal, J. Kovacevic, and M. Vetterli, "Multiple description transform coding : Robustness to erasures using tight frame expansions," in *Proc. International Symposium on Information Theory*, Cambridge, MA, août 1998, p. 408.

[279] R.M. Gray, *Source Coding Theory*, Kluwer Academic Publishers, 1990.

[280] R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2384, octobre 1998.

[281] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proc. SPIE Visual Commun. Image Proc. Conf. (VCIP)*, 1999.

[282] ISO/IEC 15444-1 :2000, "Information technology – JPEG 2000 image coding system – part 1 : Core coding system," .

[283] ISO/IEC JTC1/SC29/WG01, "VM9.0 software," Document wg1n2131, avril 2001.

[284] N. Kamaci, Y. Altunbasak, and R. Mersereau, "Multiple description coding with multiple transmit and receive antennas for wireless channels : The case of digital modulation," in *GLOBECOM*, San Antonio, TX, novembre 2001, pp. 3272–3276.

[285] J.H. Kasner, M.W. Marcellin, and B.R. Hunt, "Universal trellis coded quantization," *IEEE Transactions on Image Processing*, vol. 8, no. 12, pp. 1677–1687, décembre 1999.

[286] K. Kintzley, "An application of multiple description scalar quantizers to speech coding on correlated fading channels," M.S. thesis, Electrical Engineering Departement, Texas A&M University, College Station, TX, août 1995.

[287] A. Khodakovsky, P. Schröder, and W. Sweldens, "Progressive geometry compression," *Proceedings of SIGGRAPH*, 2000.

[288] C. Lambert-Nebout and G. Moury, "A survey of on board image compression for CNES space missions," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 1999.

[289] P. Lier, G. Moury, C. Latry, and F. Cabot, "Selection of the SPOT-5 image compression algorithm," in *Proceedings of SPIE Earth Observing Systems Conference III*, San Diego, USA, juillet 1998, vol. 3439, pp. 541–552.

[290] S.P. Lloyd, "Least squares quantization in PCM," *Unpublished Bell Laboratories Technical Note. Portions presented at the Institute of Mathematical Statistics Meeting Atlantic City New Jersey September 1957. Published in special issue on quantization, IEEE Transactions on Information Theory*, 1982.

[291] J.M. Lounsbery, *Multiresolution Analysis for Surfaces of Arbitrary Topological Type*, Ph.D. thesis, Seattle, WA, Etats-Unis, 1994.

[292] M. Lounsbery, T. DeRose, and J. Warren, "Multiresolution analysis for surfaces of arbitrary topological type," *ACM Transactions on Graphics 16,1*, vol. 99, 1997.

[293] J. Max, "Quantizing for minimum distortion," *IEEE Transactions on Information Theory*, pp. 7–12, mars 1960.

[294] A. Miguel, A. Mohr, and E. Riskin, "Spith for generalized multiple description coding," in *IEEE International Conference on Image Processing, Kobe, Japon*, octobre 1999, vol. 3, pp. 842–846.

[295] A. Mohr, E. Riskin, and R. Ladner, "Graceful degradation over packet erasures channels through forward error correcting," in *Proc. IEEE Data Compression Conf (DCC)*, 1999.

[296] B.M. Olivier, J. Pierce, and C.E. Shannon, "The philosophy of PCM," in *Proc. IRE*, novembre 1948, vol. 36, pp. 1324–1331.

[297] L. Ozarow, "On a source coding problem with two channels and three receivers," décembre 1980, vol. 59, pp. 1909–1921.

[298] P.F. Panter and W. Dite, "Quantizing distortion in pulse-count modulation with nonuniform spacing of levels," in *Proc. IRE*, janvier 1951, vol. 39, pp. 44–48.

[299] C. Parisot, *Allocations basées modèles et transformée en ondelettes au fil de l'eau pour le codage d'images et de vidéos*, Ph.D. thesis, Université de Nice - Sophia Antipolis, France, janvier 2003.

[300] F. Payan, *Compression Multirésolution de Maillages Géométriques 3D*, Ph.D. thesis, Université de Nice - Sophia Antipolis, France, soutenance prévue en décembre 2003.

[301] M. Pereira, *Compression par Descriptions Multiples pour la Transmission d'Images et de Vidéos sur Canaux Bruités*, Ph.D. thesis, Université de Nice - Sophia Antipolis, France, soutenance prévue en mai 2004.

[302] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description video coding using motion-compensated prediction," in *IEEE International Conference on Image Processing, Kobe, Japon*, octobre 1999.

[303] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distorsion sense," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 160–176, avril 1993.

[304] D. Redmill and N. Kingsbury, "Still image coding for noisy channels," in *IEEE International Conference on Image Processing*, 1994, pp. 95–99.

[305] J. Rogers and P. Cosman, "Robust wavelet zerotree image compression with fixed length packetization," *IEEE Sig. Proc. Letters*, vol. 5, no. 5, pp. 105–107, mai 1998.

[306] J. Rossignac, "Geometric simplification and compression," *ACM SIGGRAPH*, 1997.

[307] A. Said and W.A. Pearlman, "A new fast and efficient coder based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technologies*, vol. 6, pp. 243–250, juin 1996.

[308] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Transactions on Information Theory*, vol. 22, pp. 162–169, Mars 1976.

[309] S. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," in *IEEE International Conference on Image Processing, Chicago, Etats-Unis*, octobre 1998.

[310] S. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. on Image Processing*, 1999.

[311] C.E. Shannon, "A mathematical theory of communication," in *Bell Syst. Tech. J.*, 1948, vol. 27, pp. 379–423, 623–656.

[312] C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Record*, Part 4, Ed., 1959, pp. 142–163.

[313] J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445–3462, décembre 1993.

[314] W.F. Sheppard, "On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a scale," in *Proc. London Math. Soc.*, Pt. 2, Ed., 1898, vol. 24, pp. 353–380.

[315] P. Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Processing Lett.*, vol. 4, pp. 189–191, juillet 1997.

[316] P. Sherwood and K. Zeger, "Error protection for progressive image transmission over memoryless and fading channels," *IEEE Trans. on Commun.*, vol. 46, no. 12, décembre 1998.

[317] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1445–1453, septembre 1988.

[318] H. Steinhaus, "Sur la division des corps matériels en parties," *Bull. Acad Polon. Sci.*, vol. IV, pp. 801–804, 1956.

[319] N. Tanabe and N. Farvardin, "Subband image coding using entropy coded quantization over noisy channels," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 926–943, juin 1992.

[320] G. Taubin and J. Rossignac, "Geometric compression through topological surgery," *ACM Transactions on Graphics*, April 1998.

[321] G. Taubin and J. Rossignac, "3d geometry compression," *Course notes, ACM SIGGRAPH*, , no. 21, 1999.

[322] D. Taubman, "Embedded block coding with optimized truncation," ISO/IEC JTC 1/SC29/WG1 document N 1020R, octobre 1998.

[323] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, juillet 2000.

[324] C. Touma and C. Gotsman, "Triangle mesh compression," *Graphics Interface*, pp. 26–34, 1998.

[325] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proceedings of Data Compression Conference*, Snowbird, USA, mars 1996.

[326] V. Vaishampayan, "Vector quantizer design for diversity systems," in *CISS*, 1991.

[327] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.

[328] V. Vaishampayan, "Design of entropy-constrained multiple description scalar quantizers," *IEEE Trans. on Information Theory*, vol. 40, no. 1, pp. 245–250, 1994.

[329] V. Vaishampayan and S. John, "Interframe balanced-multiple-description video compression," in *Packet Video Workshop*, Columbia University, NY, 1998.

[330] V. Vaishampayan and S. John, "Interframe balanced-multiple-description video compression," in *IEEE International Conference on Image Processing, Kobe, Japon*, octobre 1999.

[331] V.A. Vaishampayan, N.J.A. Sloane, and S.D. Servetto, "Multiple description vector quantization with lattice codebooks : Design and analysis," *IEEE Trans. on Information Theory*, vol. 47, pp. 1718–1734, 2001.

[332] S. Valette, Y.S. Kim, H.Y. Jung, I. Magnin, and R. Prost, "A multiresolution wavelet scheme for irregularly subdivided 3d triangular mesh," *IEEE International Conference on Image Processing, Kobe, Japon*, vol. 1, pp. 171–174, octobre 1999.

[333] Y. Wang, M. Orchard, and A. Reibman, "Multiple description image coding for noisy channels by paring transform coefficients," in *IEEE 1997 First Workshop on Multimedia Signal Processing*, 1997.

[334] P.H. Westerink, J. Biemond, and D.E. Boekee, "An optimal bit allocation algorithm for subband coding," 1988, pp. 757–760.

[335] H.S. Witsenhausen and A.D. Wyner, "Source coding for multiple descriptions ii : A binary source," décembre 1980.

[336] H.S. Witsenhausen, "On source networks with minimal breakdown degradation," juillet-août 1980, vol. 59, pp. 1083–1087.

[337] J. Wolf, A.D. Wyner, and J. Ziv, "Source coding for multiple descriptions," octobre 1980, vol. 59, pp. 1417–1426.

[338] S. Yang and V. Vaishampayan, "Low delay communication for rayleigh fading channels : An application of the multiple description quantizer," *IEEE Trans. on Communication*, vol. 43, pp. 2771–2783, novembre 1995.

[339] P.L. Zador, "Topics in the asymptotic quantization of continuous random variables," in *Bell Laboratories Technical Memorandum*, 1966.

Chapitre 4

La quantification vectorielle et les réseaux réguliers de points

Dans ce chapitre je développe les activités de recherche que j'ai effectuées ou que j'effectue encore dans le domaine lié à la quantification vectorielle pour des applications de compression d'images 2D. Le plan de ce chapitre est le suivant. Tout d'abord j'introduis dans le paragraphe 4.1 la quantification vectorielle sur des réseaux réguliers de points ou encore quantification vectorielle algébrique. La difficulté recontrée pour la conception d'un quantificateur algébrique est liée au problème de dénombrement et d'indexage des vecteurs dans le réseau. Je présente aux paragraphes 4.2 et 4.3 les solutions que nous avons proposées pour résoudre ces problèmes dans le cas d'une source de distribution Gaussienne généralisée. Le paragraphe 4.4 présente la modélisation de la distorsion et du débit vectoriels que nous avons développé pour le réseau Z^n et une distribution source unimodale.

4.1 La quantification vectorielle algébrique

La quantification a été étudiée depuis quelques dizaines d'années maintenant et les travaux effectués ont développé de nombreux résultats aujourd'hui classiques sur la théorie débit-distorsion (deux excellents articles font un historique de la quantification [347], [359]). En particulier, il a été montré que la Quantification Vectorielle (QV[1]) possède de nombreux avantages par rapport à la Quantification Scalaire (QS) lorsqu'un codage à longueur fixe est imposé [368], [396]. De plus, les travaux de Shannon ont montré que les performances de la QV sont proches des performances théoriques optimales si la dimension n des vecteurs de quantification est suffisamment grande. Cependant, il est

[1] QV sera utilisé dans le reste du document soit pour désigner la quantification vectorielle, soit pour désigner un quantificateur vectoriel.

important de noter que la QV peut se rapprocher de ces performances optimales au prix d'une complexité calculatoire élevée ; complexité qui augmente de façon exponentielle avec la dimension des vecteurs. Généralement, la QV est effectuée en utilisant un dictionnaire non structuré construit à partir de données statistiques représentatives de la source (séquence d'apprentissage). Dans ce cas, la complexité ainsi que les besoins en stockage dus à la taille du dictionnaire peuvent devenir prohibitifs pour des applications de compression. De plus, il existe un problème de robustesse du dictionnaire qui, optimisé pour une séquence d'apprentissage donnée, donne des performances mauvaises pour une image en dehors de la séquence d'apprentissage [360]. Une solution pour surmonter ces problèmes est d'utiliser un QV structuré n-dimensionnel tel que la Quantification Vectorielle Algébrique (QVA) ou encore quantification vectorielle sur réseaux réguliers de points. Comme les vecteurs du dictionnaire sont contraints à appartenir à un réseaux régulier structuré, les performances de la QVA sont en général inférieures à celle de la QV non structurée. Mais dans la plupart des applications, ce léger désavantage est compensé par le fait que pour la QVA aucun dictionnaire a besoin d'être généré ou stocké, et que la complexité de codage est réduite.

La quantification par réseaux réguliers de points peut être vue comme une généralisation de la quantification scalaire uniforme. Comme dans le cas de la QV non structurée, la QVA prend en compte les *dépendances spatiales* entre les coefficients des vecteurs ainsi que les gains de *partitionnement* et de *forme* [368]. Quelle que soit la distribution de la source, la QVA est toujours plus efficace que la QS. Un réseaux régulier de point Λ dans $\mathbb{R}^n$ est composé par toutes les combinaisons possibles d'un ensemble de vecteurs linéairement indépendants $\mathbf{a}_i$ qui constitue la base du réseau, tel que :

$$\Lambda = \{\mathbf{y}/\mathbf{y} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + ... + u_n\mathbf{a}_n\} \tag{4.1}$$

où les coefficients u_i sont des entiers. La partition de l'espace est alors régulière et ne dépend seulement que des vecteurs de base $\mathbf{a}_i \in \mathbb{R}^m$ $(m \geq n)$ choisis. Chaque base définie un réseau régulier de points différent.

La QVA offre la possibilité de réduire considérablement les coûts calcul et stockage par rapport à un QV conçu au moyen d'algorithmes basés sur l'algorithme de Lloyd généralisé [367]. En effet, utiliser les vecteurs d'un réseau régulier comme valeurs de quantification élimine l'opération de construction d'un dictionnaire : le dictionnaire est construit implicitement de part la structure du réseau choisi. De plus, Conway et Sloane [344] ont proposé des algorithmes rapides de quantification qui utilisent simplement des opérations d'arrondis et qui ne dépendent que de la dimension n des vecteurs. Toutes ces raisons font que la QVA est devenue très populaire et a été particulièrement étudiée ces dernières années [349], [356], [396], [373], [3]. En 1979, Gersho a émis la conjecture que dans le cas asymptotique (c'est-à-dire pour des forts débits), les performances

débit-distorsion d'un QVA sont approximativement optimales [353]. Toutefois, bien que la QVA ne soit pas mathématiquement optimale pour des faibles débits, la réduction de complexité apportée par de tels quantificateurs permet d'utiliser de grandes dimensions de vecteurs, ce qui entraîne de meilleures performances expérimentales à débit donné. Dans [373], les auteurs montrent que de bonnes performances débit-distorsion peuvent être obtenues en combinant la QVA avec un codeur entropique, ce qui a motivé quelques travaux sur la QVA dans le domaine des ondelettes comme par exemple les travaux de [373], [3], [378], [364]. Dans [345], [346], Conway et Sloane ont étudié les réseaux de points qui présentaient les meilleurs gains de partionnement. Les réseaux A_n ($n \geq 1$), D_n ($n \geq 2$), E_n ($n = 6, 7, 8$), le réseau de Barnes-Wall (Λ_{16}) en dimension 16 et le réseau de Leech (Λ_{24}) en dimension 24 sont optimaux dans ce sens. D'autre part, la théorie asymptotique a permis de modéliser les réseaux réguliers de points et de mieux comprendre la QVA [354], [380], [368], [348], [362].

Cependant, peu de travaux ont été réalisés pour les bas débits sur les QVA à débit variable. Comme il a été montré dans l'article de Sripad et Snyder [375], la distortion granulaire[2] dépend du type de réseau mais aussi de la distribution de la source. Ainsi, la hiérarchie des réseaux établie dans le cas d'une théorie haute résolution, qui suppose des distributions sources constantes par morceaux, n'est plus valide. Beaucoup de travaux théoriques sur la QV ont été menés pour des sources Gaussienne et Laplacienne, cependant, dans le cas de sources de type Gaussienne généralisée avec un paramètre de décroissance inférieur à un, il a été montré par [351], [352] que le réseau cubique $\mathbb{Z}^n$ était plus performant en terme de débit-distorsion que le réseau E_8 et le réseau de Leech. Ce résultat a motivé nos travaux pour combiner la QVA avec la transformée en ondelettes.

Si la quantification par QVA est peu complexe, du fait de la structure géométrique régulière du dictionnaire, sa mise en œuvre n'est cependant pas immédiate. Elle soulève en effet un certain nombre de questions concrètes, notamment en termes de calcul et de stockage. Dans mes travaux, j'ai mis en avant deux problèmes fondamentaux dans la conception d'un QVA :

- **L'indexage.** L'indexage est une opération indépendante de la quantification, elle consiste à assigner à chaque vecteur quantifié, un indice (ou index) qui, une fois codé est transmis sur un canal jusqu'au décodeur. Cette opération est fondamentale dans la chaîne de compression. Elle conditionne en effet le débit binaire et permet le décodage sans ambiguïté des vecteurs. Différentes approches purement analytiques ont été proposées, notamment par Conway et Sloane [346], Lamblin et Adoul [365] et Fischer [349]. Ces méthodes sont généralement très peu coûteuses en

[2] Dans le cas de modèles non-asymptotiques le bruit de surcharge est négligé puisque nous supposons l'utilisation d'un codage à longueur variable et d'un dictionnaire infini.

mémoire mais présentent une complexité calculatoire non négligeable (algorithmes récursifs), ou fonctionnent seulement dans des cas particuliers (type de réseau ou de troncature particuliers). Nous avons proposé une approche plus générale qui permet l'indexage sur des distributions de type Gaussienne généralisée, réalisant un bon compromis coût mémoire/coût calcul [50], [6], [68], [13].

– **Le dénombrement.** Les méthodes d'indexage sont généralement basées sur la connaissance de la population du réseau. Ainsi, nous devons être capable de dénombrer les vecteurs du réseau sur des surfaces (ou à l'interieur de volumes) n-dimensionnels qui dépendent de la distribution de la source. Une approche classique de dénombrement est basée sur l'utilisation de séries génératrices. Dans ce formalisme, et suite à nos discussions, P. Solé a introduit les fonctions Nu qui permettent le dénombrement sur des pyramides c'est-à-dire dans le cas d'une distribution Laplacienne. Nous avons exploité ces résultats pour proposer un QVA adapté au coefficients d'ondelettes [3], [38]. De plus, nous avons introduit les fonctions *thêta modifiées* qui permettent le dénombrement sur des distributions elliptiques (Gaussienne non stationnaire) [4], [44].

Notre contribution a porté principalement sur le dénombrement et l'indexage ainsi que sur la mise en œuvre de QVA dans le cadre d'applications de compression d'images.

4.2 Le dénombrement sur des hyper-sphères

4.2.1 Problématique

Une approche classique pour le dénombrement est l'utilisation de fonctions génératrices. Ainsi, dans un réseau Λ, le nombre N_m de vecteurs d'énergie m^2 (norme L_2 au carrée par rapport à l'origine) est donné par la série *thêta* du réseau [346]. Les séries *thêta* ont été introduites à l'origine pour des normes L_2 et la théorie qui a été développée constitue un chapitre complet en mathématique. Rappelons simplement ici que ces séries sont données par :

$$\begin{aligned}\Theta_\Lambda(q) &= \sum_{\mathbf{y}\in\Lambda} q^{\|\mathbf{y}\|_2^2} \\ &= \sum_{m=0}^{+\infty} \left[q^{m^2}\right] \left\{\mathbf{y}\in\Lambda / \|\mathbf{y}\|_2^2 = m^2\right\},\end{aligned} \tag{4.2}$$

où $\left[q^{m^2}\right] f(q)$ correspond au coefficient q^{m^2} dans $f(q)$. On peut aussi écrire :

$$\Theta_{\Lambda}(q) = \sum_{m=0}^{+\infty} N_m q^{m^2}, \tag{4.3}$$

avec N_m le nombre de vecteurs *sur* la sphère d'énergie m^2.

Cependant, cette fonction est évidemment inutilisable lorsqu'il s'agit de dénombrer des vecteurs pour des métriques L_p ($\sum_{i=1}^{n} |y_i|^{\alpha}$) avec $0 < p < 2$.

4.2.2 Le cas pyramidal : les séries *Nu*

Ces travaux ont été réalisés en collaboration avec P. Solé chercheur au laboratoire I3S du CNRS et de l'Université de Nice-Sophia Antipolis. Nos travaux ont été publiés dans la revue IEEE Transactions on Image Processing en juillet 1994 : "Lattice vector quantization for multiscale image coding" [3]. Cet article est donné en annexe C du document.

4.2.2.1 Objectifs

Notre objectif était d'adapter la QVA à la statistique des images à quantifier. En effet, la construction d'un QVA est tributaire de la norme choisie pour l'indexation des vecteurs du réseau. Il a été établi que la distribution statistique des coefficients d'ondelettes est donnée par une loi Gaussienne généralisée [2] de paramètre α tel que $0 < \alpha \leq 2$. Cependant, expérimentalement il est possible de remarquer que ce paramètre est le plus fréquemment voisin de 1. Ceci nous a motivé à nous intéresser dans un premier temps par la modélisation Laplacienne de la distribution des coefficients d'ondelettes. En effet, ce modèle est beaucoup plus simple à mettre en œuvre que le modèle Gaussienne généralisée. Ainsi, dénombrer sur des pyramides permettrait de construire un QVA adapté au coefficients d'ondelettes.

En 1986, Fischer a introduit dans son article [349] une formule qui permet un dénombrement à l'intérieur de pyramides de norme L_1 égale à K :

$$P(n, K) = P(n-1, K) + 2\sum_{i=1}^{\lfloor K \rfloor} P(N-1, K-i), \tag{4.4}$$

où $P(1, K) = 2\lfloor K \rfloor + 1$ et $P(n, 0) = 1$. Cette formule récursive présente une grande complexité calculatoire dans le cas où la dimension des vecteurs n et le nombre de pyramides K deviennent grands. D'autre part, elle n'est valide que pour les réseaux réguliers $\mathbb{Z}^n$. Elle est donc inutilisable dans le cas d'applications où la taille n des vecteurs est grande et lorsque le QVA est construit à partir de réseaux autres que le réseau $\mathbb{Z}^n$. Nous avons donc été intéressé par une solution de dénombrement plus simple. Après nos nombreuses discussions, P. Solé a ainsi

introduit les fonctions Nu. Ces fonctions permettent le dénombrement sur des pyramides pour quelques réseaux classiques tels que les réseaux $\mathbb{Z}^n$, D_n, E_8 et Λ_{16}.

4.2.2.2 Les fonctions Nu

Le formalisme des fonctions Nu a été introduit par P. Solé dans [374]. Pour un réseau Λ et de façon analogue au cas sphérique, nous pouvons définir les fonctions Nu par la relation suivante :

$$\begin{aligned} \nu_\Lambda(z) &= \sum_{\mathbf{y}\in\Lambda} z^{\|\mathbf{y}\|_1} \\ &= \sum_{m=0}^{+\infty} [z^m] \left|\{\mathbf{y}\in\Lambda / \|\mathbf{y}\|_1 = m\}\right|. \end{aligned} \tag{4.5}$$

Le coefficient de z^m représente le nombre de vecteurs qui sont localisés *sur* la pyramide d'énergie m. Dans le cas du réseau $\mathbb{Z}^n$, cette formule se résume simplement à :

$$\nu_{\mathbb{Z}^n}(z) = \nu_{\mathbb{Z}}(z)^n = \left(\frac{1+z}{1-z}\right)^n. \tag{4.6}$$

Il est aussi possible de calculer facilement ce dénombrement pour un réseau D_n au moyen des séries Nu. Le réseau D_n correspond aux vecteurs de $\mathbb{Z}^n$ dont la norme L_1 est paire [346]. Ainsi, la fonction Nu d'un tel réseau est donnée par la partie paire de $\nu_{\mathbb{Z}^n}(z)$. Nous pouvons alors montrer que

$$\nu_{D_n}(z) = \frac{1}{2}\left(\nu_{\mathbb{Z}^n}(z) + \nu_{\mathbb{Z}^n}(-z)\right). \tag{4.7}$$

Le dénombrement pour des réseaux plus denses tels que les réseaux E_8 et Λ_{16} est beaucoup plus compliqué à définir. La théorie donnée par P. Solé [374] repose sur la théorie des codes blocs. Je ne détaillerai pas ces travaux ici. Un résumé se trouve dans l'article "Lattice vector quantization for multiscale image coding" [3] donné en annexe C. Nous avons exploité ces résultats pour contruire des QVA adaptés à la compression des images par tranformée en ondelettes. Quelques résultats sont présentés à la fin de ce chapitre et dans l'article donné en annexe C.

4.2.3 Le cas elliptique : les séries *thêta modifiées*

Ces travaux on été réalisés durant la thèse de Jean-Marie Moureaux (1994) [370] que j'ai co-encadré en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Une partie de nos travaux a été publiée dans la revue IEE Electronics Letters en juillet 1995 : "Counting lattice points on ellipsoïds : Application to image coding" [4].

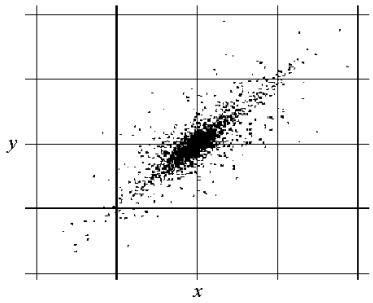


FIG. 4.1 – Exemple de distribution des coefficients d'ondelettes d'un vecteur de dimension $n = 2$. Sous-bande hg (coefficients horizontaux) à la résolution 2^{-1} issue de la transformée en ondelettes dyadique de l'image BUREAU.

4.2.3.1 Pourquoi ?

Les sources avec mémoire non stationnaires de distribution Gaussienne présentent aussi un grand intérêt en traitement des images. En effet, dans le cadre de la compression par transformée en ondelettes, dans certaines sous-bandes les coefficients d'ondelettes présentent la particularité d'être non stationnaires et très corrélés [4], [370]. Ainsi, les vecteurs n-dimensionnels constitués par des coefficients voisins sont distribués de façon elliptique autour de zéro dans une direction privilégiée (cf. figure 4.1). Une QVA elliptique sera donc plus adaptée à ce type de distribution. Nous avons donc proposé une solution au problème du dénombrement pour des distributions elliptiques et pour différents types de réseaux en introduisant les *séries thêta modifiées*.

4.2.3.2 Rappel sur les séries *thêta*

Il est important de rappeler certains résultats sur les réseaux réguliers de points, de façon à faciliter la compréhension de la construction des séries *thêta modifées*. Nous allons notamment rappeler la définition du poids de Hamming, des constructions A et B ainsi que des fonctions *thêta* de Jacobi.

Le poids de Hamming Il est utile pour calculer des séries *thêta modifées* de considérer des réseaux de points construit à partir de codes algébriques. Pour plus de détails se référer à [346]. Soit $F_2 = \{0, 1\}$ un ensemble de deux éléments. Un code binaire linéaire C composé de mots de code N-dimensionnels est un sous espace de F_2^N. Le poids d'un mot de code noté $|c|$ correspond au nombre de "uns" qu'il contient. Il est alors possible de définir le poids de Hamming

associé au code C par le polynôme :

$$W_C(x,y) = \sum_{c \in C} x^{n-|c|} y^{|c|} = \sum_{i=0}^{N} A_i x^{N-i} y^i \tag{4.8}$$

où A_i est le nombre de mots de code à la distance de Hamming i du mot de code $c \in C$. Ce poids classe les mots de code en fonction du nombre de valeurs non nulles qu'ils contiennent. La variable x compte le nombre de "zéros" et la variable y le nombre de "uns".

Les constructions *A* et *B* Les fonctions *thêta modifées* sont définies en utilisant le poids de Hamming. Ce poids est associé au réseau régulier de points en fonction du type de construction du réseau. Les constructions A et B, introduites par Leech et Sloane en 1970 [346], sont deux façons possibles pour définir des réseaux. Le lecteur intéressé peut se référer à [346] pour plus de précisions. Rappelons simplement ici ces deux constructions.

- *Construction A.* C'est la façon la plus simple pour associer un réseau à un code C. La construction $A(C)$ est définie par un sous-réseau de $\mathbb{Z}^n$ de vecteurs congrus à c modulo 2 :

$$A(C) = \{\mathbf{y} \in \mathbb{Z}^n / \exists c \in C, \mathbf{y} \equiv c \bmod (2)\} \tag{4.9}$$

- *Construction B.* Soit C un code binaire linéaire tel que ses poids sont des multiples de 4. En considérant ce code, la construction $B(C)$ est définie par :

$$B(C) = \left\{ \mathbf{y} \in \mathbb{Z}^n / \exists c \in C, \mathbf{y} \equiv c \bmod (2) ; \sum_{i=1}^{n} y_i \equiv 0 \bmod (4) \right\} \tag{4.10}$$

Les fonctions *thêta* de Jacobi Les fonctions *thêta* de Jacobi (notées θ) sont définies par [346] :

$$\theta_2(q) = \sum_{m=-\infty}^{m=+\infty} q^{\left(m+\frac{1}{2}\right)^2}, \quad \theta_3(q) = \sum_{m=-\infty}^{m=+\infty} q^{m^2}, \quad \theta_4(q) = \sum_{m=-\infty}^{m=+\infty} (-q)^{m^2}.$$

Nous les introduisons ici pour simplifier les notations dans les calculs. Il est facile de montrer que la série *thêta* du réseau $\mathbb{Z}$ est donnée par :

$$\Theta_{\mathbb{Z}}(q) = \theta_3(q), \tag{4.11}$$

et que,

$$\theta_3(q) = \theta_3\left(q^4\right) + \theta_2\left(q^4\right) \text{ et } \theta_4(q) = \theta_3\left(q^4\right) - \theta_2\left(q^4\right). \tag{4.12}$$

4.2.3.3 Définition des séries *thêta modifiées*

Soit Π une ellipse n-dimensionnelle centrée en zéro telle que :

$$\Pi = \left\{ \mathbf{y} \in \mathbb{R}^n / \left(\|\mathbf{y}\|_2^2 \right)_\Gamma = m^2 \right\}, \tag{4.13}$$

où m est une constante positive, $(\|\mathbf{y}\|_2^2)_\Gamma = \mathbf{y}\Gamma\mathbf{y}^t$ est la norme L_2 pondérée avec Γ une matrice diagonale de coefficients γ_{ij} tels que $\gamma_{ii} = 1/a_i^2$, $a_i > 0$ et $\gamma_{ij} = 0$ pour tout $i \neq j$. Ainsi, pour un réseau Λ, le nombre N_m de vecteurs situés à la distance m de l'origine (pour une métrique en norme L_2 pondérée) est donné par :

$$\begin{aligned} \Theta_\Lambda^\Gamma(q) &= \sum_{\mathbf{y}\in\Lambda} q^{\left(\|\mathbf{y}\|_2^2\right)_\Gamma} \\ &= \sum_{m=0}^{+\infty} \left[q^{m^2} \right] \left| \left\{ \mathbf{y} \in \Lambda / \left(\|\mathbf{y}\|_2^2 \right)_\Gamma = m^2 \right\} \right|. \end{aligned} \tag{4.14}$$

Le coefficient de q^{m^2} représente le nombre de vecteurs qui sont localisés *sur* l'ellipse de rayon m. Comme nous l'avons vu au paragraphe 4.2.1, Conway et Sloane ont résolu ce problème pour Γ égale à la matrice identité ($a_i = 1$, $\forall i$) et ils ont proposé une méthode pour compter les vecteurs à la surface de sphères [346]. Nous avons proposé une solution à ce problème pour tout $a_i > 0$ permettant de compter des points sur des ellipses.

4.2.3.4 Le dénombrement sur des ellipses

Série *thêta* et polynôme énumératif Deux théorèmes importants (théorèmes 3 et 15, Chapitre 7 de [346]) nous permettent de définir la série *thêta* d'un réseau Λ à partir du polynôme énumératif W_C. Cette définition dépend du type de contruction du réseau (A ou B) et du dénombrement sur les réseaux $2\mathbb{Z}$ et $2\mathbb{Z}+1$. En remarquant que

$$\begin{cases} \Theta_{2\mathbb{Z}}(q) = \theta_3\left(q^4\right) \\ \Theta_{2\mathbb{Z}+1}(q) = \theta_2\left(q^4\right) \end{cases} \tag{4.15}$$

on définit la série $\Theta_\Lambda(q)$ pour la construction A par la relation [4], [370]

$$\Theta_\Lambda(q) = W_C\left(\theta_3\left(q^4\right), \theta_2\left(q^4\right)\right). \tag{4.16}$$

De même, pour la construction B il est possible d'écrire [4], [370] :

$$\Theta_\Lambda(q) = \frac{1}{2} W_C\left(\theta_3\left(q^4\right), \theta_2\left(q^4\right)\right) + \frac{1}{2}\left[\theta_4\left(q^4\right)\right]^n. \tag{4.17}$$

Cas de la norme L_2 pondérée Dans le cas d'une norme L_2 pondérée par Γ, nous avons introduit la fonction *thêta* de Jacobi pondérée. De façon générale, on définit ces fonctions par[3] :

$$\theta_j^a(q) = \theta_j\left(q^{1/a}\right), \qquad a \neq 0. \tag{4.18}$$

En utilisant cette notation il est facile de montrer que, pour un réseau Λ donné et en considérant une norme L_2 pondérée, la relation (4.16) pour la construction A devient :

$$\Theta_\Lambda^\Gamma(q) = W_C\left(\theta_3^a\left(q^4\right), \theta_2^a\left(q^4\right)\right) \tag{4.19}$$

et la relation (4.17) pour la construction B :

$$\Theta_\Lambda^\Gamma(q) = \frac{1}{2} W_C\left(\theta_3^a\left(q^4\right), \theta_2^a\left(q^4\right)\right) + \frac{1}{2}\left[\theta_4^a\left(q^4\right)\right]^n. \tag{4.20}$$

Les séries *thêta* modifiées Le calcul de ces fonctions $\Theta_\Lambda^\Gamma(q)$ nécessite la connaissance des polynômes W_C dans le cas d'une norme L_2 pondérée. Nous pouvons montrer facilement que pour les réseaux $\mathbb{Z}^n$ et D_n construits à partir de la construction A, ces polynômes sont donnés par [4], [370] :

$$W_{C(\mathbb{Z}^n)}(x,y) = \prod_{i=1}^{n}(x_i + y_i), \tag{4.21}$$

$$W_{C(D_n)}(x,y) = \frac{1}{2}\left[\prod_{i=1}^{n}(x_i + y_i) + \prod_{i=1}^{n}(x_i - y_i)\right].$$

De plus, dans le cas du réseau E_8 et en utilisant la construction B normalisée par $1/2$ le polynôme W_C correspond à :

$$W_{C(2E_8)}(x,y) = \prod_{i=1}^{8} x_i + \prod_{i=1}^{8} y_i. \tag{4.22}$$

Nous avons maintenant toutes les données pour calculer les séries *thêta modifiées*. En effet, à partir des équations (4.12), (4.19) et (4.21), nous pouvons définir ces séries pour les réseaux $\mathbb{Z}^n$ et D_n en fonction de $\theta_j^{a_i}(q)$ donné à l'équation (4.18) par :

$$\Theta_{\mathbb{Z}^n}^\Gamma(q) = \prod_{i=1}^{n}\left(\theta_3^{a_i}\left(q^4\right) + \theta_2^{a_i}\left(q^4\right)\right) = \prod_{i=1}^{n}\theta_3^{a_i}(q) \tag{4.23}$$

[3] Notons que les équations (4.12) sont de façon évidente aussi vérifiées par ces fonctions.

et,

$$\Theta^{\Gamma}_{D_n}(q) = \frac{1}{2}\left[\prod_{i=1}^{n}\left(\theta_3^{a_i}\left(q^4\right)+\theta_2^{a_i}\left(q^4\right)\right)+\prod_{i=1}^{n}\left(\theta_3^{a_i}\left(q^4\right)-\theta_2^{a_i}\left(q^4\right)\right)\right] \quad (4.24)$$
$$= \frac{1}{2}\left[\prod_{i=1}^{n}\theta_3^{a_i}(q)+\prod_{i=1}^{n}\theta_4^{a_i}(q)\right]$$

De même, à partir des équations (4.12), (4.20) et (4.22), la série *thêta modifiée* du réseau E_8 est donnée par :

$$\Theta^{\Gamma}_{2E_8}(q) = \frac{1}{2}\left[\prod_{i=1}^{8}\theta_2^{a_i}\left(q^4\right)+\prod_{i=1}^{8}\theta_3^{a_i}\left(q^4\right)+\prod_{i=1}^{8}\theta_4^{a_i}\left(q^4\right)\right] \quad (4.25)$$

ou encore,

$$\Theta^{\Gamma}_{E_8}(q) = \frac{1}{2}\left[\prod_{i=1}^{8}\theta_2^{a_i}(q)+\prod_{i=1}^{8}\theta_3^{a_i}(q)+\prod_{i=1}^{8}\theta_4^{a_i}(q)\right]. \quad (4.26)$$

Les séries (4.23), (4.24) et (4.26) permettent respectivement de calculer le nombre de vecteurs d'un réseau $\mathbb{Z}^n$, D_n et E_8 contenus à la surface d'ellipses de demi-axes $a_i m$ pour $i \in [1, ..., n]$. Ainsi, pour compter des vecteurs sur une surface elliptique de rayon m il faut développer la série $\Theta^{\Gamma}_{\Lambda}(q)$ jusqu'à l'ordre m^2.

Remarque : *Les formules sont données pour des ellipses non orientées.*

4.2.3.5 Exemple de dénombrement sur un réseau D_4

Dans le cas d'une ellipse de paramètres $a_1^2 = 0,5$, $a_2^2 = 0,0625$, $a_3^2 = 0,25$ et $a_4^2 = 0,125$, la série *thêta modifiée* du réseau D_4 est donnée par un développement de la série (4.24) :

$$\Theta^{\Gamma}_{D_4}(q) = 1 + 4q^6 + 2q^8 + 4q^{10} + 4q^{12} + 2q^{16} + ...$$

Il y a donc 1 vecteur sur l'ellipse $m^2 = 0$, 4 vecteurs sur l'ellipse $m^2 = 6$, 2 vecteurs sur l'ellipse $m^2 = 8$ etc.

Une comparaison entre les modèles de distribution elliptique et sphérique est donné figure 4.2. Le gain apporté en QVA par un modèle mieux adapté à la distribution réelle est dans ce cas de 1,6 dB en terme de Rapport Signal-à-Bruit pic (PSNR).

FIG. 4.2 – Exemple d'image codée/décodée avec une QVA à 0,8 bpp. Image du haut : originale. Image du milieu : modèle de distribution elliptique - PSNR=27,7 dB. Image du bas : modèle de distribution sphérique - PSNR=26,1 dB.

4.3 L'indexage pour des distributions de type Gaussienne généralisée

J'ai réalisé ces travaux en collaboration avec J.M. Moureaux Maître de Conférences au CRAN - Université de Nancy 1 et P. Loyer ingénieur chez Alcatel à Cannes. Nos travaux ont été publiés dans la revue IEEE Transactions on Communications en 1998 : "Low Complexity Indexing Method for $\mathbb{Z}^n$ and D_n Lattice Quantizers" et dans la revue IEEE Transactions on Information Theory en février 2003 : "Lattice codebook enumeration for generalized Gaussian source" [13]. Ce deuxième article est donné en annexe D du document.

4.3.1 Problématique

De nombreux travaux ont été développés pour *dénombrer* ou *indexer* des distributions Gaussienne ou Laplacienne. Nous pouvons citer par exemple les travaux de [346], [350], [50], [6], [3], [4] mais aucun d'entre eux traite directement des distributions de type Gaussienne généralisée avec des paramètres différents de 1 et 2. En effet, il n'existe que quelques travaux sur le problème de dénombrement dans le cas Gaussien généralisé comme ceux de Laroia et Farvardin [366] ou encore de Chen et Villasenor [342]. Notre principale contribution a été de proposer un algorithme d'indexage pour p compris dans l'intervalle $0 < p \leq 2$ ayant une faible complexité même pour des dictionnaires de grande taille. L'indexage consiste à assigner à chaque vecteur du réseau un index unique qui permet d'identifier le vecteur. Cette opération est très importante pour des applications de compression car elle permet de retrouver au décodeur les vecteurs du réseau qui ont été utilisés lors de la quantification. La méthode que nous avons proposé est basée sur une interprétation géométrique du réseau et sur l'utilisation des séries *thêta* (ou des séries *Nu* dans le cas Laplacien). L'introduction des séries *thêta* offre de nombreux avantages et principalement elle permet de réduire la complexité de l'algorithme dans les cas $p = 1$ et $p = 2$ grâce à l'utilisation de produits de convolution. Nous avons développé l'approche dans le cas du réseau $\mathbb{Z}^n$, mais une extension à d'autres réseaux comme le réseau D_n est possible.

4.3.2 Les fonctions *thêta généralisées*

La fonction *thêta généralisée* du réseau cubique $\mathbb{Z}^n$ est donnée par la relation :

$$\theta_n(z) = \sum_r \#S_n(r) z^{r^p}, \tag{4.27}$$

où $\#S_n(r)$ donne le nombre de vecteurs[4] situés sur la L_p-sphère de rayon r définie par :

$$S_k(r) = \left\{ \mathbf{y} / \left\| \mathbf{y} \right\|_p = r \right\}, \text{ avec } k \leq n. \tag{4.28}$$

On a en particulier $\#S(0) = 1$. De façon équivalente, il est possible d'écrire :

$$\#S_n(r) = \left[z^{r^p} \right] \theta_n(z). \tag{4.29}$$

Puisque les coordonnées des vecteurs dans un réseau $\mathbb{Z}^n$ sont indépendantes, les séries *thêta généralisées* possèdent la propriété suivante :

$$\theta_{n+1}(z) = \theta_n(z)\theta_1(z).$$

Cette formule récursive permet de calculer toutes les séries *thêta* à partir de la série $\theta_1(z)$. Il est important de noter que dans les cas Gaussien et Laplacien, les exposants r^p dans la série sont entiers et donc, il est possible de calculer $\theta_{n+1}(z)$ par un *produit de convolution* entre $\theta_1(z)$ et $\theta_n(z)$ [369]. Cette propriété est essentielle pour réduire la complexité de l'algorithme, principalement sur des architectures à base de DSP où les produits de convolution sont faciles à implémenter et ne nécessitent que quelques opérations arithmétiques.

4.3.3 Principe de la méthode proposée

L'idée est basée sur la remarque que tout hyperplan perpendiculaire au premier axe $x_1 = x$ avec $|x| \leq r$ coupe la L_p-sphère $S_n(r)$ par une L_p-sphère de dimension $n-1$ et de rayon $(r^p - |x|^p)^{1/p}$ centrée sur le vecteur nul de dimension $n-1$. Un exemple donné à la figure 4.3 illustre ce principe. Le dénombrement et l'indexation consistent alors à compter les vecteurs "avant"[5] un vecteur M donné qui appartient à la L_p-sphère $S_n(r)$. Ainsi :

- Pour les vecteurs tels que $x_1 = m_1$, nous sommes ramenés au problème initial, c'est-à-dire compter les vecteurs "avant" $M_2\,(m_2, ..., m_n)$ en dimension $n-1$ avec M_2 situé sur la L_p-sphère $S_{n-1}((r^p - |m_1|^p)^{1/p})$;
- Pour les vecteurs tels que $x_1 < m_1$, il faut compter le nombre de vecteurs sur toutes les surfaces de dimension $n-1$ "avant" M en faisant la somme des coefficients $\#S_{n-1}(r)$ de la série $\theta_{n-1}(z)$:

$$\sum_{x<m_1} \#S_{n-1}(\sqrt[p]{r^p - |x|^p}) \tag{4.30}$$

[4] Si l'on reprend la notation utilisée dans les paragraphes précédents et pour des énergies entières m^p, on peut écrire $\#S_n(m) = N_m$ pour la dimension n.

[5] Un vecteur du réseau $M\,(m_1, m_2, ..., m_n)$ est dit "avant" $M'\,(m'_1, m'_2, ..., m'_n)$ si :
- $m_1 < m'_1$ ou
- $m_1 = m'_1$ et $m_2 < m'_2$ ou
- $m_1 = m'_1$ et $m_2 = m'_2$ et $m_3 < m'_3$ etc.

Le nombre de vecteurs situés sur la L_p-sphère de rayon r est donc donné par la relation :

$$\#S_n(r) = \sum_{x_1=-\lfloor r \rfloor}^{\lfloor r \rfloor} \#S_{n-1}(\sqrt[p]{r^p - |x_1|^p}). \tag{4.31}$$

Remarque : *Dans le cas particulier p=1, la formule (4.31) nous permet de retrouver les équations de Fischer et Pan [350] :*

$$\#S_n(r) = \#S_{n-1}(r) + 2(\#S_{n-1}(r-1) + ... + \#S_{n-1}(1) + 1). \tag{4.32}$$

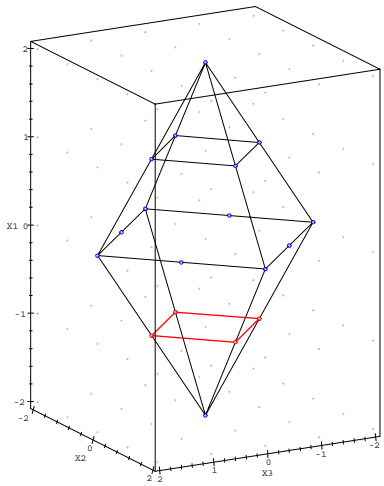


FIG. 4.3 – Interprétation géométrique du principe de la méthode dans le cas de la norme L_1. Le plan auquel appartient la pyramide *rouge* (de dimension 2 et de rayon 1 centrée sur le vecteur (0,0)) coupe la surface de la pyramide de dimension 3. L'intersection, c'est-à-dire le nombre de points en surface de la pyramide rouge, donne le nombre de points sur la pyramide en dimension 3 pour la coordonnée $x_1 = -1$. Faire varier x_1 de -2 à 2 et compter le nombre de points sur chaque pyramide correspondante en dimension 2, donne ici le nombre total de points sur la surface de la pyramide en dimension 3 de rayon $r = 2$.

4.3.4 Algorithme d'indexage proposé

L'algorithme que nous avons développé ainsi que l'algorithme inverse qui permet d'obtenir le vecteur à partir de l'index sont présentés en détail dans l'article "Lattice codebook enumeration for generalized Gaussian source" [13] donné en annexe D de ce document. Nous indiquons ici les grandes lignes de l'indexage. Pour cela, on note :

- M_k le vecteur $M_k\,(m_k, ..., m_n)$;
- $Index(k)$ la fonction qui calcule l'index du vecteur M_k ;
- $Avant(M_k)$ la fonction qui calcule le nombre de vecteurs tels que $x_k < m_k$.

D'après ce que nous avons dit au paragraphe 4.3.3 précédent, il est possible d'écrire la relation

$$Index(M_1) = Index(M_2) + Avant(M_1), \tag{4.33}$$

et de façon récursive, l'index du vecteur $M_1\,(m_1, ..., m_n)$ de dimension n est donné par :

$$Index(M_1) = Index(M_n) + \sum_{k=1}^{n-1} Avant(M_k), \tag{4.34}$$

qui est la formule clé de notre algorithme. En effet, tout vecteur situé sur une L_p-sphère de rayon r (pour $0 < p \leq 2$) peut recevoir un index au moyen de cette formule. Cet algorithme est inversible, c'est-à-dire qu'au moyen d'un index il est possible de retrouver de façon unique au décodeur le vecteur du réseau correspondant. La complexité de cet algorithme est étudiée en détail dans l'article donné en annexe D.

4.4 Quantification vectorielle et compression

Ces travaux on été réalisés durant la thèse de Philippe Raffy (1997) [429] que j'ai co-encadré en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Une partie de nos travaux a été publiée dans la revue IEEE Transactions on Image Processing en 2000 : "Distortion-rate models for entropy coded lattice vector quantization" [11]. Cet article est donné en annexe E du document.

4.4.1 Conception d'un QVA

4.4.1.1 Les différentes étapes

Comme nous l'avons vu précédemment, la compression qui utilise la QVA présente de nombreux avantages. Cependant, sa mise œuvre n'est pas immé-

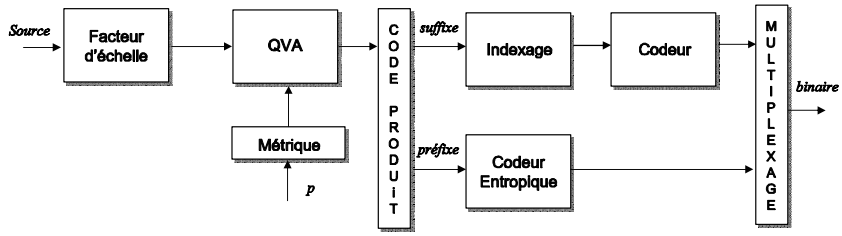


FIG. 4.4 – Schéma général de principe d'un QVA.

diate. En effet, elle soulève un certain nombre de questions concrètes notamment en termes de calcul et de stockage. De façon générale et pour une source donnée, on peut définir la QVA en cinq étapes :

1. Choix d'un réseau et d'une métrique L_p ;
2. Normalisation de la source ;
3. Quantification de la source normalisée au moyen d'algorithmes rapides ;
4. Indexage des vecteurs quantifiés ;
5. Codage des index.

Le choix d'une métrique L_p dépend de la distribution statistique de la source. Un choix judicieux est important car c'est à partir du type de métrique que l'on va définir les méthodes de dénombrement (séries *thêta*, séries *Nu* ou autres) et d'indexage. La normalisation de la source correspond de manière équivalente à une "mise à l'échelle" du réseau régulier. Elle conditionne le bruit granulaire généré par le réseau car elle agit directement sur le moment d'ordre 2 du réseau [346]. Le choix du facteur de normalisation se fait de façon identique à celui du pas de quantification, c'est-à-dire par une méthode d'allocation de débits. L'indexage est une opération indépendante de la quantification. C'est une opération nécessaire et importante qui doit permettre un décodage unique.

4.4.1.2 Utilisation d'un code préfixe

Schématiquement, après l'étape de quantification, chaque vecteur quantifié appartient au réseau régulier et est localisé sur une surface centrée en zéro[6] de rayon donné. Les vecteurs qui appartiennent à une même surface possèdent donc la même énergie. La structure du réseau permet alors d'assigner un *code*

[6] Nous nous plaçons dans le cas de source de moyenne nulle : les hypersphères définies dans le réseau sont concentriques et centrées en zéro.

préfixe à chacun des vecteurs du réseau. Ainsi, un vecteur $\mathbf{y}$ du réseau peut être défini par un couple (m, i) où :

- m est le préfixe du code associé à $\mathbf{y}$. Il correspond à l'index du rayon r de la L_p-sphère ;
- i est le suffixe du code associé à $\mathbf{y}$. Il correspond à la position de $\mathbf{y}$ sur l'hyper-sphère de rayon r. Cet index de position est donné par un algorithme d'indexage (cf. paragraphe 4.3).

Un schéma général de QVA est donné sur la figure 4.4. Il a été montré par Fischer que les codes préfixes sont bien adaptés à la structure des réseaux [349]. Dans ce contexte, le code binaire associé à un vecteur est donné par la concaténation du code binaire associé au rayon (préfixe) avec le code binaire associé à l'index de position (suffixe). Le codage du préfixe est une opération relativement simple qui peut être faite en utilisant un codeur entropique du type codeur arithmétique. En effet, le nombre de rayons différents mis en jeux lors de la quantification reste relativement faible et donc ce type de codeur est très performant. Le codage du suffixe peut se faire de deux façons différentes. Soit au moyen d'un code à longueur fixe en supposant que tous les vecteurs qui appartiennent à une surface de rayon constant sont équiprobables (ce qui reste vrai uniquement dans un cas asymptotique), soit en ne considérant pas l'équiprobabilité ce qui est plus réaliste dans la pratique. Dans ce second cas, on peut mettre en œuvre un codage entropique pour la position.

4.4.2 Modélisations de la distorsion et du débit

4.4.2.1 Problématique

Un point important en QVA est le réglage du "facteur d'échelle" ou pas de quantification. En effet, le pas de quantification agit directement sur le volume des cellules de quantification [376], [362] et donc sur la distorsion générée par le quantificateur. Les travaux effectués sur l'estimation de la distorsion dans le cas vectoriel sont généralement basés sur la théorie asymptotique et ont introduit des bornes minimales et maximales. Ils donnent une bonne idée au niveau de la performance des quantificateurs mais demeurent généralement inadaptés dans les cas non asymptotiques ou bas débits. De plus, la connaissance de la fonction débit-distorsion d'un quantificateur vectoriel ne donne pas directement une façon de faire pour optimiser ses paramètres (par exemple le facteur d'échelle pour un QVA). Un état de l'art complet sur la quantification se trouve dans l'article de Gray et Neuhoff [359].

Contrairement aux approches existantes basées sur l'hypothèse de haute résolution (hauts débits), nous avons donné une formulation n-dimensionnelle de la distorsion non asymptotique développée par Sripad et Snyder dans [375] pour le cas scalaire. Nous avons aussi proposé un modèle non asymptotique

pour le débit dans le cas de la QVA et pour différents réseaux ($\mathbb{Z}^n$, D_n, E_n). Les formules que nous avons développées sont basées sur l'estimation d'un modèle pour la distribution de la source et peuvent facilement s'étendre au cas des mixtures.

4.4.2.2 Cas de la distorsion non asymptotique pour un réseau $\mathbb{Z}^n$

Le modèle de distorsion que nous avons proposé se base sur les travaux de [372], [343], [377], [375]. Cette approche "exacte" a aussi été étudiée dans le cadre des modulateurs Delta et Delta-Sigma [361], [341], [357].

A partir des travaux de Clavier [343] et de Widrow [377], nous avons donné une expression analytique de la distorsion granulaire pour un réseau régulier $\mathbb{Z}^n$. Cette expression correspond à la version vectorielle de la formule de distorsion donnée par les travaux de Sripad et Snyder [375]. Elle dépend de la *fonction caractéristique conjointe* $\Phi_{X_1 \ldots X_n}$ de la source [379], [340], [363]. Nous la présentons ici sous la forme de la proposition 4 suivante.

Proposition 4 *Soit* $\mathbf{x} = (x_1, ..., x_n)$ *un vecteur aléatoire de dimension n ayant pour densité de probabilité conjointe* $f_{X_1,...,X_n}(x_1, ..., x_n)$ *et pour fonction caractéristique marginale* $\Phi_{X_i}(u_i)$*. Alors, la distorsion granulaire exprimée par vecteur issue de la QVA sur un réseau cubique* $\mathbb{Z}^n$ *avec une mesure de distorsion donnée par l'EQM et un pas de quantification q est :*

$$D_g(q) = \frac{q^2}{12} \sum_{i=1}^{n} \left[1 + \frac{6}{\pi^2} \sum_{k \neq 0} \frac{(-1)^k}{k^2} \Phi_{X_i} \left(\frac{2\pi k}{q} \right) \right].$$

La preuve de cette proposition peut être établie immédiatement à partir des travaux de [375]. Le modèle proposé ne nécessite aucune hypothèse sur la distribution de la source mais seulement la connaissance de la fonction caractéristique marginale donnée par $\Phi_{X_i}(u_i) = \Phi_{X_1,...,X_i,...,X_n}(0, ..., u_i, ..., 0)$. Il est important de noter que la proposition 4 est aussi vraie pour des sources non stationnaires, et c'est là tout son intérêt[7] ! Nous avons étudié la validité de ce modèle durant la thèse de Raffy [429]. Des résultats sont présentés dans l'article "Distortion-rate models for entropy coded lattice vector quantization" [11] donné en annexe E. La courbe 4.5 suivante permet de valider le modèle proposé dans le cas d'une source non stationnaire et une QVA effectuée sur un réseau $\mathbb{Z}^2$. Ici, la source est une séquence i.i.d. de vecteurs alétoires (x, y) de

[7] Un développement plus poussé de cette formule pourrait permettre de donner une estimation de la distorsion dans le cas où un pas de quantification différent est appliqué sur chaque composante des vecteurs sources, exploitant ainsi pleinement la non-stationnarité de la source.

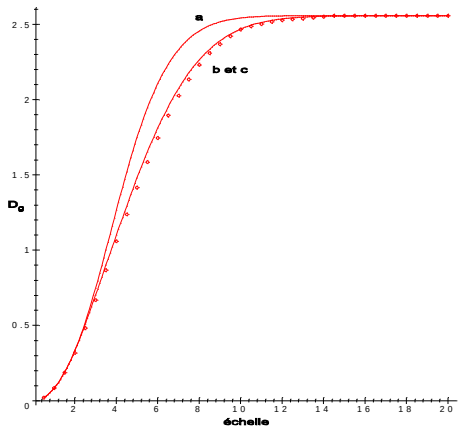


FIG. 4.5 – Source synthétique elliptique centrée ($\sigma_x = 1$ et $\sigma_y = 2$) et QVA sur un réseau $\mathbb{Z}^2$. (a) modèle de distorsion de Widrow ; (b) modèle de distorsion proposé ; (c) distorsion expérimentale (points).

distribution Gaussienne ; la moyenne est $(0, 0)$ et la matrice de covariance est donnée par K avec $K_{11} = \sigma_x^2 = 1$ et $K_{22} = \sigma_y^2 = 4$ et $K_{12} = K_{21} = 0$. On peut noter que pour des statistiques elliptiques, l'extension proposée donne de meilleurs résultats comparativement aux méthodes existantes [362], [377].

4.4.2.3 Cas du débit non asymptotique pour un réseau $\mathbb{Z}^n$

Contexte A notre connaissance, l'estimation non asymptotique du débit (bas débits) pour un QVA à débit variable a été peu investiguée et peu de travaux peuvent être trouvés dans la littérature. Citons entre autres nos travaux [47] et ceux de Kim, Hu et Nguyen [364]. Ces derniers traitent le cas d'une source de distribution Laplacienne. Nous avons proposé un modèle de débit non asymptotique pour des *codes préfixes* valides pour le réseau $\mathbb{Z}^n$ et quelle que soit la distribution de la source. Nous avons aussi proposé une extension géométrique de ce modèle valide quel que soit le réseau Λ mais uniquement pour des distributions de type Gaussienne. Le détail de la démonstration pour ce deuxième cas de figure se trouve dans la thèse de Raffy [429] et dans l'article "Distortion-rate models for entropy coded lattice vector quantization" [11] donné en annexe E du document. Nous nous focalisons ici sur le cas du réseau $\mathbb{Z}^n$.

Définition du code préfixe Un code préfixe est construit en concaténant le mot de code utilisé pour l'énergie r^p du vecteur avec le mot de code utilisé pour sa position sur l'hyper-sphère d'énergie r^p. La longueur d'un mot de code

dépend donc de la structure du réseau ainsi que de la métrique L_p choisie pour le dénombrement[8]. Nos travaux ont permi de proposer une modélisation de la longueur moyenne des mots de code du code préfixe en partant de l'expression analytique de l'entropie des vecteurs du réseau $\mathbf{y} = Q(\mathbf{x})$ utilisés pour la quantification d'une source vectorielle $\mathbf{x}$ donnée. Soit l'entropie conditionnelle des vecteurs $\mathbf{y}$ sachant le rayon r donnée par la relation [358] :

$$H(\mathbf{y}|r) = -\sum_i \sum_j p(r_i) p(\mathbf{y}_j|r_i) \log_2 p(\mathbf{y}_j|r_i), \tag{4.35}$$

où $p(\mathbf{y}_j|r_i) = \Pr\{\mathbf{y} = \mathbf{y}_j / r = r_i\}$ est la probabilité qu'un vecteur source $\mathbf{x}$ soit quantifié par le vecteur $\mathbf{y}_j$ du réseau appartenant à la surface de rayon $r = r_i$ (le rayon r est considéré comme une variable aléatoire de probabilité $p(r_i) = \Pr\{r = r_i\}\}$). Notons que la fonction $r(\mathbf{y}_j)$ qui donne le rayon r_i en fonction d'un vecteur $\mathbf{y}_j$ du réseau est une fonction déterministe de $\mathbf{y}_j$ avec $p(\mathbf{y}_j|r_i) = 0$ quand $r_i \neq r(\mathbf{y}_j)$.

Avec ces notations et en introduisant l'entropie $H(r)$ du rayon r cette entropie conditionnelle devient [358] :

$$H(\mathbf{y}|r) = H(\mathbf{y}, r) - H(r) \tag{4.36}$$

avec $H(r) = -\sum_i p(r_i) \log_2 p(r_i)$ et $H(\mathbf{y}, r) = H(\mathbf{y})$ puisque r est une fonction déterministe de $\mathbf{y}$. Il est alors possible d'écrire l'entropie des vecteurs de quantification du réseau sous la forme :

$$H(\mathbf{y}) = -\sum_i p(r_i) \left\{ \log_2 p(r_i) + \sum_j p(\mathbf{y}_j|r_i) \log_2 p(\mathbf{y}_j|r_i) \right\}. \tag{4.37}$$

Le premier terme de cette formule est l'entropie des rayons ou surfaces utilisés par la quantification. Le second terme est l'entropie des vecteurs de quantification situés sur une surface de rayon r_i donnée. Le modèle de débit que nous avons proposé se déduit directement de l'entropie $H(\mathbf{y})$ donnée par la formule (4.37). Il suppose donc la connaissance des probabilités $p(r_i)$ et des probabilités conditionnelles $p(\mathbf{y}|r)$.

Approximation du débit Pour des grandes dimensions n il est bien connu que la densité de probabilité d'une source Gaussienne ou Laplacienne est constante sur une hyper-sphère de rayon donné. Ceci signifie que les vecteurs source sont uniformément distribués sur la surface de l'hyper-sphère. Il est alors raisonnable de considérer que la probabilité qu'un vecteur source d'énergie r^p (métrique L_p), quantifié par un vecteur du réseau pris parmi l'ensemble des vecteurs

[8] Rappelons que le choix de la métrique dépend de la distribution de la source [349], [3], [4].

appartenant à la surface d'énergie r^p, est constante et égale à $\frac{1}{\#S_n(r)}$. La quantité $\#S_n(r)$ correspond au nombre de vecteurs du réseau d'énergie r^p donné par les séries génératrices. Cette hypothèse entraîne que $H(\mathbf{y}|r)$ est maximum. Dans ce cas, la formule (4.37) se simplifie et nous pouvons approximer le débit granulaire par :

$$R_g = -\sum_{i=0}^{+\infty} p(r_i) \log_2 p(r_i) + \sum_{i=1}^{+\infty} p(r_i) \log_2 \#S_n(r_i). \tag{4.38}$$

Si l'on introduit le fait que le réseau est "mis à l'échelle" au moyen d'un pas de quantification q (estimé par une allocation des ressources binaires), l'énergie des vecteurs du réseau est multiplié par q^p et la formule (4.38) peut s'écrire sous la forme :

$$R_g(q) = -\sum_{i=0}^{+\infty} p(qr_i) \left[\log_2 p(qr_i) - \log_2 \#S_n(r_i)\right]. \tag{4.39}$$

Modélisation de la loi du rayon $p(r_i)$ La formule (4.39) nécessite la connaissance de la loi de distribution du rayon qui dépend du modèle de distribution de la source et donc de la métrique L_p utilisée. Nous avons proposé une modélisation de cette loi pour des réseaux $\mathbb{Z}^n$ valide à la fois pour des bas et des forts débits. Cette approximation se base sur la connaissance d'un modèle de la densité de probabilité conjointe $f_{X_1,...,X_n}(x_1,...,x_n)$ de la source. Elle est donné ici sous la forme de la proposition 5 et sa démonstration se trouve dans la thèse de Raffy [429] et dans l'article "Distortion-rate models for entropy coded lattice vector quantization" [11] en annexe E du document.

Proposition 5 *Soit* $\mathbf{x} = (x_1,...,x_n)$ *un vecteur aléatoire de dimension n ayant pour densité de probabilité conjointe* $f_{X_1,...,X_n}(x_1,...,x_n)$ *et soit une métrique* L_p. *La probabilité de la surface de rayon r qui résulte de la quantification de* $\mathbf{x}$ *par un QVA avec un facteur d'échelle q sur un réseau cubique* $\mathbb{Z}^n$ *est donnée par :*

$$\Pr\left\{\|\mathbf{y}\|_p = qr\right\} = \sum_{m_1=-\infty}^{+\infty} \cdots \sum_{m_n=-\infty}^{+\infty} \delta\Big(\sum_{k=1}^{n} |m_k|^p - r^p\Big) \times \int_{-q/2}^{q/2} \cdots \int_{-q/2}^{q/2} f_{X_1 \ldots X_n}(x_1 - m_1 q, \ldots, x_n - m_n q)\, dx_1 \ldots dx_n$$

avec δ *le symbol de Kronecker tel que* $\delta(u) = 1$ *si* $u = 0$ *et* 0 *sinon.*

Nous avons montré expérimentalement que ce modèle est très précis quel que soit le facteur d'échelle q. Par exemple, dans le cas d'une source Laplacienne i.i.d. synthétique et d'un réseau $\mathbb{Z}^8$ l'erreur moyenne d'approximation du débit réel est inférieure à 0,32% [11].

4.5 Conclusion-Synthèse

Nous avons développé une nouvelle méthode géométrique de codage des coefficients d'ondelettes qui utilise la quantification vectorielle par réseaux réguliers de points, connue sous le nom de quantification vectorielle algébrique (ou encore dans la littérature par le terme anglais "lattice VQ"). Cette méthode permet de limiter le coût calcul de codage à celui d'un quantificateur scalaire. Nos travaux se sont orientés suivant plusieurs objectifs et principalement, nous avons proposé des solutions pour le dénombrement des vecteurs dans un réseau régulier, pour l'indexage de ces vecteurs dans le cas d'une distribution Gaussienne généralisée et pour la modélisation de la distorsion de quantification et du débit dans un cadre non asymptotique. Ces travaux ont conduit à la définition d'un algorithme de quantification vectorielle algébrique à entropie contrainte pour la compression multirésolution des images.

4.6 Références

[340] J.A. Bucklew and G.L. Wise, "Multidimensional asymptotic quantization theory with $r^t h$ power distortion measures," *IEEE Transactions on Information Theory*, vol. 28, pp. 239–247, mars 1982.

[341] J.C. Candy and O.J. Benjamin, "The structure of quantization noise from sigma-delta modulation," *IEEE Transactions on Communications*, vol. 29, pp. 316–1323, septembre 1981.

[342] F. Chen, Z. Gao, and J. Villasenor, "Lattice vector quantization of generalized gaussian sources," *IEEE Trans. Inform. Theory*, vol. 43, pp. 92–103, janvier 1997.

[343] A.G. Clavier, P.F. Panter, and D.D. Grieg, "Distortion analysis," *Electrical Engineering*, pp. 1110–1122, novembre 1947.

[344] J.H. Conway and N.J.A. Sloane, "Fast quantizing and decoding algorithms for lattice quantizers and codes," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 227–232, mars 1982.

[345] J.H. Conway and N.J.A. Sloane, "Voronoi region of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 211–226, mars 1982.

[346] J.H. Conway and N.J.A. Sloane, *Spheres Packings, Lattices and Groups*, Springer-Verlag, 1st edition, 1988.

[347] P.C. Cosman, R. M. Gray, and M. Vetterli, "Vector quantization of image subbands : A survey," *IEEE Transactions on Image Processing*, vol. 5, no. 2, pp. 202–225, 1996.

[348] M.V. Eyuboglu and G.D. Forney, "Lattice and trellis quantization with lattice- and trellis-bounded codebooks–high-rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, no. 1, pp. 46–59, janvier 1993.

[349] T.R. Fischer, "A pyramid vector quantizer," *IEEE Trans. on Inform Theory*, vol. 32, no. 4, pp. 568–583, juillet 1986.

[350] T.R. Fischer and J. Pan, "Enumeration encoding and decoding algorithms for pyramid cubic lattice and trellis codes," *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 2056–2061, November 1995.

[351] Z. Gao, F. Chen, B. Belzer, and J. Villasenor, "A comparison of the z, e_8 and leech lattices for image subband quantization," in *IEEE Data Compression Conference, Snowbird, Etats-Unis*, mars 1995.

[352] Z. Gao, F. Chen, B. Belzer, and J. Villasenor, "A comparison of the z, e_8 and leech lattices of low-shape-parameter generalized gaussian sources," *IEEE Signal Processing Letters*, vol. 2, no. 10, pp. 197–199, octobre 1995.

[353] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. 25, pp. 373–380, 1979.

[354] A. Gersho, "On the structure of vector quantizers," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 157–166, mars 1982.

[355] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[356] J.D. Gibson and K. Sayood, "Lattice quantization," *Advances in Electronics and Electron Physics*, vol. 72, pp. 259–331, juin 1988.

[357] R.M. Gray, "Quantization noise spectra," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1220–1244, novembre 1990.

[358] R.M. Gray, *Source Coding Theory*, Kluwer Academic Publishers, 1990.

[359] R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2384, octobre 1998.

[360] R.M. Gray and T. Linder, "Mismatch in high rate entropy constrained vector quantization," *à paraître dans IEEE Transactions on Information Theory*, 2003.

[361] J.E. Iwersen, "Calculated quantizing noise of single-integration delta-modulation coders," *Bell Syst. Technical Journal, septembre 1969*, vol. 48, pp. 2359–2389, septembre 1969.

[362] D.G. Jeong and J.D. Gibson, "Uniform and piecewise uniform lattice vector quantization for memoryless gaussian and laplacian sources," *IEEE Trans. on Inform. Theory*, pp. 786–804, May 1993.

[363] J.C. Kieffer, "Stochastic stability for feedback quantization schemes," *IEEE Trans. on Inform. Theory*, vol. 28, pp. 248–254, mars 1982.

[364] W.H. Kim, Y.H. Hu, and T. Nguyen, "Joint optimization of lattice vector quantizer and entropy coder in subband coding," in *Proceedings of SPIE*, 1997, vol. 3078.

[365] C. Lamblin and J.P. Adoul, "Algorithme de quantification vectorielle sphérique à partir du réseau de gosset d'ordre 8," *Anna. Télécommun.*, vol. 43, no. 3-4, pp. 172–186, 1988.

[366] R. Laroia and N. Farvardin, "A structured fixed-rate vector quantizer derived from variable-length scalar quantizer : Part i-memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 851–867, mai 1993.

[367] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. on Comm.*, vol. 28, no. 1, pp. 84–95, 1980.

[368] T. Lookabaugh and R.M. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 1020–1033, septembre 1989.

[369] P. loyer, *Réseaux Arithmétiques et Aspects des Communications Numériques*, Ph.D. thesis, Université de Nice - Sophia Antipolis, France, 1995.

[370] J.M. Moureaux, *Quantification Vectorielle Algébrique pour la Compression d'Images. Application à l'Imagerie Radar à Synthèse d'Ouverture (SAR)*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, décembre 1994.

[371] P. Raffy, *Modélisation, Optimisation et Mise en Oeuvre de Quantificateurs Bas Débits pour la Compression d'Images Utilisant une Transformée en Ondelettes*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, décembre 1997.

[372] S.O. Rice, *Mathematical Analysis of Random Noise*, in Selected Papers on Noise and Stochastic Processes, N. Wax Eds, New York : Dover, 1954. Reprinted from Bell Syst. Tech., vol. 23, pp. 282-332, 1944 and vol. 24, pp. 46-156, 1945.

[373] T. Senoo and B. Girod, "Vector quantization for entropy coding of image subbands," *IEEE Transactions on Image Processing*, vol. 1, no. 4, pp. 526–532, octobre 1992.

[374] P. Solé, "Counting lattice points in pyramids," *Actes du Congrès Séries Formelles et Combinatoire Algébrique. Montréal, QC, Canada : LACIM*, pp. 343–355, 1992.

[375] A.B. Sripad and D.L. Snyder, "A necessary and sufficient condition for quantization errors to be uniform and white," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 25, no. 5, pp. 442–448, octobre 1977.

[376] P. Swaszek, "A vector quantizer for the laplacian source," *IEEE Transactions on Information Theory*, vol. 37, pp. 1355–1365, septembre 1991.

[377] B. Widrow, "Statistical analysis of amplitude quantized sampled data systems," .

[378] Z.M Yusof and T. Fischer, "An entropy-coded lattice vector quantizer for transform and subband image coding," *IEEE Transactions on Image Processing*, vol. 5, pp. 289–298, février 1996.

[379] P.L. Zador, "Topics in the asymptotic quantization of continuous random variables," *Bell Laboratories Technical Memorandum*, 1966.

[380] P.L. Zador, "Asymptotic quantization error of continuous signals and their quantization dimension," *IEEE Transactions on Information Theory*, vol. 28, 1982.

Chapitre 5

Le problème de décodage optimal

Dans ce chapitre je développe les activités de recherche que j'ai effectuées dans le domaine lié au décodage optimal d'images 2D et de séquences d'images (vidéos). Le plan de ce chapitre est le suivant. Tout d'abord j'introduis dans le paragraphe 5.1 le besoin d'un décodage d'images et de vidéos efficace et je mets en avant la notion de bruit borné qui caractérise le bruit de quantification ainsi que les bruits liés à la chaîne instrumentale d'acquisition. La solution que nous avons proposé pour résoudre ce problème est donnée dans le paragraphe 5.2.1 pour les images fixes. Elle est basé sur un filtrage inverse non linéaire dynamique. Nous avons étendue cette solution dans le cas des vidéos et la méthode proposée est présentée dans le paragraphe 5.2.2 suivant. Dans ce cas, nous avons introduit un critère basé sur la segmentation spatio-temporelle des sequences et sur un traitement séparé du fond et des pixels en mouvement. Enfin, le paragraphe 5.3 résume l'extension de notre approche pour le contrôle et de la réduction des erreurs liées à la transmission ou au stockage des vidéos sur des médias bruités.

5.1 Le besoin d'un décodage efficace

Les systèmes d'imagerie (caméras et appareils photo numériques, satellites) entraînent une perte de l'information due à l'optique (flou, artefacts dus aux capteurs électroniques). De plus, la capacité mémoire liée au stockage ou à la transmission étant limitée, l'utilisation d'un codeur est nécessaire afin de réduire le débit binaire. L'usage d'un codeur a pour conséquence d'engendrer la création d'artefacts nuisibles à la qualité visuelle. La suppression de ces artefacts permet un meilleur confort visuel dans la perception des données restaurées.

Les méthodes classiques de décodage et restauration procèdent en deux étapes : dans un premier temps l'image est décodée puis, afin de supprimer le flou et d'éliminer les artefacts l'image subit un post-traitement. Dans ce

document, l'objectif n'est pas de faire un état de l'art complet de ce genre de méthodes. Cependant, parmi les méthodes les plus récentes de post-traitements il est important de distinguer les approches basées sur les *Equations aux Dérivées Partielles (EDP)*. Un récapitulatif des méthodes de restauration par EDP est dressé par Deriche et Faugeras dans [392]. Parmi les techniques les plus récentes, nous pouvons citer les travaux de Charbonnier [386], [387], [388] qui introduisent une méthode de débruitage basée sur la régularisation semi-quadratique dans laquelle le problème est formulé comme une minimisation d'énergie, ainsi que les travaux de Vese et Aubert [441], [381] qui proposent des approches variationnelles pour la reconstruction d'images. On peut noter aussi les travaux de Osher et Rudin sur les filtres de choc [426]. D'autre part, Katsaggelos a présenté différents travaux sur la restauration d'images décodées : une formulation générale pour la restauration sous contraintes est donnée dans [409], et des algorithmes de restauration itératifs dans [410] et [411]. Enfin, dans [412] et [414], Kornprobst réalise une étude et une comparaison des méthodes de débruitage existantes et déduit un formalisme commun à toutes ces méthodes. Les approches stochastiques sont également diverses et variées ; les travaux de Besag [383], Geman et Geman [395] et Azencot [382] permettent de lier ces approches aux approches déterministes. De façon générale, ces approches ne prennent pas en compte les caractéristiques du système d'acquisition/transmission et du bruit qu'il engendre, ainsi que la non-linéarité du quantificateur. Les images décodées souffrent d'artefacts gênants pour les photo-interprètes, spécialement à des forts taux de compression. En effet, les méthodes basées sur les transformées à reconstruction exacte éliminent les problèmes d'aliasing uniquement dans le cas ou il n'y a pas de quantification [158], [416] et [421]. Une transformée linéaire ne peut en aucun cas prendre en compte l'aspect non linéaire du quantificateur. La quantification est un problème complexe ; c'est une transformée non linéaire dont l'erreur est difficile à modéliser. De plus, l'optimisation de cette opération au sens du meilleur compromis débit-distorsion, se fait généralement dans le domaine transformé. Ainsi, le fait d'appliquer une transformée inverse se traduit par des artefacts non contrôlés (en particuliers des oscillations sur les contours dans l'image).

Une façon de prendre en compte la non-linéarité du quantificateur consiste alors à *optimiser le filtrage de synthèse*. Ainsi, Chen et Lin [389] introduisent des filtres de Wiener dans un schéma de codage multirésolution par ondelettes de façon à restaurer des signaux fractals qui ont souffert d'une distorsion due au bruit canal. En présence de quantification, Dembo et Malah ont utilisé un modèle statistique de bruit pour la conception des filtres optimaux au sens du minimum d'erreurs quadratiques moyennes [391]. Les filtres sont obtenus à partir d'un ensemble d'équations linéaires. Le but est d'améliorer les performances des transformées multirésolutions non orthogonales pendant le codage d'images. Cependant, Moulin a montré que la faible performance de tels codeurs ne peut pas être attribuée à l'utilisation d'une transformée non orthogonale mais à l'uti-

lisation d'une structure inadaptée des codeurs [425]. Il utilise des techniques numériques pour minimiser la distorsion globale sans imposer aucune restriction aux modèles des quantificateurs. Kovacevic dans [417], [400] et Haddad dans [401], sacrifient l'orthogonalité et la reconstruction parfaite afin d'optimiser les filtres de synthèse de façon à minimiser l'erreur de reconstruction. Ces approximations utilisent des filtres d'analyse fixes et des quantificateurs selon le modèle de Lloyd-Max qui soit maximisent le SNR, soit réduisent les dépendances du signal reconstruit. En particulier, Delopoulos et Kollias développent une technique de filtrage basée sur la minimisation de l'erreur de reconstruction en connaissant les statistiques du second ordre du signal et du bruit de quantifcation [390]. Enfin, Gosse et Duhamel ont proposé un algorithme itératif d'optimisation des coefficients des filtres pour le cas monodimensionnel (codage de la parole) [397], [398]. Les filtres de synthèse sont optimisés au sens du minimum de l'erreur quadratique moyenne. Les modèles de bruit blanc et coloré sont alors utilisés de façon à étudier leurs performances. Les résultats montrent que le modèle de bruit blanc se comporte aussi bien que le modèle de bruit coloré. Cependant, ces approches ne donnent pas des gains très significatifs.

Les travaux que nous avons effectués ont pour objectif de remettre en cause les filtres linéaires à reconstruction parfaite. Nous proposons de combiner la prise en compte de la non-linéarité du quantificateur ainsi qu'une optimisation de l'opération de décodage par une approche variationnelle. Nous avons proposé un algorithme de décodage adapté à la problématique globale posée. En effet, le compresseur étant imposé, nous proposons l'optimisation du décodeur en prenant à la fois en compte les caractéristiques du bruit de quantification pour un codeur spécifié ainsi que les caractéristiques du système d'acquisition : connaissance de la Fonction de Transfert de Modulation (FTM) de l'optique ainsi que du bruit électronique des capteurs. Le principe de cette nouvelle méthode est basé sur le calcul d'un "filtre de synthèse" au décodeur, par optimisation d'un critère. Ce critère prend en compte la façon de quantifier/coder ainsi que des *a priori* sur l'image reconstruite introduits sous la forme de contraintes. Le double objectif est de réduire les artéfacts dus à l'opération de codage (effets de blocs, détérioration des contours...), ainsi que de déconvoluer et restaurer l'image acquise en prenant en compte les caractéristiques du système d'acquisition. Ces deux opérations sont effectuées conjointement et non pas successivement. Un point novateur est aussi la prise en compte dans la chaîne de traitement du bruit caractérisé par un *bruit borné non stationnaire*. Notre approche a été soit multirésolution pour le décodage optimal d'images fixes avec une application dans le cadre de l'imagerie satellitaire, soit basée sur la DCT pour le décodage optimal de vidéos codées MPEG.

5.2 La réduction des artefacts de compression

5.2.1 Dans le cadre des images fixes

Ces travaux ont été réalisés durant la thèse de Stéphane Tramini (1999) [438] que j'ai co-encadrée en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. A cette occasion, nous avons collaboré avec le Professeur G. Aubert du laboratoire J.A. Dieudonné de l'Université de Nice-Sophia Antipolis. Une partie de nos travaux a été publiée dans la revue International Journal of Imaging Systems and Technology en octobre 1998 : "Intraframe image decoding based on a nonlinear variational approach" [9].

5.2.1.1 Problème de reconstruction au décodage

Problématique Décoder une image comprimée est un réel problème. En effet, la difficulté majeure est de prendre en compte les caractéristiques du système de compression et du bruit qu'il engendre. Nous avons proposé de traiter simultanément le décodage et le débruitage sur des images ayant subi une opération de compression [56], [57], [61]. Le but de nos travaux est la suppression des artefacts liés à l'acquisition et la compression par la prise en compte de la globalité du système. Un simple filtrage linéaire ne prend pas en compte la non-linéarité de l'opération de quantification et la non-stationnarité des images. C'est pourquoi nous proposons de concevoir le décodeur en prenant en compte des *a priori* sur la solution et la connaissance du codeur (transformation et quantification), afin d'atténuer les effets dus au bruit de quantification. La solution peut être considérée comme un problème inverse avec optimisation de la structure TRANSFORMATION - QUANTIFICATION - DECODAGE formulée par une approche variationnelle. Cette nouvelle approche de décodage se traduit par un filtrage inverse non linéaire dynamique.

Notations Une image en niveaux de gris (et en variables continues) peut être représentée par la relation suivante :

$$\begin{cases} f : \Omega \subset \mathbb{R}^2 \longmapsto \mathbb{R} \\ (x, y) \longmapsto f(x, y) \end{cases} \tag{5.1}$$

où $f(x, y)$ correspond à l'intensité du pixel au site (x, y). Considérons l'opérateur "transformée" T défini par

$$\begin{cases} T : L^2(\Omega) \longmapsto L^2\left(\Omega^T\right) \\ f(x, y) \longmapsto Tf(u, v) \end{cases} \tag{5.2}$$

supposé linéaire et continu. L'image $p = Tf$ est l'image dans le domaine transformé $\Omega^T \subset \mathbb{R}^2$. Définissons par $\widehat{p}$ l'image f_o d'origine quantifiée dans le do-

maine transformé. La fonction $\widehat{p}\,(u,v)$ donne la valeur du pixel de l'image quantifiée pour (u,v) appartenant au domaine transformé Ω^T. En supposant que l'erreur de quantification ε est additive [396], et que f_o et ε sont décorrélés, il est possible d'écrire la relation :

$$\widehat{p} = Q\,(T f_o) = T f_o + \varepsilon, \tag{5.3}$$

avec Q l'opérateur de quantification et T l'opérateur de transformation supposés connus.

Critère proposé Le problème principal au décodage est de reconstruire l'image f^* qui approxime l'image d'origine f_o en utilisant la connaissance de $\widehat{p}$ et T. Rao, Miller, Rose et Gersho ont proposé une solution à ce problème en introduisant la quantification vectorielle généralisée [431], [432]. Pour T et sa transformation duale T^* connues, la solution qu'ils proposent consiste à optimiser conjointement les dictionnaires de codage et de décodage, ce qui modifie les valeurs de $\widehat{p}$ en fonction de la minimisation d'un certain critère. Dans notre cas, nous proposons de résoudre le problème de décodage formulé par l'équation (5.3) en minimisant une fonctionnelle J sur l'espace Ω. L'idée générale est de trouver f^* qui minimise J pour T et $\widehat{p}$ fixés, telle que :

$$f^* = \arg \min_{f \in L^2(\Omega)} J\,(f)\,, \tag{5.4}$$

où nous avons défini J par

$$J\,(f) = \int_{\Omega^T} (Tf - \widehat{p})^2 \, d\Omega^T. \tag{5.5}$$

Afin de résoudre ce problème inverse mal posé et éviter une solution dominée par le bruit, nous régularisons la solution du critère J donné équation (5.5) de façon à rendre le problème bien posé. La régularisation assure l'existence, l'unicité et la stabilité de la solution. On peut établir une régularisation en introduisant une fonction de potentiel [388] de façon à éviter un lissage isotropique. En effet, ce terme permet de supprimer l'effet néfaste du bruit sur la solution obtenue tout en préservant les contours. Le choix ainsi que les propriétés de la fonction de potentiel φ comme fonction de régularisation[1] préservant les contours sont indiqués dans les travaux de Charbonnier [387], [386], [388]. La fonction φ choisie est une fonction convexe de classe C^2. Le critère régularisé est alors donné par :

$$J\,(f) = \int_{\Omega^T} (Tf - \widehat{p})^2 \, d\Omega^T + \lambda^2 \int_{\Omega} \varphi\,(|\nabla f|)\, d\Omega. \tag{5.6}$$

[1] Notons que $\varphi\,(x) = x^2$ correspond à une régularisation au sens de Tikhonov.

Dans le cas où T est une transformée en ondelettes sur N niveaux, $J(f)$ devient :

$$J(f) = \sum_{k=1}^{3N+1} \pi_k \int_{\Omega_k^T} (Tf - \widehat{p})^2 \, d\Omega_k^T + \lambda^2 \int_{\Omega} \varphi(|\nabla f|) \, d\Omega, \tag{5.7}$$

où les pondérations π_k sont introduites par la biorthogonalité des filtres (cf. chapitre 2) et où Ω_k^T est le domaine des coefficients transformés appartenant à la sous-bande k avec,

$$\bigcup_k \Omega_k^T = \Omega^T. \tag{5.8}$$

5.2.1.2 A priori sur le bruit de quantification : bruit borné

Formalisation du problème Le terme d'attache aux données de l'équation (5.5) suppose que la solution f^* continue peut être déterminée au sens des moindres carrés par la mesure directe de $\Pr(f/\widehat{p})$, grâce à la connaissance de la quantité observée $\widehat{p}$ et en supposant que f est gaussien au voisinage de $\widehat{p}$ [438]. Cependant, ce terme n'est pas suffisant pour assurer une bonne modélisation du quantificateur : il ne suppose pas que le bruit de quantification ε est un *bruit borné*[2]. En effet, si nous prenons le cas simple d'un quantificateur scalaire uniforme de pas q on a :

$$-\frac{q}{2} \leq \varepsilon < \frac{q}{2}, \tag{5.9}$$

ou encore,

$$Tf \in I_q = \left[\widehat{p} - \frac{q}{2}, \widehat{p} + \frac{q}{2}\right[\tag{5.10}$$

où q désigne le pas de quantification utilisé au site (u, v) pour quantifier $Tf(u, v)$. Ceci nous amène à reformuler le problème de minimisation de la fonctionnelle J définie par l'équation (5.7) par le problème d'optimisation (P) sous contrainte suivant :

$$(P) \begin{cases} \text{minimiser } J(f) \\ \text{sous la contrainte } Tf \in I_q \end{cases} \tag{5.11}$$

Satisfaire la condition $Tf \in I_q$ est équivalent à vérifier les deux inégalités :

$$\begin{cases} g_{1,q}(f) = Tf - \widehat{p} - \frac{q}{2} < 0 \\ g_{2,q}(f) = \widehat{p} - Tf - \frac{q}{2} \leq 0 \end{cases} \tag{5.12}$$

pour tout $Tf \in \Omega_k^T$ et $k \in \{1, ..., N\}$. L'ensemble F des solutions de (P) est alors défini par

$$F = \left\{ f \in L^2(\Omega) / g_{1,q}(f) < 0 \text{ et } g_{2,q}(f) \leq 0 \right\}. \tag{5.13}$$

[2] Cette idée était apparue parallèlement dans [446] pour la restauration d'images codée JPEG et mise en œuvre par une méthode de projection.

Méthode basée sur la dualité lagrangienne Nous avons ramené la résolution de ce problème d'optimisation sous contraintes en un problème d'optimisation sans contraintes en utilisant la dualité lagrangienne. La notion de dualité lagrangienne constitue un outil particulièrement adapté pour l'optimisation des fonctions convexes ou localement convexes [424]. Le Lagrangien associé au problème (P) est alors donné par

$$\mathcal{L}(f,\mu) = J(f) + \sum_{k=1}^{3N+1} \sum_{i=1}^{2} \int_{\Omega_k^T} \mu_{i,k} g_{i,q_k}(f)\, d\Omega_k^T, \tag{5.14}$$

avec $\mu_{i,k}$ des nombres réels positifs ou nuls. Prouver l'existence et l'unicité d'une solution à ce critère est un problème difficile. Une étude théorique complète de ce critère est donnée dans la thèse de Tramini [438]. La solution optimale au problème (P) est obtenue en déterminant un point col de la fonction de Lagrange. La recherche du point col se fait en résolvant le système (Q) dual de (P) suivant :

$$(Q) \begin{cases} \max\limits_{\mu} w(\mu) = \max\limits_{\mu} \left[\min\limits_{f \in L^2(\Omega)} \mathcal{L}(f,\mu) \right] \\ \mu \in L^2\left(\Omega^T\right)^+ \end{cases} \tag{5.15}$$

- $w(\mu)$ est une fonction concave de μ. Cette propriété est absolument générale et ne suppose rien sur la convexité du critère J et sur les fonctions $g_{i,q}$, ni sur la convexité de l'ensemble des solutions F. La concavité de $w(\mu)$ permet d'affirmer que tout optimum local μ^* de $w(\mu)$ est un optimum global. La preuve de la concavité en μ de la fonction duale est donnée dans la thèse de Tramini (annexe E) [438].
- Lorsqu'il existe un point-col (en particulier dans le cas convexe) et que le minimum en f de la fonction de Lagrange $\mathcal{L}(f,\mu^*)$ est unique pour $\mu{=}\mu^*$ (optimum du dual), la résolution du problème dual (Q) permet d'obtenir une solution optimale du problème (P). Dans les cas où il n'existe pas de point-col, la résolution du problème dual procure une solution approchée.

Notre objectif était de développer un algorithme permettant d'approcher une solution de (P). C'est dans ce sens que nous avons introduit l'algorithme de décodage MORPHE (pour "**M**ethod for **O**ptimal **R**econstruction including **P**rojection and **H**yperparameters **E**stimation") [438].

5.2.1.3 MORPHE : Filtrage inverse non linéaire dynamique

Minimisation du Lagrangien Elle constitue la première étape de l'optimisation qui est une étape cruciale pour la régularisation. Supposons que la

solution f^* au problème de minimisation existe. Cette solution f^* vérifie alors

$$\left.\frac{\partial \mathcal{L}(f,\mu)}{\partial f}\right|_{f=f^*} = 0. \tag{5.16}$$

Dans [438], nous avons calculé les équations d'Euler Lagrange associées. Pour tout $k \in \{1,...,N\}$, ces équations sont données par :

$$\frac{\partial \mathcal{L}(f,\mu)}{\partial f} = T^*\pi_k\left(Tf-\widehat{p}\right) - \lambda^2 div\left(\frac{\varphi'\left(|\bigtriangledown f|\right)}{2\,|\bigtriangledown f|}\cdot\bigtriangledown f\right) + \frac{1}{2}T^*\left(\mu_{1,k}-\mu_{2,k}\right), \tag{5.17}$$

où T^* est l'opérateur transformée dual, avec les conditions aux bords de Neumann :

$$\left.\frac{\partial f}{\partial n}\right|_{\partial\Omega} = 0.$$

La résolution du système (5.17) nous permet d'obtenir la solution f^*, elle détermine ainsi la fonction duale $w(\mu)$.

Recherche de la variable duale optimale Cette deuxième étape de recherche de l'optimum de la fonction duale ne pose pas de réelles difficultés en raison de la propriété de concavité de $w(\mu)$. En pratique μ^* n'étant pas connu *a priori*, nous proposons d'appliquer une méthode itérative pour engendrer une suite $(\mu^m)_{m\geq 0}$ convergeant vers μ^*. On définit alors μ^{m+1} par la relation de récurrence donnée par

$$\mu^{m+1} = P_+\left(\mu^m + \eta_m \bigtriangledown w\left(\mu^m\right)\right) \tag{5.18}$$

où $P_+(\mu) = \max(\mu, 0)$. Dans cet algorithme, les pas de déplacement η_m peuvent être choisis *a priori* : il s'agit alors d'une méthode de gradient à pas prédéterminés ou de sous-gradient, appliquée à la fonction duale (le lecteur intéressé dans la stratégie des choix des pas peut se référer à l'ouvrage [424]).

Algorithme proposé Les méthodes qui consistent à résoudre le problème dual exploitent la propriété de concavité de la fonction duale w et le fait que l'on peut déduire un gradient du calcul de la valeur $w(\mu)$. Tout algorithme de gradient peut être utilisé. Nous proposons d'appliquer le schéma d'optimisation alterné donné par l'algorithme suivant basé sur la méthode de Usawa [439]. On note :

$$C(f,\mu) = \sum_{k=1}^{3N+1}\sum_{i=1}^{2}\int_{\Omega_k^T}\mu_{i,k}g_{i,q_k}(f)\,d\Omega_k^T \tag{5.19}$$

ALGORITHME

1. **Initialisation :** *Choisir* $\mu_1^0 \geq 0$, $\mu_2^0 \geq 0$ et $m = 0$
2. *Chercher* $f^{*(m+1)} = \arg \min_{f \in L^2(\Omega)} \mathcal{L}(f, \mu^m)$ en résolvant[3] le système (5.17).
3. *Calculer* $w(\mu^m) = J\left(f^{*(m+1)}\right) + C\left(f^{*(m+1)}, \mu^m\right)$
4. *Evaluer* $\mu^{m+1} = P_+(\mu^m + \eta_m \bigtriangledown w(\mu^m))$
5. **Test d'arrêt : Si** $C\left(f^{*(m+1)}, \mu^{m+1}\right) \simeq 0$ **Alors** *arrêter* **Sinon** $m \leftarrow m+1$ et *aller à l'étape 2.*

Propriété de conservation du train binaire Afin de supprimer les artefacts liés à une compression de données, les méthodes "classiques" de post-traitement ont tendance à trop lisser l'image décomprimée et donc à perdre une partie de l'information qu'elle contient. Cette perte d'information entraîne alors une *non conservation du train binaire* de l'image comprimée. En effet, si l'image décomprimée traitée par le post-traitement est passée une nouvelle fois dans le système de compression avec les mêmes paramètres (transformée et quantificateur identiques), l'image en sortie du codeur est différente de celle trouvée lors du premier codage. La prise en compte de la caractéristique de *bruit borné* permet de remédier à ce problème. En effet, si après le post-traitement on assure que la solution vérifie la contrainte (5.10), alors le système devient robuste à un nombre "infini" de compression / décompression / post-traitement successifs.

5.2.1.4 Résultats

Sur la figure 5.1 nous présentons un résultat de décodage d'image au moyen de l'algorithme MORPHE. Les résultats montrent un gain en rapport Signal-à-Bruit Pic (PSNR) d'environ 1,03 dB par rapport à un décodage "classique" qui consiste à appliquer le filtrage inverse linéaire au moyen des filtres duaux de synthèse T^*. L'algorithme de compression qui a été utilisé est celui que nous avons développé au cours de la thèse de Raffy [8], [429] avec les filtres "9-7" (cf. chapitre 2). Une étude expérimentale complète se trouve dans la thèse de Tramini [438].

5.2.1.5 Application à l'imagerie satellitaire

Problématique Dans le cadre d'une étude avec le Centre National d'Etudes Spatiales (CNES) sur la déconvolution d'images satellitaires haute résolution, nous avons adapté la méthode MORPHE au cadre de la restauration d'images

[3] La recherche du minimum du Lagrangien en f est faite au moyen d'un algorithme de régularisation semi-quadratique développé dans [388].

FIG. 5.1 – Décodage optimal de l'image LENA comprimée avec un taux de compression de 86 :1 (0,093 bpp). Images de gauche : décodage standard par filtrage inverse linéaire - PSNR=30,18 dB. Images de droite : décodage par MORPHE - PSNR=31,21 dB.

floues altérées par un bruit borné. Les instruments optiques d'aujourd'hui doivent être performants et peu coûteux. L'amélioration du rapport performance-coût dépend en grande partie de la synergie de l'ensemble du système : instrument-traitement sol. De plus, l'importance des coûts instrumentaux à bord est un facteur décisif, d'où la nécessité du report au sol des traitements complexes. Depuis la génération SPOT5, des modes hautes résolutions produisent des images sans repliement spectral. Cette propriété est essentielle afin d'effectuer la déconvolution-restauration dans de bonnes conditions. La connaissance fine des caractéristiques instrumentales dont dispose le CNES (FTM, caractéristiques du bruit d'acquisition ...) permet d'effectuer ces déconvolution-restauration en introduisant des données *a priori*. De multiples travaux de déconvolution-restauration existent, mais nous ne proposons pas dans ce document un rappel ou une étude de ces méthodes. Le lecteur intéressé pourra se référer aux travaux de [408], [433], [434], [435], [420]. Notre tâche a été de tirer parti des informations fournies par le CNES (dans le cas d'images satellites) ou par le système d'acquisition, et MORPHE correspond parfaitement à cette attente, surpassant tant au niveau visuel qu'en terme de SNR les méthodes classiques de déconvolution-restauration [438].

En effet, les systèmes d'imagerie satellitaire entraînent l'apparition de flou lié à l'optique du satellite, l'introduction de bruit électronique et de numérisation. De plus, la capacité mémoire liée au stockage ou à la transmission étant limitée, l'utilisation d'un codeur est nécessaire afin de réduire le débit binaire. L'usage d'un codeur couplé au système d'acquisition a pour conséquence d'engendrer la création d'artefacts nuisibles à la qualité visuelle. Ces opérations se caractérisent par une perte d'information, ainsi que par l'apparition d'artefacts (moutonnement, ringing pour les ondelettes, effets de bloc pour JPEG). La suppression de ces artefacts permettrait un meilleur confort visuel dans la perception des données restaurées ou d'envisager des taux de compression supérieurs.

Contrairement aux méthodes classiques, qui effectuent séparément le décodage, le débruitage et la déconvolution, nous avons proposé une méthode de DECODAGE - DEBRUITAGE - DECONVOLUTION conjoint pour la reconstruction d'images. Ces travaux ont fait l'objet des publications suivantes : [57], [62], [61], [64], [71].

Modélisation du processus d'acquisition Les imperfections des capteurs de nombreux systèmes d'imagerie introduisent une dégradation que l'on peut modéliser par une convolution avec un filtre passe-bas de réponse impulsionnelle H et le cumul d'un bruit additif b correspondant au bruit des capteurs. Le filtre H, qui correspond à la Fonction de Transfert de Modulation (FTM) de l'optique, est un filtre bidimensionnel supposé être anti-aliasing [418]. Dans ce

cas de figure, l'image acquise peut s'exprimer de la manière suivante :

$$f_b = Hf_o + b. \tag{5.20}$$

Selon les experts de systèmes d'imagerie, le bruit d'acquisition global peut être assimilé à un bruit borné centré et non stationnaire. Ainsi, si l'on note $b_{\min}$ et $b_{\max}$ les bornes du signal $b(x, y)$, la connaissance de f_b version floue et bruitée de f_o nous permet de dire que la solution recherchée f^* doit vérifier :

$$Hf^* \in I_b = [f_b - b_{\min}, f_b + b_{\max}] \tag{5.21}$$

Contrairement aux bornes du bruit de quantification qui dépendent de chaque sous-bande k, les bornes $b_{\min}$ et $b_{\max}$ ne dépendent que du pixel observé et leur détermination est faite directement à partir de la distribution du signal b. Nous avons considéré dans nos travaux une modélisation uniforme de la distribution de ce bruit [438].

Modélisation de la dégradation acquisition/compression En combinant les deux processus d'acquisition et de compression, il est possible de modéliser la transformation complète permettant de passer de l'image réelle f_o recherchée à l'image observée $\widehat{p}$. Ce modèle est de la forme :

$$\widehat{p} = Q(THf_o + Tb) = THf_o + \eta, \tag{5.22}$$

où $\eta = Tb + \varepsilon$ représente le bruit dû à la fois à la quantification et aux imperfections du système d'acquisition. Dans ce cas de figure, la fonctionnelle J introduite à l'équation (5.7) devient :

$$J(f) = \sum_{k=1}^{3N+1} \pi_k \int_{\Omega_k^T} \left(THf - \widehat{p}\right)^2 d\Omega_k^T + \lambda^2 \int_{\Omega} \varphi\left(|\nabla f|\right) d\Omega. \tag{5.23}$$

De plus, nous avons montré dans [438] que dans l'espace transformé la condition (5.21) devenait :

$$THf^* \in I_\theta = [\widehat{p} - \theta, \widehat{p} + \theta] \tag{5.24}$$

où les bornes θ sont définies pour une sous-bande k par la somme de deux bornes :

$$\theta = \frac{q}{2} + b_T. \tag{5.25}$$

La détermination de la borne b_T qui dépend du bruit d'acquisition n'est pas un problème facile. Le calcul est donné dans la thèse de Tramini [438]. Notons simplement que cette borne est maintenant multirésolution et dépend

de la sous-bande k considérée. La formulation du problème de minimisation de la fonctionnelle J définie par l'équation (5.23) devient alors :

$$(P)\begin{cases} \text{minimiser } J(f) \\ \text{sous la contrainte } THf \in I_\theta \end{cases} \tag{5.26}$$

Une étude complète de ce problème est donnée dans la thèse de Tramini [438] ainsi que de nombreux résultats visuels et numériques sur des images satellites fournie par le CNES Toulouse. L'algorithme que nous avons développé pour ce problème reste basé sur la méthode MORPHE décrite au paragraphe 5.2.1.3.

5.2.2 Dans le cadre des vidéos

Ces travaux on été réalisés durant la thèse de Joël Jung (2000) [407] que j'ai co-encadrée en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Une partie de nos travaux a été publiée dans la revue IEEE Transactions on Multimedia en juin 2003 : "Optimal decoder for block-transform based video coders" [14]. Cet article est donné en annexe G du document. Nous avons de plus breveté nos travaux avec le CNRS [22].

5.2.2.1 Problématique

Le problème de l'effet de blocs La plupart des algorithmes de compression de vidéos numériques standardisés et utilisés dans des systèmes d'acquisition (appareils photos numériques, caméras vidéos numériques) ou de transmission (télévision numérique) sont basés sur le standard JPEG et donc sur des transformées en blocs (DCT). Dans toutes ces applications, la qualité des images fournies par les décodeurs est un point clé. Or, ces images souffrent généralement d'artefacts tels que les *effets de blocs* introduits par la quantification et la transformée. Ainsi, les constructeurs doivent inclure des algorithmes de post-traitements dans les systèmes de façon à améliorer au mieux la qualité des images et des vidéos décomprimées. De nombreuses approches ont été proposées dans la littérature pour réduire les effets de blocs. Historiquement, Gersho [430] a initié ces approches sur les images fixes en utilisant un filtrage non linéaire adaptatif. Par la suite, d'autres méthodes ont été suggérées basées sur des filtrages globaux et locaux [419], du seuillage dans le domaine des ondelettes [402], ou encore sur la minimisation d'un critère [443]. Les méthodes de décodage avancées adoptent une approche globale pour le décodage au moyen de la minimisation d'un critère qui modélise les effets de blocs [423], [58]. Elles permettent de diminuer les effets indésirables de lissage liés au post-traitement. Néanmoins, le traitement de la vidéo nécessite une prise en compte des caractéristiques temporelles de celle-ci [14] et non pas un traitement indépendant sur chacune des images consécutives. Dans cette optique, des travaux récents ont

introduit des méthodes de suppression des effets de bloc pour les vidéos codées MPEG [444], [442], [385], [393], [427], [445], [14]. Notons que l'algorithme de post-traitement préconisé par la norme MPEG-4 et basé sur [428] effectue une détection des artefacts dans le domaine spatial alors que la correction se fait dans le domaine transformé DCT par des opérations de filtrage. Cet algorithme constitue actuellement la référence [403].

Les méthodes orientées objet La prise en compte de l'aspect temporel a aussi été étudiée au moyen de méthodes spatio-temporelles *orientées objet*. Ce problème a été largement investigué et de nombreuses approches différentes ont été proposées dans la littérature. Par exemple, nous pouvons citer les travaux de [404], [406] pour la segmentation spatio-temporelle d'objets en mouvement et ceux de [399], [405] pour le suivi d'objets ("tracking"). La plupart des méthodes exploite les informations spatiales et temporelles dans les vidéos pour séparer le fond fixe de l'objet en mouvement [440], [394]. Dans [384] Katsaggelos proposa une méthode qui permet de traiter simultanément le problème de suppression de bruit et d'estimation du mouvement. Le même problème a été étudié dans [415] ou les auteurs traitent la segmentation spatio-temporelle et la restauration de la vidéo de façon couplée.

Notre contribution La plupart de ces méthodes n'exploite pas toute l'information apportée par le train binaire comprimé, comme par exemple la connaissance des pas de quantification, les vecteurs mouvements, le type de macrobloc (I ou P) etc. En conséquence, d'un côté les résultats de la segmentation sont sévèrement altérés par les effets de blocs et d'un autre côté les méthodes utilisées pour supprimer les effets de blocs ne peuvent bénéficier d'une segmention spatio-temporelle efficace pour traiter séparément les objets en mouvement et le fond fixe. Dans cet optique, nous avons proposé un algorithme de suppression des effets de blocs orienté objet qui exploite les informations contenues par le train binaire M-JPEG ("Motion-JPEG") ou MPEG. Cet algorithme se résume en trois points :

1. Le fond fixe de la scène est estimé ainsi que les pixels en mouvement. Les effets de blocs et le bruit de quantification sont traités sur le fond fixe de façon conjointe à la procédure de segmentation spatio-temporelle.
2. Chaque objet en mouvement est ensuite isolé et un suivi d'objet est effectué de façon à générer des séquences indépendantes d'objets en mouvement ("Video Object Plane" - VOP). Les effets de blocs et le bruit de quantification sont traités sur chaque VOP de façon indépendante.
3. La séquence débruitée est reconstruite à partir du fond fixe et des VOPs.

Notre approche et nos contributions dans ce domaine sont décrites dans les paragraphes suivants.

5.2.2.2 La séparation du fond et des objets en mouvement

Le critère de segmentation spatio-temporel proposé Le problème posé consiste à séparer les pixels en mouvement c_k du fond f_k^* estimé à l'image k dans la séquence. L'idée principale de notre approche est basée sur l'observation que les objets en mouvement dans la séquence sont caractérisés par de fortes discontinuités temporelles. Ainsi, afin de séparer les objets en mouvement du fond, nous avons proposé d'effectuer un lissage temporel de la séquence sur tous les pixels qui présentent un fort gradient temporel. Ce problème est formulé comme un problème inverse en introduisant la fonctionnelle (5.27) suivante :

$$J\left(f_k, c_k\right) = \underbrace{\int_{\Omega} c_k^2 \left(f_k - T^* \widehat{p}_k\right)^2 d\Omega}_{(1)} + \underbrace{\alpha_p \int_{\Omega} \left(c_k - 1\right)^2 \left\|\nabla_T f_k\right\|^2 d\Omega}_{(2)}, \tag{5.27}$$

où T^* représente la DCT inverse (T est ici l'opérateur DCT), $T^*\widehat{p}_k$ est l'image observée et ∇_T symbolise le gradient temporel entre l'image f_k et une image de moyenne temporelle m_k définie par la relation suivante :

$$m_k = \frac{1}{\sum_{i=1}^{k-1} c_i} \sum_{i=1}^{k-1} c_i f_i, \quad \text{pour} \sum_{i=1}^{k-1} c_i \neq 0. \tag{5.28}$$

On définit le gradient temporel par l'équation (5.29) :

$$\nabla_T f_k = \left|f_k - m_k\right| \tag{5.29}$$

Notons que dans la relation (5.27), α_p est un coefficient qui permet de contrôler la convergence du critère vers $T^*\widehat{p}_k$ ou m_k. Pour obtenir des fonds différents sur chaque image :

- Les zones de l'image correspondant aux objets vont être remplacées par une information temporelle, correspondant à une moyenne évoluée des pixels des images précédentes, sur cette zone (terme (2) dans l'équation (5.27)) ;
- Les zones correspondant au fond sont remplacées par des pixels issus de l'image observée (terme (1) dans l'équation (5.27)).

Ce principe est schématisé dans la thèse de Jung [407] et dans notre article "Optimal decoder for MPEG-based video coder" [14] donné en annexe G du document (figure 2).

La solution Pour chaque image k de la séquence, l'image de fond f_k^* minimise J pour T, $\widehat{p}_k$ et m_k fixés, telle que :

$$f_k^* = \arg \min_{f_k, c_k} J\left(f_k, c_k\right). \tag{5.30}$$

La solution optimale de ce problème est donnée pour :

$$\frac{\partial J\left(f_k, c_k\right)}{\partial f_k} = 0 \text{ et } \frac{\partial J\left(f_k, c_k\right)}{\partial c_k} = 0 \tag{5.31}$$

ce qui est équivalent à résoudre le système suivant pour chaque image k :

$$\begin{cases} \left(c_k^2 + \alpha_p \left(c_k - 1\right)^2\right) f_k = c_k^2 T^* \widehat{p}_k + \alpha_p \left(c_k - 1\right)^2 m_k & \text{(a)} \\ \text{et } c_k = \frac{\alpha_p (f_k - m_k)^2}{\alpha_p (f_k - m_k)^2 + (f_k - T^* \widehat{p}_k)^2} & \text{(b).} \end{cases} \tag{5.32}$$

L'interprétation de cette dérivée permet de distinguer clairement que f_k va tendre vers $T^*\widehat{p}_k$, vers m_k, ou vers un compromis des deux, en fonction des valeurs de c_k . De plus, on observe que si f_k est proche de m_k, c'est-à-dire que l'on se trouve sur un pixel appartenant à un objet en mouvement, c_k tend vers 0. Inversement, si f_k est proche de $T^*\widehat{p}_k$, c_k tend vers 1.

Algorithme proposé Nous avons proposé l'algorithme de minimisations alternées suivant pour résoudre le système (5.32) :

ALGORITHME

1. **Initialisation :** $c_k = 1$
2. **Répéter jusqu'à convergence du critère en f_k :**
   (a) ETAPE 1 : *résoudre l'équation (a) du système (5.32) en* f_k avec c_k fixé.
   (b) ETAPE 2 : *résoudre l'équation (b) du système (5.32) en* c_k avec f_k fixé.

Au cours des itérations, la résolution du système (5.32) fournie une estimée f_k^* du fond et de l'image c_k des pixels en mouvement telle que $c_k \in [0, 1]$. Grâce à ce principe de minimisations alternées, où le fond et les pixels en mouvement sont calculés en parallèle et bénéficient chacun de la connaissance de l'autre, les "objets" sont efficacement détectés et retirés du fond.

5.2.2.3 Les contraintes liées au débruitage du fond

Régularisation du critère La minimisation du critère (5.27) proposé effectue une segmentation spatio-temporelle mais n'assure pas la suppression du bruit de quantification ni des effets de blocs introduits par le codage par DCT. De la même façon que pour le décodage d'images fixes et pour supprimer l'effet

néfaste du bruit sur la solution obtenue tout en préservant les contours et éviter ainsi un filtrage isotropique, nous introduisons un terme de régularisation au critère (5.27). Ce terme est donné par :

$$C_1(f_k) = \lambda^2 \int_\Omega \varphi_1(|\nabla f_k|)\, d\Omega \tag{5.33}$$

Suppression des effets de blocs L'objectif est de réduire l'effet de bloc [58], [65] présent sur l'image décodée, et plus particulièrement d'atténuer l'aspect spatial de cet artefact : au sein d'une même image des sauts d'intensité apparaissent à la frontière des blocs DCT, et ce d'autant plus que le débit est faible. L'action que nous devons mener se situe donc spatialement sur le bord de ces blocs DCT. Nous avons proposé de localiser spatialement les effets de blocs de sorte de les éliminer sans altérer l'information contenue dans l'image. Pour cela, la transformée en ondelettes est un outil parfaitement adapté. En effet, l'idée est de seuiller dans le domaine transformé les coefficients d'ondelettes correspondant aux bords des blocs, tout en conservant les autres coefficients qui représentent les détails de l'image. Le processus de décision pour réaliser la sélection est donc primordial, afin de ne pas altérer l'image. Cette sélection, *a priori* délicate, est simplifiée par le fait que l'on connait de façon précise la localisation de l'effet de bloc (la taille des blocs DCT est généralement connue) et par la connaissance de la géométrie de l'effet de bloc qui se caractérise par des discontinuités horizontales et verticales et apparaît donc sous forme de lignes de coefficients d'ondelettes horizontales ou verticales, isolées des autres coefficients.

Pour supprimer les effets de blocs, nous avons introduit une nouvelle contrainte qui réalise un seuillage doux des coefficients d'ondelettes [9]. Elle est donnée par la relation :

$$C_2(f_k) = \eta^2 \int_{\Omega^R} \varphi_2\left(\frac{|Rf_k|}{\delta}\right) d\Omega^R \tag{5.34}$$

où δ est un seuil qui dépend de l'amplitude de l'effet de bloc, R représente l'opérateur de la transformée en ondelettes et Ω^R est le support de l'image dans le domaine transformé en ondelettes. La valeur de η indique, pour chaque pixel, si le coefficient doit être seuillé ou non. Cette variable a été positionnée pour chaque pixel par un processus de sélection décrit dans la thèse de Jung [407] et qui tient compte des remarques précédentes, de sorte que $\eta = 1$ si le coefficient doit être seuillé, et $\eta = 0$ s'il doit être préservé. En pratique, la contrainte C_2 réalise par conséquent un seuillage doux dans le domaine spatio-fréquentiel. Une étude complète sur la façon de procéder est donnée dans la thèse de Jung [407].

Prise en compte du bruit borné de quantification Lors du décodage, les matrices de quantification utilisées au codeur et présentes dans la trame

binaire sont connues pour chaque bloc DCT. De la même manière que pour les images fixes, les valeurs de la transformée DCT de l'image f_k doivent vérifier :

$$Tf_k \in I_q = \left[\widehat{p}_k - \frac{q}{2}, \widehat{p}_k + \frac{q}{2}\right[. \tag{5.35}$$

où q désigne le pas de quantification utilisé au site (u, v) pour quantifier $Tf_k(u, v)$ qui correspond à la DCT de l'image f à l'instant k. Satisfaire la condition $Tf_k \in I_q$ est équivalente à vérifier les inégalités données par le système d'équations (5.12).

5.2.2.4 Le critère spatio-temporel régularisé

La fonctionnelle complète permettant de réaliser à la fois le débruitage et la suppression des effets de blocs du fond, ainsi que l'extraction des pixels en mouvement est donnée par l'équation suivante [65], [63] :

$$\begin{aligned} J(f_k, c_k) = &\int_\Omega c_k^2 \left(f_k - T^*\widehat{p}_k\right)^2 d\Omega + \alpha_p \int_\Omega (c_k - 1)^2 (f_k - m_k)^2 \, d\Omega \\ &+ \lambda^2 \int_\Omega \varphi_1 (|\nabla f_k|) \, d\Omega + \eta^2 \int_{\Omega^R} \varphi_2 \left(\frac{|Rf_k|}{\delta}\right) d\Omega^R \end{aligned} \tag{5.36}$$

Le rajout de la contrainte de bruit borné donné au paragraphe 5.2.2.3 précédent nous amène a formaliser le problème de minimisation de cette fonctionnelle par le problème (P) suivant :

$$(P) \begin{cases} \text{minimiser } J(f_k, c_k) \\ \text{sous la contrainte } Tf_k \in I_q \end{cases} \tag{5.37}$$

Ce problème peut être résolu en utilisant la dualité Lagrangienne pour c_k fixé et dans ce cas, le Lagrangien $\mathcal{L}(f_k, c_k, \mu)$ associé au problème (P) s'écrit :

$$\mathcal{L}(f_k, c_k, \mu) = J(f_k, c_k) + \sum_{n=1}^{N} \sum_{i=1}^{2} \int_{\Omega_n^T} \mu_{i,n} g_{i,q_n}(f_k) \, d\Omega_n^T \tag{5.38}$$

où N désigne ici le nombre de coefficients dans un bloc DCT et Ω_n^T est le support de l'image dans le domaine transformé DCT qui regroupe les coefficients DCT de même fréquence. Approcher une solution de (P) peut être obtenu au moyen de l'algorithme MORPHE défini au paragraphe 5.2.1.3 précédent. Les équations d'Euler Lagrange associées sont données par :

$$\begin{aligned} &\frac{\partial \mathcal{L}(f_k, c_k, \mu)}{\partial f_k} = c_k^2 \left(f_k - T^*\widehat{p}_k\right) + \alpha_p (c_k - 1)^2 (f_k - m_k) \\ &- \lambda^2 div \left(\frac{\varphi_1'(|\nabla f_k|)}{2|\nabla f_k|} \cdot \nabla f_k\right) + \eta^2 R^* \frac{\varphi_2'(|Rf_k|/\delta)}{2|Rf_k|/\delta} Rf_k + \frac{1}{2} T^* (\mu_{1,k} - \mu_{2,k}) \end{aligned} \tag{5.39}$$

FIG. 5.2 – De gauche à droite : l'image originale, l'image des pixels en mouvement c_k et l'image des objets indexés.

La dérivation complète de la fonctionnelle par le théorème des accroissements finis est donné dans l'annexe B de la thèse de Jung [407]. Ainsi, l'ETAPE 1 de l'algorithme donné au paragraphe 5.2.2.2 revient à chercher la solution du système d'équations non linéaires (5.39).

5.2.2.5 Le traitement des objets en mouvement

Extraction des VOP La segmentation spatio-fréquentielle donnée par la résolution du système (5.39) et l'algorithme proposé au paragraphe 5.2.2.2 permet :

- D'estimer l'image de fond f_k pour chaque image k de la séquence ;
- D'obtenir une estimation des pixels en mouvement (images c_k) robuste aux effets de blocs qui apparaissent dans la séquence codée MPEG.

A partir des images c_k de pixels en mouvement on déduit la liste des objets en mouvement (VOP) avec leurs caractéristiques spatiales et temporelles au moyen d'une segmentation spatiale et d'un tracking [407]. L'objectif de nos travaux n'est pas de proposer de nouvelles méthodes pour l'indexation et le tracking. Il existe de nombreuses méthodes qui font encore actuellement l'œuvre de travaux spécifiques, telles que [436], [413], [422]. Un exemple d'extraction d'objets en mouvement à partir des images c_k est donné sur la figure 5.2.

Traitement des VOP Grâce à la connaissance des caractéristiques spatiales et temporelles des objets, chacun d'entre eux va être traité indépendamment, en fonction de ses propres caractéristiques. Le traitement repose sur la minimisation d'une fonctionnelle, dont le comportement est proche de celui appliqué sur le fond. Les contraintes spatiales sont similaires, par contre en temporel la segmentation est remplacée par un moyennage temporel sur $2n+1$ images, avec compensation de mouvement. Ainsi, pour chaque objet O_k^l, et pour chaque

image k, la nouvelle représentation restaurée de l'objet est donnée par la minimisation de la fonctionnelle (5.40) suivante :

$$J\left(O_k^l\right) = \sum_{i=-n}^{n} \int_{\Omega_{O_k^l}} \left(O_k^l - T^* \widehat{p}_{k+i}\left(x + u_{k,k+i}^l, y + v_{k,k+i}^l\right)\right)^2 d\Omega_{O_k^l} + C_1\left(O_k^l\right) + C_2\left(O_k^l\right) \quad (5.40)$$

où $\left(u_{i,j}^l, v_{i,j}^l\right)$ est le vecteur mouvement de l'objet O_i^l, entre l'image i et l'image j. Un tel vecteur mouvement est recherché pour chaque objet, et pour chaque image de la séquence. Une technique classique telle que le "block-matching" peut être utilisée pour réaliser cette opération. Dans le cas d'une séquence MPEG, cette information est directement fournie par le codeur dans le train binaire.

L'introduction d'une contrainte sur le bruit borné de quantification peut être traitée par l'algorithme MORPHE. Nous ne détaillerons pas ici les calculs du débruitage de l'objet dont l'approche est similaire à celle du fond f présentée précédemment. Le détail de la méthode est développé dans la thèse de Jung [407] et dans l'article "Optimal decoder for block-transform based video coders" [14] donné en annexe G.

5.2.2.6 Reconstruction de la séquence

Les objets débruités de la séquence appartenant à l'image k sont projetés dans l'espace Ω afin de constituer une image d'objets débruités O_k^*. L'image finale restaurée I_k^* est obtenue par l'équation (5.41) suivante. Pour chaque pixel (x, y) on a :

$$I_k^* = \frac{c_k^2 f_k^* + \left(c_k - 1\right)^2 O_k^*}{c_k^2 + \left(c_k - 1\right)^2} \quad (5.41)$$

avec $c_k \in [0, 1]$.

5.2.2.7 Algorithme

L'algorithme complet de décodage que nous avons développé repose sur quatre étapes. Cet algorithme porte le nom de OMD (pour "**O**ptimal **M**PEG **D**ecoding").

ALGORITHME

1. Pour chaque image k de la séquence faire :
   (a) ETAPE 1 : à l'aide de l'image observée $\widehat{p}_k$, et de la moyenne m_k calculée avec les images précédentes, le fond f_k est extrait et restauré

simultanément, et une image correspondante représentant les pixels en mouvement est calculée.

(b) ETAPE 2 : le pré-traitement de ces images de pixels en mouvement permet la création d'une carte d'identité pour chaque objet en mouvement, contenant des informations spatiales et temporelles, grâce à une indexation et un suivi d'objets. A partir de cet instant, la notion d'objet correspond à celle de VOP, et concerne une même entité au cours du temps.

(c) ETAPE 3 : la suppression des effets de bloc est réalisée indépendamment sur chaque objet O_k^l en fonction de leurs caractéristiques intrinsèques spatiales et temporelles.

(d) ETAPE 4 : la reconstruction de la séquence fusionne le fond et les objets restaurés.

5.2.2.8 Complexité

La méthode par optimisation que nous proposons s'appuie sur des minimisations alternées à la fois pour trouver les fonds et les objets restaurés. Ces algorithmes itératifs sont évidemment longs et coûteux du fait des nombreux systèmes d'équation à résoudre. L'étude théorique complète de la complexité de l'algorithme OMD n'est pas simple. A titre informatif, nous donnons ici des mesures expérimentales qui reflètent bien le comportement de l'algorithme. Les calculs ont été réalisés sur un Pentium II 450Mhz, disposant de 512Mo de mémoire vive. Nous ne donnons pas le temps de traitement total, peu significatif étant donné que le code et les algorithmes n'ont pas été optimisés pour la rapidité d'exécution. Seuls des rapports face à la méthode de décodage standard sont donnés.

Nous considérons ici le traitement de la séquence HALL sur 100 images, codées par l'algorithme M-JPEG. Le temps de calcul CPU requis par OMD a été mesuré 11,2 fois supérieur à celui de la méthode standard avec deux objets détectés et traités. Notons que ce rapport augmente avec le nombre d'objets, mais que le traitement objet est entièrement parallélisable. Le temps de calcul est réparti de la manière suivante entre les différents modules : 57% du temps est passé à l'ETAPE 1 de l'algorithme, 1% à l'ETAPE 2 sans tracking, 7% est utilisé par le tracking, 34% à l'ETAPE 3 (avec deux objets traités), et 1% à l'ETAPE 4.

5.2.2.9 Prise en compte des images I, P et B de MPEG

L'algorithme OMD présenté ici traite des séquences M-JPEG ou des images Intra (I) issues de codecs MPEG. Si l'on veut prendre en compte les images prédites (P et B) de la trame MPEG, quelques modifications doivent être appliquées à l'algorithme. Notamment, les images I seront traitées par les contraintes

spatiales alors que les images P seront décodées par le schéma suivant. Dans un schéma MPEG, l'image $\widehat{p}_k(x,y)$ restituée à l'instant k correspond à l'image décodée à l'instant précédent $\widehat{p}_{k-1}(x,y)$ compensée en mouvement par le vecteur (u,v), et à laquelle on ajoute l'image d'erreur quantifiée à l'instant k, notée $\widehat{\varepsilon}_k(x,y)$, soit :

$$\widehat{p}_k(x,y) = \widehat{p}_{k-1}(x-u, y-v) + \widehat{\varepsilon}_k(x,y) = \overline{\widehat{p}}_{k-1}(x,y) + \widehat{\varepsilon}_k(x,y). \quad (5.42)$$

Dans le cadre MPEG, c'est l'image d'erreur ε qui est codéee JPEG et qui par conséquent contient les artefacts de compression (bruit de quantification et effets de blocs). Le traitement OMD doit donc s'appliquer maintenant à ε_k et non plus à f_k comme pour M-JPEG. Le nouveau critère s'écrit donc :

$$J(\varepsilon_k, c_k) = \int_\Omega c_k^2 \left(\varepsilon_k + \overline{f}^*_{k-1} - \widehat{p}_k\right)^2 d\Omega + \alpha_p \int_\Omega (c_k - 1)^2 \left(\varepsilon_k + \overline{f}^*_{k-1} - m_k\right)^2 d\Omega + \lambda^2 \int_\Omega \varphi_1\left(\left\|\nabla\left(\overline{f}^*_{k-1} + \varepsilon_k\right)\right\|\right) d\Omega + \eta^2 \int_{\Omega^R} \varphi_2\left(\frac{|R\varepsilon_k|}{\delta}\right) d\Omega^R, \quad (5.43)$$

où $\overline{f}^*_k$ représente l'image décodée par OMD à l'instant k et compensée en mouvement, telle que :

$$f^*_k(x,y) = \overline{f}^*_{k-1}(x,y) + \varepsilon^*_k(x,y),$$

avec ε^*_k solution de la minimisation du problème. L'introduction d'une contrainte sur le bruit borné dû à la quantification du résiduel ε_k nous amène à résoudre le problème (P) suivant :

$$(P) \begin{cases} \text{minimiser } J(\varepsilon_k, c_k) \\ \text{sous la contrainte } T\varepsilon_k \in I_q = \left[\widehat{p}_k - \overline{\widehat{p}}_{k-1} - \frac{q}{2}, \widehat{p}_k - \overline{\widehat{p}}_{k-1} + \frac{q}{2}\right[\end{cases} \quad (5.44)$$

La résolution de ce problème en ε_k peut être traitée par MORPHE afin d'obtenir l'image f^*_k décodée OMD qui reproduira au mieux le fond sous les contraintes suivantes :

- L'image f^*_k restituée est lissée, et ses discontinuités sont préservées ;
- Les effets de bloc sont supprimés de l'image d'erreur ε^*_k ;
- Le bruit de quantification sur l'image d'erreur est réduit.

Un critère similaire est introduit sur les objets.

5.2.2.10 Résultats

Nous présentons ici des tests effectués sur la séquence HALL qui est une séquence de test retenue par le projet européen COST211, dont l'objectif est de

proposer des outils et des algorithmes dédiés aux formats MPEG4 et MPEG7. La figure (5.3) est un exemple de restauration avec à gauche la séquence décodée par le décodeur MPEG1, et à droite la séquence décodée par OMD[4]. Il est possible de remarquer que l'effet de bloc a été réduit à la fois sur le fond et sur l'objet en mouvement, et que les discontinuités sont conservées franches, même au bord de l'objet. Mais c'est en observant la séquence complète à 30 images par seconde que l'on perçoit l'amélioration principale provenant du traitement temporel.

La figure (5.4) compare en terme de Rapport Signal-à-Bruit Pic (PSNR) le décodage MPEG1 avec le décodage OMD proposé, pour l'ensemble des images de la séquence HALL. Nous remarquons que les deux courbes évoluent en parallèle, avec un écart de l'ordre de 0,6 dB en faveur d'OMD.

5.3 La réduction des bruits de transmission, d'acquisition et de stockage

Ces travaux on été réalisés durant la thèse de Joël Jung (2000) [407] que j'ai co-encadrée en collaboration avec le Professeur Michel Barlaud à l'Université de Nice-Sophia Antipolis. Une partie de nos travaux a été publiée dans la revue IEEE Transactions on Multimedia en juin 2003 : "Optimal decoder for block-transform based video coders" [14]. Cet article est donné en annexe G du document.

Le problème du contrôle et de la réduction des erreurs liées à la transmission prend une ampleur de plus en plus importante du fait de la multiplication actuelle des transmissions vidéos sur les réseaux. L'effet de ces pertes sur les séquences comprimées peut être catastrophique : la perte d'un paquet est d'autant plus grave, en raison de l'intégration de la compression dans le schéma de transmission. En effet, plus le taux de compression est fort, plus la quantité de données perdues est grande. De plus, la perte d'un paquet dans une séquence implique souvent une propagation temporelle des artefacts du fait de la prédiction temporelle présente dans les codages de type MPEG. Le problème se complique également si l'on considère des applications temps-réel, interactives, ou encore des communications multipoints. D'autre part, en plus des difficultés liées aux transmissions, des artefacts peuvent également provenir de l'acquisition, du stockage ou de la lecture par un matériel déficient.

Il existe deux[5] grands types d'approches qui ont été proposés à ce jour pour

[4] Remarquons que le traitement est uniquement réalisé sur la luminance. La couleur est obtenue en projetant les chrominances issues de la séquence compressée sur la luminance traitée. Deux extraits ont été choisis, l'un appartenant au fond, et l'autre à l'objet en mouvement.

[5] Nous ne parlerons pas ici d'une troisième catégorie de méthodes qui utilise l'interactivité entre la source et la destination, et qui requiert une retransmission des paquets perdus,

FIG. 5.3 – Comparaison de l'image décodée par l'algorithme MPEG1 classique à gauche, et par la méthode proposée OMD à droite, pour deux extraits de la séquence HALL (fond et objet en mouvement).

l'amélioration de la qualité visuelle des séquences :

– La première approche est de type "réseau". Elle a pour objectif de diminuer le mieux possible la quantité de pertes lors de la transmission en rendant le flot plus robuste. Lors du transport de vidéo, deux sources d'erreurs s'accumulent : les erreurs dues à la compression, et celles dues à la transmission. L'idée d'un codeur conjoint qui minimise simultanément ces deux erreurs a été longuement étudiée. Le théorème de Shannon affirme qu'il est possible de séparer le codage source et le codage canal en ayant une performance optimale [437]. Cependant, ce théorème suppose que la complexité et le temps de calcul peuvent être infinis, hypothèses non vérifiées pour la plupart des applications de transmission vidéo.

puisque cette démarche est peu adaptée à la transmission temps-réel des séquences.

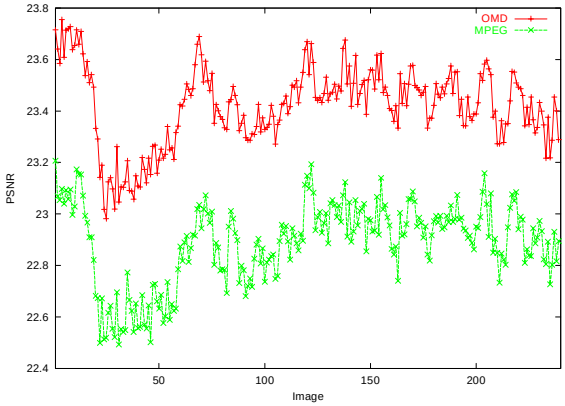


FIG. 5.4 – Evolution du PSNR sur la séquence HALL : comparaison de la méthode de décodage MPEG1 avec la méthode OMD.

– La seconde approche est de type "image". Elle a pour objectif de masquer au mieux l'effet de ces pertes lors du décodage et de réduire leur impact visuel sur la séquence restituée. Les techniques employées pour la correction des erreurs par dissimulations permettent en général de diminuer l'impact visuel des blocs qui ont été perdus, malgré les techniques de codage canal employées. La plupart des techniques pour les codeurs par blocs reposent sur trois hypothèses : premièrement, les images naturelles sont majoritairement composées de basses fréquences, deuxièmement l'oeil tolère plus facilement de la distorsion dans les hautes fréquences, et troisièmement si les paquets ont été ordonnés judicieusement, une tranche perdue doit être isolée spatialement et temporellement. Il s'agit par conséquent le plus souvent de techniques par interpolations appliquées aux macroblocs ou aux vecteurs-mouvement. Notons toutefois que les techniques abordées ne sont pas réservées aux pertes dues à des transmissions défectueuses.

De très nombreux travaux ont été fait dans ce domaine et l'objectif ici n'est pas de les recenser. Un etat de l'art complet est fait dans la thèse de Jung [407]. Notre contribution dans ce domaine a été d'adapter la méthode OMD de façon à prendre en compte les artefacts qui peuvent apparaître lors de l'acquisition, de la transmission, du stockage et de la lecture des vidéos. Cette adaptation et une étude complète de la méthode sont décrites dans la thèse de Jung [407] et dans l'article "Optimal decoder for block-transform based video coders" [14] donné en annexe G. Un exemple de décodage d'une trame MPEG1 au moyen d'OMD est présenté dans notre article en annexe G (figure 8) du document. Un perte de paquets a été simulée sur une trame MPEG, contenant la séquence

HALL. Le code binaire a subi des pertes aléatoires. Afin que la séquence puisse être décodable, les en-têtes des images ont été préservées. Nous remarquons que la perte d'une partie du code a engendré de fortes dégradations : nous pouvons supposer qu'une partie des vecteurs mouvement a été touchée, ainsi que les images d'erreurs.

5.4 Conclusion-Synthèse

Les travaux que nous avons effectué ont eu pour objectif de remettre en cause les filtres linéaires à reconstruction parfaite. Nous avons proposé un algorithme de décodage adapté à la problématique globale posée. En effet, le compresseur étant imposé, nous avons proposé l'optimisation du décodeur en prenant à la fois en compte les caractéristiques du bruit de quantification ainsi que les caractéristiques du système d'acquisition. Un point novateur de notre approche est l'introduction dans la chaîne de traitement du bruit de quantification et du bruit électronique caractérisés par des bruits bornés non stationnaires. Contrairement à un post-traitement "classique", notre méthode de décodage permet la conservation du train binaire initial JPEG, M-JPEG ou MPEG1, MPEG2.

5.5 Références

[381] G. Aubert and L. Vese, "A variational method in image recovery," *SIAM Journal of numerical analysis*, vol. 34, no. 5, pp. 1948–1979, octobre 1997.

[382] R. Azencot, "Markov fields and image analysis," in *Proceeding du 6ème congrès de Reconnaissance de Formes et Intelligence Artificielle, AFCET-INRIA*, novembre 1987, vol. 2, pp. 1183–1191.

[383] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of Royal Statistical Society*, vol. 2, pp. 192–236, 1974.

[384] J.C. Brailean and A.K. Katsaggelos, "Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences," *IEEE Transactions on Image Processing*, vol. 4, no. 9, pp. 1231–1251, septembre 1995.

[385] J.E. Caviedes and J. Jung, "No-reference metric for a video quality control loop," in *5th World Multiconference on Systemics, Cybernetics and Informatics (SCI), Orlando, Etats-Unis*, juillet 2001.

[386] P. Charbonnier, *Reconstruction d'Image : Régularisation avec Prise en Compte des Discontinuités*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, 1994.

[387] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *IEEE International Conference On Image Processing, Austin, Etats-Unis*, novembre 1994.

[388] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *Transactions on Image Processing*, vol. 6, no. 2, pp. 298–311, février 1997.

[389] B.S. Chen and C. W. Lin, "Multiscale wiener filter for the restoration of fractal signals : Wavelet filter bank approach," *IEEE Transactions on Signal Processing*, vol. 42, pp. 2972–2982, 1994.

[390] A.N. Delopoulos and S.D. Kollias, "Optimal filter banks for signal reconstrucɲtion from noisy subband component," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 212–224, 1996.

[391] A. Dembo and D. Malah, "Statistical design of analysis/synthesis systems with quantization," *IEEE Transactions on Signal Processing*, vol. 36, no. 3, pp. 328–341, 1988.

[392] R. Deriche and O. Faugeras, "Les edp en traitement des images et vision par ordinateur," *Traitement du Signal*, vol. 13, no. 6, 1996.

[393] C. Derviaux, F.X. Coudoux, M.G. Gazalet, and P. Corlay, "Blocking artifact reduction of dct coded image sequences using a visually adaptive postprocessing," in *International Conference on Image Processing, Lausanne, Suisse*, septembre 1996, vol. 1, pp. 5–8.

[394] M. Ebbecke, M.B.H. Ali, and A. Dengel, "Real time object detection, tracking and classification in monocular image sequences of road traffic scenes," in *IEEE International Conference on Image Processing, Santa Barbara, Etats-Unis*, octobre 1997, vol. 2, pp. 402–405.

[395] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restauration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 6, no. 6, pp. 721–741, 1984.

[396] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[397] K. Gosse, *Bancs de Filtres pour le Codage de Source : Reconstruction Parfaite ou Minimisation d'un Critère de Distorsion*, Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications de Paris, France, décembre 1996.

[398] K. Gosse and P. Duhamel, "Perfect reconstruction versus mmse filter banks in source coding," *IEEE Transactions on Signal Processing*, vol. 45, no. 9, pp. 2188–2202, septembre 1997.

[399] C. Gu, T. Ebrahimi, and M. Kunt, "Morphological object segmentation and tracking for content based video coding," *Multimedia Communication and Video Coding, New-York*, novembre 1995.

[400] R.A. Haddad and N. Uzun, "Modeling, analysis and compensation of quantization effects in m-band subband codecs," in *IEEE International Conference on Audio Speech and Signal Processing*, 1994, vol. 3, pp. 145–148.

[401] Haddad and N. Uzun, "Cyclostationnary modeling, analysis and optimal compensations of quantization errors in subband coders," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, pp. 2109–2119, septembre 1995.

[402] T. Hsung, D. Lun, and W. Siu, "A deblocking technique for jpeg decoded image using wavelets transform modulus maxima representation," in *IEEE International Conference on Image Processing, Lausanne, Suisse*, septembre 1996, vol. 2, pp. 561–564.

[403] International Organization for Standardization, "Coding of moving pictures and audio," in *ISO/IEC JTC 1/SC 29/WG 11, Maui, Etats-Unis*, décembre 1999.

[404] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Region-based active contours for video object segmentation with camera compensation," in *IEEE International Conference on Image Processing, Thessalonique, Grèce*, octobre 2001.

[405] S. Jehan-Besson, M. Barlaud, and G. Aubert, "An object-based motion method for video coding," in *IEEE International Conference on Image Processing, Thessalonique, Grèce*, octobre 2001.

[406] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Deformable regions driven by an eulerian accurate minimization method for image and video segmentation, application to face detection in color video sequences," in *European Conference on Computer Vision, Copenhagen, Danemark*, 2002.

[407] J. Jung, *OMD : Une Méthode de Décodage Optimal Orientée Objet pour les Séquences d'Images*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, 2000.

[408] J. Kalifa, S. Mallat, F. Falzon, and B. Rougé, "High resolution satellite image restoration with frames," in *SPIE Conference on Wavelets Applications in Signal and Image Processing IV*, 1996.

[409] A.K. Katsaggelos, J. Biemond, R.M. Mersereau, and R.W. Schafer, "A general formulation of constrained iterative image restoration," in *IEEE International Conference on Audio Speech and Signal Processing*, mars 1985, vol. 3, pp. 700–703.

[410] A.K. Katsaggelos, J. Biemond, R.M. Mersereau, and R.W. Shafer, "A regularized iterative image restoration algorithm," *IEEE Transactions on Signal Processing*, vol. 39, pp. 914–929, avril 1991.

[411] A.K. Katsaggelos Ed., *Digital Image Restoration*, vol. 23, Springer Series in Information Sciences, new york, springer edition, 1991.

[412] P. Kornprobst, R. Deriche, and G. Aubert, "Image coupling, restoration and enhancement via pde's," in *IEEE International Conference on Image Processing, Santa Barbara, Etats-Unis*, octobre 1997, pp. 458–461.

[413] I. Kompatsiaris and M.G. Strintzis, "Spatiotemporal segmentation and tracking of objects in image sequences," in *IEEE International Conference on Image Processing, Kobe, Japon*, octobre 1999.

[414] P. Kornprobst, *Contribution à la restauration d'images et à l'analyse de séquences : Approches variationnelles et solutions de viscosité*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, 1998.

[415] P. Kornprobst, R. Deriche, and G. Aubert, "Image sequence restoration : A pde based coupled method for image restoration and motion segmentation," in *European Conference on Computer Vision, Fribourg, Suisse*, juin 1998, pp. 548–562.

[416] J. Kovacevic and M. Vetterli, *Wavelet and Subband Coding*, Englewood Cliffs, N.J. : Prentice-Hall, 1995.

[417] J. Kovacevic, "Subband coding systems incorporating quantizer models," *IEEE Transactions on Image Processing*, vol. 4, no. 5, pp. 543–553, 1995.

[418] C. Latry and B. Rougé, "Spot5 thr mode," in *SPIE Visual Communication and Image Processing, USA*, 1998.

[419] Y.L. Lee, H.C. Kim, and H.W. Park, "Blocking effect reduction of jpeg images by signal adaptative filtering," *IEEE Transactions on Image Processing*, vol. 7, no. 2, février 1998.

[420] P.L. Lions, S. Osher, and L. Rudin, "Denoising and deblurring images with constrained nonlinear partial differential equations," in *preprint of SINUM*.

[421] S. Mallat, "A theory for multiresolution signal decomposition : The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, 1989.

[422] F. Marques and J. Llach, "Tracking of generic object for video object generation," in *International Conference on Image Processing, Chicago, Etats-Unis*, octobre 1998.

[423] S. Minami and A. Zakhor, "An optimization approach for removing blocking effects in transform coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 2, pp. 74–82, avril 1995.

[424] M. Minoux, *Mathematical Programming : Theory and Algorithms*, Wiley : New York, 1986.

[425] P. Moulin, "A multiscale relaxation algorithm for snr maximization in nonorthogonal subband coding," *IEEE Transactions on Image Processing*, 1995.

[426] S. Osher and L. Rudin, "Feature oriented image enhancement using shock filters," *SIAM Journal of numerical analysis*, vol. 27, no. 4, pp. 919–940, août 1990.

[427] H. Paek, J.W. Park, and S.U. Lee, "Non-iterative post-processing technique for transform coded image sequence," in *IEEE International Conference on Image Processing, Washington, Etats-Unis*, octobre 1995, vol. 3, pp. 208–211.

[428] H.W. Park and Y.L. Lee, "A postprocessing method for reducing quantization effects in low bit-rate moving picture coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 161–171, février 1999.

[429] P. Raffy, *Modélisation, Optimisation et Mise en Oeuvre de Quantificateurs Bas Débits pour la Compression d'Images Utilisant une Transformée en Ondelettes*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, décembre 1997.

[430] B. Ramamurthi and A. Gersho, "Nonlinear space-variant postprocessing of block coded images," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 5, octobre 1986.

[431] A. Rao, D. Miller, K. Rose, and A. gersho, "Generalized vector quantization : Jointly optimal quantization and estimation," in *IEEE International Symposium on Information Theory*, Whistler B.C., Canada, septembre 1995.

[432] A. Rao, D. Miller, K. Rose, and A. gersho, "Generalized vq method for combined compression and estimation," in *IEEE International Conference on Audio Speech and Signal Processing, Atlanta, Etats-Unis*, mai 1996.

[433] B. Rougé, J. Kalifa, and S. Mallat, "Restoration d'images par paquets d'ondelettes," in *Seizième colloque GRETSI*, 1997.

[434] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *IEEE International Conference on Image Processing, Austin, Etats-Unis*, novembre 1994, vol. 1, pp. 31–35.

[435] L. Rudin, S. Osher, and C. Fu, "Total variation based image restoration of noisy blurred images," in *preprint of SINUM*.

[436] S. Sista and R.L. Kashyap, "Unsupervised video segmentation and object tracking," in *IEEE International Conference on Image Processing, Kobe, Japon*, octobre 1999.

[437] C.E. Shannon, "A mathematical theory of communication," in *Bell Syst. Tech. J.*, 1948, vol. 27, pp. 379–423, 623–656.

[438] S. Tramini, *Problèmes Inverses et EDP pour le Décodage et la Déconvolution d'Images*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, novembre 1999.

[439] H. Uzawa, *Iterative Methods for Concave Programming - Chap.10 in Studies in Linear and Nonlinear Programming*, Stanford University Press, 1958.

[440] J. Vass, K. Palaniappan, and X. Zhuang, "Automatic spatio-temporal video sequence segmentation," in *International Conference on Image Processing, Chicago, Etats-Unis*, octobre 1998, vol. 1, pp. 958–962.

[441] L. Vese, *Problèmes Variationnels et EDP pour l'Analyse d'Image et l'Evolution de Courbes*, Ph.D. thesis, Université de Nice-Sophia Antipolis, France, 1996.

[442] J.L.H. Webb, "Postprocessing to reduce blocking artifacts for low bit-rate video coding using chrominance information," in *IEEE International Conference on Image Processing, Santa Barbara, Etats-Unis*, octobre 1997, vol. 2, pp. 9–12.

[443] Y. Yang, N. Galatsanos, and A. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 6, pp. 421–432, décembre 1993.

[444] Y. Yang and N. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," *IEEE Transactions on Image Processing*, vol. 6, no. 10, pp. 1345–1357, octobre 1997.

[445] S.F. Yang, "Linear restoration of block-transform coded motion video," in *Visual Communications and Image Processing, San José, Etats-Unis*, février 1997.

[446] S. Zhong, "Image coding with optimal reconstruction," in *IEEE International Conference on Image Processing, Santa Barbara, Etats-Unis*, octobre 1997, vol. 1, pp. 161–164.

Chapitre 6

Conclusion et projet de recherche

Conclusion Mon travail de recherche s'est axé durant ces dix dernières années sur la compression et le codage des images et des vidéos dans le cadre de l'analyse multirésolution. Notre effort de recherche a porté sur la conception d'un système de compression/décompresion multirésolution adapté aux caractéristiques des signaux sources et des canaux de transmission. Ma recherche s'est décomposée en une partie fondamentale et une partie applicative avec des retombées industrielles dans le domaine de la normalisation JPEG-2000 et dans le domaine spatial puisque notre algorithme de compression a été retenu pour être embarqué dans les futurs génération de satellites d'observation de la Terre lancés par le Centre National d'Etudes Spatiales.

Dans ce document j'ai développé les travaux de recherches les plus importants que j'ai effectués depuis mon entrée au CNRS en faisant référence dans certains cas à quelques unes de mes publications que j'ai annexé au document. Principalement, les travaux que j'ai mené sur l'insertion de la TRANSFORMEE EN ONDELETTES dans un schéma de compression d'images ont permis la construction des filtres dits "9-7" qui fournissent à l'heure actuelle les meilleurs résultats en compression d'image. Ces filtres ont fait l'objet d'une implantation sur circuit intégré commercialisé par Analog Device sous le nom de ADV601 ainsi que d'une implantation sur DSP par Texas Instrument. Ils sont de plus retenus par toutes les propositions de pointe pour la future norme de compression d'images fixes JPEG-2000 et constituent un des filtres de référence pour cette nouvelle norme. L'algorithme d'allocation des ressources binaires que nous avons développé dans ce cadre multirésolution est basé sur des modèles théoriques de distorsion et de débit. Cet algorithme d'allocation dynamique a été introduit dans le cadre du codage par transformée en ondelettes "au fil de l'eau" et permet soit un contrôle du débit binaire, soit un contrôle de la "qualité" image en terme d'erreur quadratique moyenne ou de rapport signal-à-bruit. Cette mé-

thode nous a servi comme brique de base pour construire un algorithme de codage source/canal par descriptions multiples efficace pour la transmission de données images ou vidéos sur des réseaux Internet ou encore de troisième génération (UMTS). De plus, nous avons montré qu'il était aussi efficace pour la compression de maillages 3D et pouvait concurrencer en terme de compromis DEBIT-DISTORSION-COMPLEXITE les meilleures méthodes de compression actuelles telles que le standard JPEG-2000.

Parallèlement à cette approche scalaire, nous avons développé une nouvelle méthode géométrique de codage des coefficients d'ondelettes qui utilise la QUANTIFICATION VECTORIELLE par réseaux réguliers de points. Nos travaux se sont orientés suivant plusieurs objectifs et principalement, nous avons proposé des solutions pour le dénombrement des vecteurs dans un réseau régulier, pour l'indexage de ces vecteurs dans le cas d'une distribution Gaussienne généralisée et pour la modélisation de la distorsion de quantification et du débit dans un cadre non asymptotique. Ces travaux ont conduit à la définition d'un algorithme de quantification vectorielle algébrique à entropie contrainte pour la compression multirésolution des images.

L'aspect DECODAGE est aussi très important. Les travaux que nous avons effectué sur ce sujet de recherche ont eu pour objectif de remettre en cause les filtres linéaires à reconstruction parfaite. Nous avons proposé un algorithme de décodage adapté à la problématique globale posée. En effet, le compresseur étant imposé, nous avons proposé l'optimisation du décodeur en prenant à la fois en compte les caractéristiques du bruit de quantification ainsi que les caractéristiques du système d'acquisition. Un point novateur de notre approche a été l'introduction dans la chaîne de traitement du bruit de quantification et du bruit électronique caractérisés par des bruits bornés non stationnaires. Contrairement à un post-traitement "classique", notre méthode de décodage permet la conservation du train binaire initial JPEG, M-JPEG ou MPEG1, MPEG2.

Projet de recherche De façon naturelle, mes recherches s'orientent vers la compression de données IMAGES et MULTIMEDIA pour leur transmission sur des réseaux à pertes et leur archivage. Mon projet de recherche est résumé ci-après avec ses applications directes et ses implications dans le monde socio-économique. Il concerne l'étude de méthodes de compression performantes et adaptées aux données multimédia images et vidéos (2D ou 3D), en vue de leur transmission sur des systèmes de communication à pertes ou de leur archivage. Il a pour objectif le regroupement et l'interaction entre les différents domaines de recherches que sont la représentation multirésolution des images 2D (ou 3D) et vidéo 2D+t (ou 3D+t), le codage de source, le codage de canal. Il porte principalement sur trois points :

– **Introduction de méthodes de compression basées sur des ana-**

lyses multirésolutions de seconde génération pour la compression de séquences d'images dans l'optique MPEG-4. L'accès à l'information est devenu un élément essentiel de l'activité économique mondiale, et l'une des raisons de l'expansion et du dynamisme du secteur des télécommunications. Les évolutions technologiques sur le plan des composants et des infrastructures ont conduit d'une part au développement extraordinaire du réseau fédérateur Internet, et d'autre part à l'introduction de plus en plus importante des contenus audiovisuels au niveau des serveurs d'information, rendant l'information essentiellement multimédia et permettant la création de contenus de plus en plus riches. La nouvelle norme MPEG-4 a été définie pour le codage et la transmission bas débit de flux multimédia. Elle utilise la notion d'objet audio/vidéo mais ne spécifie pas la façon dont ces objets sont générés. Ce nouveau standard en cours d'élaboration permet la représentation, le codage, la manipulation du contenu des documents vidéo. Il implique l'existence de méthodes de segmentation automatique performantes de la vidéo numérique pour des applications de visiophonie ou téléconférence, de production et diffusion de programmes TV, de télésurveillance active, de télémédecine, etc. La segmentation est un préalable à une représentation et compression orientée objet. Les travaux que nous nous proposons de poursuivre dans ce domaine concernent à la fois la segmentation spatio-temporelle et la compression des différents éléments décrivant la scène contenue par la séquence. Les méthodes de compression que nous envisageons, contrairement aux standards existants (MPEG), sont basées sur des méthodes de type ondelettes telles que celle proposée dans la future norme JPEG-2000 et plus précisément sur des transformées en ondelettes mise en œuvre par des schéma *liftings* de deuxième génération adaptés à des supports ou des grilles d'échantillonnage non rectangulaires. L'objectif étant de générer une analyse multirésolution sur des objets de forme quelconque. L'introduction d'une telle transformée dans un système de compression orienté objet de type MPEG-4, conjointement à la segmentation spatio-temporelle, permettrait de toute évidence d'améliorer l'efficacité de la méthode de compression en terme de compromis QUALITE/DEBIT et en terme de FONCTIONNALITES MPEG-4. Les fonctionnalités MPEG-4 (c'est-à-dire, par exemple, la manipulation d'objets rendue possibles à l'utilisateur ou encore la SCALABILITE) seront d'une grande importance pour les systèmes de communications virtuelles tels que les futures vidéo-conférences où les interlocuteurs d'une réunion peuvent tous se voir en même temps et dans un même lieu. En effet, la manipulation d'objets en mouvement sera rendue possible de part l'opération de segmentation spatio-temporelle et son codage efficace grâce à des méthodes qui prennent en compte la nature spatio-fréquentielle de l'objet (forme et texture de l'objet). Aujourd'hui la future norme MPEG-4 suscite l'inté-

rêt de nombreux chercheurs. Sa conception définitive est loin d'être finie et laisse la place à de nombreux progrès nécessaires dans ce domaine de recherche si l'on veut disposer un jour de systèmes de compression vidéo performants dans des plages de débits relativement faibles.

– **L'optimisation conjointe du codeur et du décodeur avec prise en compte du canal de transmission dans un schéma de compression d'images pour les télécommunications** (Internet, radio-mobile 3G et 4G). Le problème du contrôle et de la réduction des erreurs liées à la compression et à la transmission prend une ampleur de plus en plus importante, du fait de la multiplication actuelle des transmissions vidéo sur les réseaux. L'effet de ces pertes sur les séquences comprimées peut être catastrophique et la perte d'un paquet est d'autant plus grave en raison de l'intégration de la compression dans le schéma de transmission. En effet, plus le taux de compression est fort, plus la quantité de données perdues sera grande. De plus, la perte d'un paquet dans une séquence implique souvent une propagation temporelle des artefacts du fait de la prédiction temporelle présente dans les codages de type MPEG. Le problème se complique également si l'on considère des applications temps-réel, interactives ou encore des communications multipoints. Aujourd'hui, le problème de la transmission de données image sur des réseaux hétérogènes reçoit une attention considérable. En effet, un scénario typique qui nécessiterait la transmission de données au travers de deux canaux de capacités différentes (le premier ayant une capacité plus grande que le second) entraînerait la perte de paquets transmis de façon à s'adapter à la capacité du second canal. Si le réseau est capable d'effectuer des traitements adaptés selon les paquets, alors, l'utilisation d'un codage source de type multirésolution, allouant les priorités des paquets en fonction de leur contenu, serait une solution évidente. Cependant, si le réseau ne regarde pas le contenu des paquets il y a une sélection aléatoire des paquets supprimés, entraînant une dégradation non contrôlée du signal reconstruit. C'est essentiellement le type de pertes produites par le réseau Internet. Il n'est alors pas évident de concevoir un codeur source adapté à ce problème. L'idée fondamentale pour contrôler ce problème serait de permettre au décodeur de retrouver l'information perdue, ou une partie de cette information, connaissant les paquets qui lui sont parvenus. Ce problème correspond à la généralisation du problème de "descriptions multiples". Introduire du codage de source robuste aux erreurs de transmission ou pertes par paquets, ou encore des techniques de répartition optimisée de bits entre source et canal, traduit une tendance très forte à l'heure actuelle en télécommunication qui est celle de mieux adapter codage source et codage canal. Dans un souci de performance et/ou de réduction de complexité, notre projet en codage/décodage conjoint peut nous amener,

en fonction de ces résultats, à franchir une étape supplémentaire avec l'introduction d'un système codeur/décodeur unique pour source et canal. Ceci aurait naturellement un impact conséquent sur les équipements de réseaux et des terminaux futurs. Il est donc nécessaire d'identifier les verrous technologiques dans le domaine de la représentation et de la compression des signaux vidéo dans un contexte de transmission sur réseaux hétérogènes. Ces verrous concernent la REPRESENTATION SCALABLE des flux, caractéristique clé pour permettre une adaptation des débits des flux aux caractéristiques non stationnaires et hétérogènes des réseaux d'une part et aux caractéristiques hétérogènes des terminaux d'autre part.

– **La compression géométrique de maillages 3D ou encore de séquences d'images 3D et la prise en compte de la qualité image**. L'objectif est le développement d'algorithmes de compression géométrique de MAILLAGES 3D pour l'imagerie en général et plus précisemment l'imagerie médicale, et la définition de méthodes d'EVALUATION OBJECTIVE de la qualité des images décomprimées. Le domaine est d'actualité du fait du développement des méthodes d'imagerie numérique qui entraîne un accroissement constant du volume des données à archiver et à transmettre. Ce volume considérable est l'un des facteurs de ralentissement de la diffusion des systèmes de communication et d'archivage (PACS). Si de nombreux algorithmes ont été proposés pour coder les images en général et si même des standards comme JPEG et MPEG ont été adoptés, l'utilisation en imagerie médicale de méthodes de codage irréversible pose toujours des problèmes non résolus. L'évaluation de l'application des méthodes irréversibles aux images et aux séquences d'images est donc une étape primordiale pour leur acceptation et leur utilisation en clinique par exemple. L'inadéquation des critères classiques de QUALITE D'UNE IMAGE, les difficultés et les limites de tests psychophysiques ne permettent pas d'évaluer complètement les performances d'une méthode de codage et encore moins d'inclure les outils, même d'évaluation classique, dans le processus de codage. Notre projet concerne l'étude et la mise au point de stratégies d'évaluation de ces méthodes de compression et vise à trouver un compromis entre le taux de compression et la qualité de l'image restituée. Notre contribution consisterait essentiellement à définir une chaîne complète de compression la mieux adaptée aux maillage 3D. Nous envisageons de nous intéresser en particulier à la représentation multiéchelles des maillages ainsi qu'à la notion de zones d'intérêt et de contours dans l'image. Ceci permettra la définition d'une approche de compression davantage ciblée sur l'application. Les méthodes que nous souhaitons mettre en œuvre seront basées sur des analyses multirésolutions *liftings* 3D de seconde génération.

Evidemment ces recherches se feront en collaboration avec d'autres laboratoires Français, Européens et Internationaux.

Troisième partie

Mes publications significatives

Liste des articles significatifs

1. **Page 217 :** M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies "Image Coding Using Wavelet Transform", IEEE Transaction on Image Processing, Vol.1, No.2, pp. 205-220, avril 1992.

2. **Page 235 :** C. Parisot, M. Antonini, M. Barlaud, "3D Scan Based Wavelet Transform and Quality Control for Video Coding", EURASIP Journal on Applied Signal Processing vol.2003, no.2, pp. 56-65, janvier 2003.

3. **Page 247 :** M. Barlaud, P. Solé, T. Gaidon, M. Antonini, P. Mathieu, "Pyramidal Lattice Vector Quantization for Multiscale Image Coding", IEEE Transaction on Image Processing, Vol.3, No.4, pp. 367-381, juillet 1994.

4. **Page 265 :** P. Loyer, J.M. Moureaux, M. Antonini, "Lattice Codebook Enumeration for Generalized Gaussian Source", IEEE Transactions on Information Theory vol.49, no.2, pp. 521-528, février 2003.

5. **Page 275 :** P. Raffy, M. Antonini, M. Barlaud, "Distortion-Rate Models for Entropy Coded Lattice Vector Quantization", IEEE Transactions on Image Processing, Vol.9, No.12, pp. 2006-2017, décembre 2000.

6. **Page 289 :** F. Davoine, M. Antonini, J.M. Chassery, M. Barlaud, "Fractal Image Compression based on Delaunay Triangulation and Vector Quantization", IEEE Transaction on Image Processing, Vol.5, No.2, Special issue on vector quantization, pp. 338-346, février 1996.

7. **Page 301 :** J. Jung, M. Antonini, M. Barlaud, "Optimal Decoder for Block–Transform Based Video Coders", IEEE Transactions on Multimedia vol.5, no.2, pp. 145-160, juin 2003.

Annexe A

Image Coding Using Wavelet Transform

IEEE Transactions on Image Processing vol.1, no.2, pp. 205-220, avril 1992

J'ai co-écrit cet article avec Michel Barlaud, Pierre Mathieu et Ingrid Daubechies.

Résumé Dans cet article, nous avons proposé un nouveau schéma de compression d'images qui prend en compte les caractéristiques psychovisuelles à la fois dans le domaine spatial et dans le domaine transformé. Cette nouvelle méthode consiste en deux étapes. Premièrement, elle utilise une transformée en ondelettes dans le but d'obtenir un ensemble de sous-bandes biorthogonales : l'image originale est décomposée à différentes résolutions en utilisant un algorithme pyramidal. La décomposition s'effectue le long des lignes et des colonnes de l'image et maintient constant le nombre de pixels contenu dans l'image. Deuxièmement, selon la théorie débit-distorsion de Shannon, les coefficients d'ondelettes sont quantifiés vectoriellement au moyen d'un dictionnaire multirésolution. De plus, nous avons proposé une procédure d'allocation des débits adaptée à la sensibilité de l'œil humain, et développé un schéma de transmission progressif dans le but de permettre au récepteur de reconnaître une image le plus rapidement possible avec un coût de transmission minimal. C'est dans cet article que nous avons intoduit les filtres "9-7" et étudié leurs performances dans un schéma complet de compression.

Image Coding Using Wavelet Transform

Marc Antonini, Michel Barlaud, *Member, IEEE,* Pierre Mathieu, and Ingrid Daubechies, *Member, IEEE*

***Abstract*—Image compression is now essential for applications such as transmission and storage in data bases. This paper proposes a new scheme for image compression taking into account psychovisual features both in the space and frequency domains; this new method involves two steps. First, we use a *wavelet transform* in order to obtain a set of biorthogonal subclasses of images; the original image is decomposed at different scales using a pyramidal algorithm architecture. The decomposition is along the vertical and horizontal directions and maintains constant the number of pixels required to describe the image. Second, according to Shannon's rate distortion theory, the wavelet coefficients are *vector quantized* using a multiresolution codebook. Furthermore, to encode the wavelet coefficients, we propose a noise shaping bit allocation procedure which assumes that details at high resolution are less visible to the human eye. Finally, in order to allow the receiver to recognize a picture as quickly as possible at minimum cost, we present a progressive transmission scheme. It is shown that the wavelet transform is particularly well adapted to progressive transmission.**

***Keywords*—Wavelet, biorthogonal wavelet, multiscale pyramidal algorithm, vector quantization, noise shaping, progressive transmission.**

I. Introduction

In many different fields, *digitized images* are replacing conventional analog images as photograph or x-rays. The volume of data required to describe such images greatly slow transmission and makes storage prohibitively costly. The information contained in the images must, therefore, be compressed by extracting only the visible elements, which are then encoded. The quantity of data involved is thus reduced substantially.

A fundamental goal of data compression is to reduce the bit rate for transmission or storage while maintaining an acceptable fidelity or image quality. Compression can be achieved by transforming the data, projecting it on a basis of functions, and then encoding this transform. Because of the nature of the image signal and the mechanisms of human vision, the transform used must accept nonstationarity and be well localized in both the space and frequency domains. To avoid redundancy, which hinders compression, the transform must be at least biorthogonal and lastly, in order to save CPU time, the corresponding algorithm must be fast. The two-dimensional wavelet transform defined by Meyer and Lemarié [31], [24], [25], together with its implementation as described by Mallat [27], satisfies each of these conditions.

The compression method we have developed associates a wavelet transform and a vector quantization coding scheme. The wavelet coefficients are coded considering a noise shaping bit allocation procedure. This technique exploits the psychovisual as well as statistical redundancies in the image data, enabling bit rate reduction.

Section II describes the wavelet transforms used in this paper. After a quick review of wavelets in general, we explain in more detail the properties and construction of regular biorthogonal wavelet bases. We then extend this one-dimensional construction to a two-dimensional scheme with separable filters. The new coding scheme is next presented in Section III. We focus particularly in this section on the statistical properties of wavelet coefficients, on the asymptotic coding gain that can be achieved using vector quantization in the subimages, and on the optimal allocation across the subimages. Experimental results are given in Section IV for images taken within and outside of the training set.

II. Wavelets

A. A Short Review of Wavelet Analysis

Wavelets are functions generated from one single function ψ by dilations and translations

$$\psi^{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right).$$

(For this introduction we assume t is a one-dimensional variable). The *mother wavelet* ψ has to satisfy $\int dx\ \psi(x) = 0$, which implies at least some oscillations. (Technically speaking, the condition on ψ should be $\int d\omega\ |\Psi(\omega)|^2\ |\omega|^{-1} < \infty$, where Ψ is the Fourier transform of ψ; if $\psi(t)$ decays faster than $|t|^{-1}$ for $t \to \infty$, then this condition is equivalent to the one above). The definition of wavelets as dilates of one function means that high frequency wavelets correspond to $a < 1$ or narrow width, while low frequency wavelets have $a > 1$ or wider width.

The basic idea of the wavelet transform is to represent any arbitrary function f as a superposition of wavelets. Any such superposition decomposes f into different scale levels, where each level is then further decomposed with a resolution adapted to the level. One way to achieve such a decomposition writes f as an integral over a and b of $\psi^{a,b}$ with appropriate weighting coefficients [22]. In practice, one prefers to write f as a discrete superposition (sum rather than integral). Therefore, one introduces a discre-

Manuscript received February 7, 1990; revised March 26, 1991.

M. Antonini, M. Barlaud, and P. Mathieu are with LASSY 13S CNRS, Universite de Nice-Sophia Antipolis, 06560 Valbonne, France.

I. Daubechies is with AT&T Bell Laboratories, Murray Hill, NJ 07974.

IEEE Log Number 9106073.

tization, $a = a_0^m$, $b = nb_0a_0^m$, with $m, n \in \mathbb{Z}$, and $a_0 > 1$, $b_0 > 0$ fixed. The wavelet decomposition is then

$$f = \sum c_{m,n}(f)\psi_{m,n} \tag{1}$$

with $\psi_{m,n}(t) = \psi^{a_0^m, nb_0a_0^m}(t) = a_0^{-m/2}\psi(a_0^{-m}t - nb_0)$. Decompositions of this type were studied in [14], [15]. For $a_0 = 2$, $b_0 = 1$ there exist very special choices of ψ such that the $\psi_{m,n}$ constitute an orthonormal basis, so that

$$c_{m,n}(f) = \langle \psi_{m,n}, f \rangle = \int dx\, \psi_{m,n}(x) f(x)$$

in this case. Different bases of this nature were constructed by Stromberg [36], Meyer [31], Lemarié [24], Battle [7], and Daubechies [16]. All these examples correspond to a multiresolution analysis, a mathematical tool invented by Mallat [27], which is particularly well adapted to the use of wavelet bases in image analysis, and which gives rise to a fast computation algorithm.

In a multiresolution analysis, one really has *two* functions: the mother wavelet ψ and a *scaling function* ϕ. One also introduces dilated and translated versions of the scaling function, $\phi_{m,n}(x) = 2^{-m/2}\phi(2^{-m}x - n)$. For fixed m, the $\phi_{m,n}$ are orthonormal. We denote by V_m the space spanned by the $\phi_{m,n}$; these spaces V_m describe successive approximation spaces, $\cdots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \cdots$, each with resolution 2^m. For each m, the $\psi_{m,n}$ span a space W_m which is exactly the orthogonal complement in V_{m-1} of V_m; the coefficients $\langle \psi_{m,n}, f \rangle$, therefore, describe the information lost when going from an approximation of f with resolution 2^{m-1} to the coarser approximation with resolution 2^m. All this is translated into the following algorithm for the computation of the $c_{m,n}(f) = \langle \psi_{m,n}, f \rangle$ (for more details, see [27]):

$$\begin{aligned} c_{m,n}(f) &= \sum_k g_{2n-k}a_{m-1,k}(f) \\ a_{m,n}(f) &= \sum_k h_{2n-k}a_{m-1,k}(f) \end{aligned} \tag{2}$$

where $g_l = (-1)^l h_{-l+1}$ and $h_n = 2^{1/2}\int dx\, \phi(x - n)\phi(2x)$. In fact the $a_{m,n}(f)$ are coefficients characterizing the projection of f onto V_m. If the function f is given in sampled form, then one can take these samples for the highest order resolution approximation coefficients $a_{0,n}$, and (2) describes a subband coding algorithm on these sampled values, with low-pass filter h and high-pass filter g. Because of their association with orthonormal wavelet bases, these filters give exact reconstruction, i.e.:

$$a_{m-1,l}(f) = \sum_n [h_{2n-l}a_{m,n}(f) + g_{2n-l}c_{m,n}(f)]. \tag{3}$$

Most of the orthonormal wavelet bases have infinitely supported ψ, corresponding to filters h and g with infinitely many taps. The construction in [16] gives ψ with finite support, and therefore, corresponds to FIR filters. It follows that the orthonormal bases in [16] correspond to a subband coding scheme with exact reconstruction property, using the same FIR filters for reconstruction as for decomposition. Such filters are well known since the work of Smith and Barnwell [35] and of Vetterli [37]. The extra ingredient in the orthonormal wavelet decomposition is that it writes the signal to be decomposed as a superposition of reasonably *smooth* elementary building blocks. The filters must satisfy the additional condition:

$$\prod_{k=1}^{\infty} H(2^{-k}\xi)$$

decay faster than $C(1 + |\xi|)^{-\epsilon-0.5}$ as $|\xi| \to \infty$, for some $\epsilon > 0$, where

$$H(\xi) = 2^{-1/2}\sum_n h_n e^{-jn\xi}.$$

This extra regularity requirement is usually not satisfied by the exact reconstruction filters in the ASSP literature.

B. Applications of Wavelet Bases to Image Analysis

1) Biorthogonal Wavelet Bases: Since images are mostly smooth (except for occasional edges) it seems appropriate that an exact reconstruction subband coding scheme for image analysis should correspond to an orthonormal basis with a reasonably smooth mother wavelet. In order to have fast computation, the filters should be short (short filters lead to less smoothness, however, so they cannot be too short). On the other hand it is desirable that the FIR filters used be linear phase, since such filters can be easily cascaded in pyramidal filter structures without the need for phase compensation. Unfortunately, there are no nontrivial orthonormal linear phase FIR filters with the exact reconstruction property [35], regardless of any regularity considerations. The only symmetric exact reconstruction filters are those corresponding to the Haar basis, i.e., $h_0 = h_1 = 2^{1/2}$ and $g_0 = -g_1 = 2^{1/2}$, with all other h_n, $g_n = 0$.

One can preserve linear phase (corresponding to symmetry for the wavelet) by relaxing the orthonormality requirement, and using biorthogonal bases. It is then still possible to construct examples where the mother wavelets have arbitrarily high regularity.

In such a scheme, we still decompose as in (2), but reconstruction becomes

$$a_{m-1,l}(f) = \sum_n [\tilde{h}_{2n-l}a_{m,n}(f) + \tilde{g}_{2n-l}c_{m,n}(f)] \tag{4}$$

where the filters $\tilde{h}$, $\tilde{g}$ may be different from h, g. In order to have exact reconstruction, we impose:

$$\begin{aligned} \tilde{g}_n &= (-1)^n h_{-n+1} \\ g_n &= (-1)^n \tilde{h}_{-n+1} \end{aligned} \quad \sum_n h_n\tilde{h}_{n+2k} = \delta_{k,0}. \tag{5}$$

So far, we have not performed anything differently from the usual exact reconstruction subband coding schemes with synthesis filters different from the decomposition filters. If the filters satisfy the additional condition that:

$$\prod_{k=1}^{\infty} \tilde{H}(2^{-k}\xi) \quad \text{and} \quad \prod_{k=1}^{\infty} H(2^{-k}\xi) \tag{6a}$$

decay faster than $C(1 + |\xi|)^{-\epsilon-0.5}$ as $|\xi| \to \infty$, for some $\epsilon > 0$, where

$$\tilde{H}(\xi) = 2^{-1/2} \sum_n \tilde{h}_n e^{-jn\xi} \quad H(\xi) = 2^{-1/2} \sum_n h_n e^{-jn\xi} \tag{6b}$$

then we can give the following interpretation to (2) and (4). Define functions ϕ and $\tilde{\phi}$ by

$$\phi(x) = \sum_n h_n \phi(2x - n) \quad \text{and} \quad \tilde{\phi}(x) = \sum_n \tilde{h}_n \tilde{\phi}(2x - n).$$

Their Fourier transforms are exactly the infinite products (6a), and they are, therefore, well-defined square integrable functions, compactly supported if the filters h and $\tilde{h}$ are FIR. Define also

$$\psi(x) = \sum_n g_n \phi(2x - n) \quad \text{and} \quad \tilde{\psi}(x) = \sum_n \tilde{g}_n \tilde{\phi}(2x - n).$$

Then, the $a_{m,n}(f)$ and $c_{m,n}(f)$ in (2) can be rewritten as:

$$a_{m,n}(f) = \langle \phi_{m,n}, f \rangle = 2^{-m/2} \int dx\, \phi_{m,n}(x) f(x)$$

$$c_{m,n}(f) = \langle \psi_{m,n}, f \rangle = 2^{-m/2} \int dx\, \psi_{m,n}(x) f(x)$$

and reconstruction is simply:

$$f = \sum_{m,n} \langle \psi_{m,n}, f \rangle \tilde{\psi}_{m,n}. \tag{7}$$

The filter bank structure with the associating wavelets and scaling functions is depicted on the following subband coding scheme (Fig. 1).

If the infinite products in (6a) decay even faster than imposed above, then ϕ and $\tilde{\phi}$ and consequently ψ and $\tilde{\psi}$ will be reasonably smooth. Note that (7) is very similar to the orthonormal decomposition described in Section II-A; the only difference is that the expansion of f with respect to the basis $\tilde{\psi}_{m,n}$ uses coefficients computed via the dual basis $\psi_{m,n}$ with $\tilde{\psi}$ different from ψ. This interpretation is not possible for all exact reconstruction subband coding schemes; in particular, convergence of the infinite products (6a) is only possible if

$$\sum_n h_n = 2^{1/2} \quad \text{and} \quad \sum_n \tilde{h}_n = 2^{1/2}.$$

Moreover, (7) can only hold if

$$\sum_n (-1)^n h_n = 0 \quad \text{and} \quad \sum_n (-1)^n \tilde{h}_n = 0.$$

Most exact reconstruction subband coding schemes do not satisfy these conditions.

Biorthogonal bases of wavelets have recently been constructed, with regularity simultaneously but independently, by Cohen, Daubechies and Feauveau [12] and by Herley and Vetterli [38]. Reference [12] contains a detailed mathematical study, with proofs that, under the conditions stated above, the wavelets do indeed constitute numerically stable bases (Riesz bases) and a discussion of necessary and sufficient conditions for regularity. In [18]

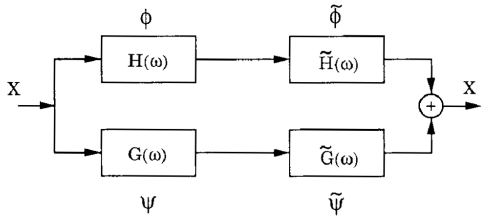


Fig. 1. Filter bank structure and the associating wavelets.

Feauveau explores the construction from the point of view of multiresolution spaces rather than from the filters. Basically one has two hierarchies of spaces in the biorthogonal case, each corresponding to one pair of filters.

It is shown in [12] that arbitrarily high regularity can be achieved by both ψ and $\tilde{\psi}$, provided one chooses sufficiently long filters. In particular, if the functions ψ and $\tilde{\psi}$ are, respectively, $(k - 1)$ and $(\tilde{k} - 1)$ times continuously differentiable, then the trigonometric polynomials $H(\xi)$ and $\tilde{H}(\xi)$ have to be divisible by $(1 + e^{-j\xi})^k$ and $(1 + e^{-j\xi})^{\tilde{k}}$, respectively, so that the length of the corresponding filters h, $\tilde{h}$ has to exceed k, $\tilde{k}$.

By (5), divisibility of $\tilde{H}(\xi)$ by $(1 + e^{-j\xi})^{\tilde{k}}$ means that ψ will have $\tilde{k}$ consecutive moments zero:

$$\int dx\, x^l \psi(x) = 0, \qquad \text{for} \quad l = 0, 1, \cdots, \tilde{k} - 1.$$

For more details concerning this discussion, see [12].

It is well known (and it can easily be checked by using Taylor expansions) that if ψ has $\tilde{k}$ moments zero, then the coefficients $\langle \psi_{m,n}, f \rangle$ will represent functions f, which are $\tilde{k}$ times differentiable, with a high compression potential (many coefficients will be negligibly small).

Many examples of biorthogonal wavelet bases with reasonably regular ψ and $\tilde{\psi}$ can be constructed; for our applications, the regularity of the elementary building blocks $\tilde{\psi}_{m,n}$, which is linked to the number of zero moments of ψ, is more important than the regularity of the $\psi_{m,n}$ or the number of zero moments of $\tilde{\psi}$. Within the limits imposed by the support widths, we will, therefore, try to choose $\tilde{k}$ as large as possible.

In terms of trigonometric polynomials $H(\xi)$ and $\tilde{H}(\xi)$, the exact reconstruction requirement condition on h and $\tilde{h}$ given in (5) reduces to (for symmetric filters)

$$H(\xi)\tilde{H}(\xi) + H(\xi + \pi)\tilde{H}(\xi + \pi) = 1. \tag{8}$$

Together with divisibility of H and $\tilde{H}$, respectively, by $(1 + e^{-j\xi})^k$ and $(1 + e^{-j\xi})^{\tilde{k}}$, this leads to (see [12])

$$H(\xi)\tilde{H}(\xi) = \cos(\xi/2)^{2l} \left[\sum_{p=0}^{l-1} \binom{l - 1 + p}{p} \cdot \sin(\xi/2)^{2p} + \sin(\xi/2)^{2l} R(\xi) \right] \tag{9}$$

where $R(\xi)$ is an odd polynomial in $\cos(\xi)$, and where $2l = k + \tilde{k}$ (symmetry of h and $\tilde{h}$ forces $k + \tilde{k}$ to be even).

TABLE I
FILTER COEFFICIENTS FOR THE SPLINE FILTERS WITH $l = 3$, $k = 4$, $\tilde{k} = 2$

n	0	± 1	± 2	± 3	± 4
$2^{-1/2}h_n$	45/64	19/64	$-1/8$	$-3/64$	3/128
$2^{-1/2}\tilde{h}_n$	1/2	1/4	0	0	0

Many examples are possible. We have studied in particular the following three examples, which belong to three different families.

2) Spline Filters: One can choose, e.g., $R \equiv 0$, with $\tilde{H}(\xi) = \cos(\xi/2)^{\tilde{k}} e^{-j\kappa\xi/2}$ where $\kappa = 0$ if $\tilde{k}$ is even, $\kappa = 1$ if $\tilde{k}$ is odd. This corresponds to the filters called "spline filters" in [12] (because the corresponding function $\tilde{\phi}$ is a B-spline function) or "binomial filters" in [38] (because the $\tilde{h}$ are simply binomial coefficients). It then follows that:

$$H(\xi) = \cos(\xi/2)^{2l-\tilde{k}} e^{j\kappa\xi/2} \cdot \left[\sum_{p=0}^{l-1} \binom{l-1+p}{p} \sin(\xi/2)^{2p}\right]. \quad (10)$$

We have looked at one example from this family; it corresponds to $l = 3$, $\tilde{k} = 2$. The coefficients h_n and $\tilde{h}_n$ are listed in Table I; the corresponding scaling functions and wavelets are plotted in Fig. 2.

It is clear that the two filters in the first example have very uneven length. This is typical for all the examples in this family of "spline filters."

3) A Spline Variant with Less Dissimilar Lengths: This family still uses $R \equiv 0$ in (9), but factorizes the right-hand side of (9), breaking up the polynomial of degree $l-1$ in $\sin(\xi/2)$ into a product of two polynomials in $\sin(\xi/2)$ with real coefficients, one to be allocated to H, the other to $\tilde{H}$, so as to make the lengths of h and $\tilde{h}$ as close as possible.

The example presented here is the "smallest" one in this family (shortest h and $\tilde{h}$); it corresponds to $l = 4$ and $k = 4$. The filter coefficients are listed in Table II; the corresponding scaling functions and wavelets are plotted in Fig. 3.

Note that, unlike examples 1 and 3 where the $2^{-1/2}h_n$, $2^{-1/2}\tilde{h}_n$ are rational, the entries in Table II are truncated decimal expansions of irrational numbers. The functions ϕ in examples 1 and 2 look very similar (compare Figs. 2(a) and 3(a)); a more detailed analysis shows that the one in example 2 is more regular, however. Both correspond to 4 vanishing moments for $\tilde{\psi}$.

4) Filters Close to Orthonormal Filters: Finally, there exist many examples for which $R \neq 0$. In particular there exists a special choice of R for which the two filters are very close to each other, and both very close to an orthonormal wavelet filter.

Surprisingly, for the first example of this series, one of the two filters is a Laplacian pyramid filter proposed in [9]. It corresponds to $l = 2$, $k = 2$ and $R(\xi) = 48\cos(\xi)/175$. The filter coefficients are listed in Table III; the corresponding scaling functions and wavelets are plotted in Fig. 4. It is clear that the scaling functions ϕ and $\tilde{\phi}$ are very similar, corresponding to very similar ψ and $\tilde{\psi}$. Note that in this case, the filter coefficients are again rational.

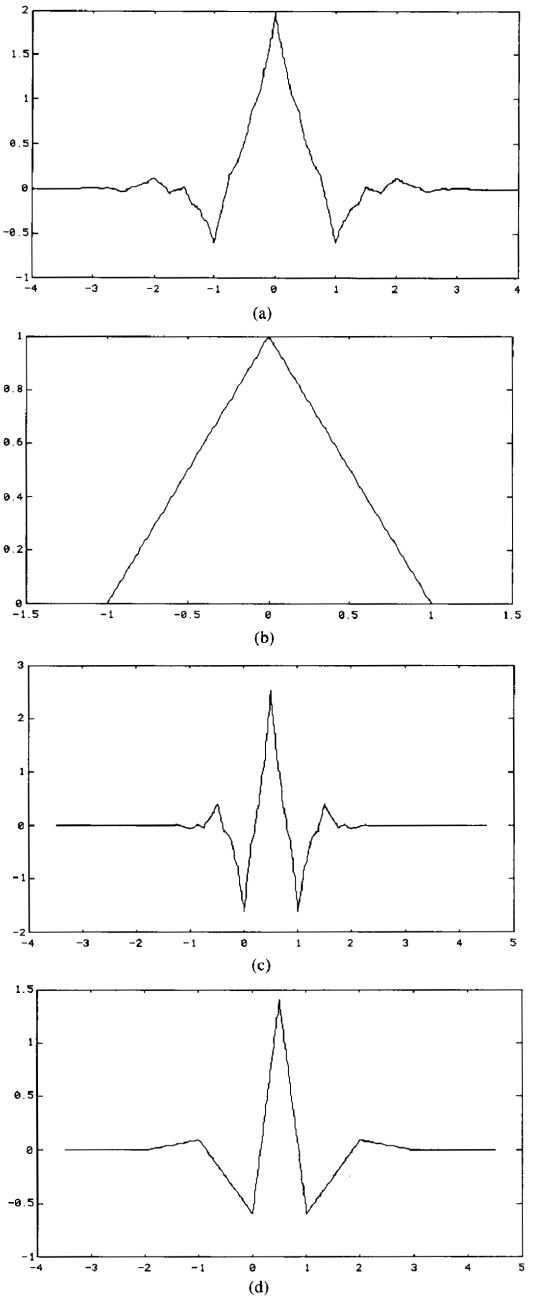


Fig. 2. Scaling functions ϕ, $\tilde{\phi}$ and wavelets ψ, $\tilde{\psi}$ for example 1 (spline filters with $l = 3$, $k = 4$, $\tilde{k} = 2$). (a) Scaling function ϕ. (b) Scaling function $\tilde{\phi}$. (c) Wavelet ψ. (d) Wavelet $\tilde{\psi}$.

TABLE II
FILTER COEFFICIENTS FOR THE SPLINE VARIANT WITH LESS DISSIMILAR LENGTHS, WITH $l = 4 = k$, $\tilde{k} = 4$

n	0	± 1	± 2	± 3	± 4
$2^{-1/2}h_n$	0.602 949	0.266 864	−0.078 223	−0.016 864	0.026 749
$2^{-1/2}\tilde{h}_n$	0.557 543	0.295 636	−0.028 772	−0.045 636	0

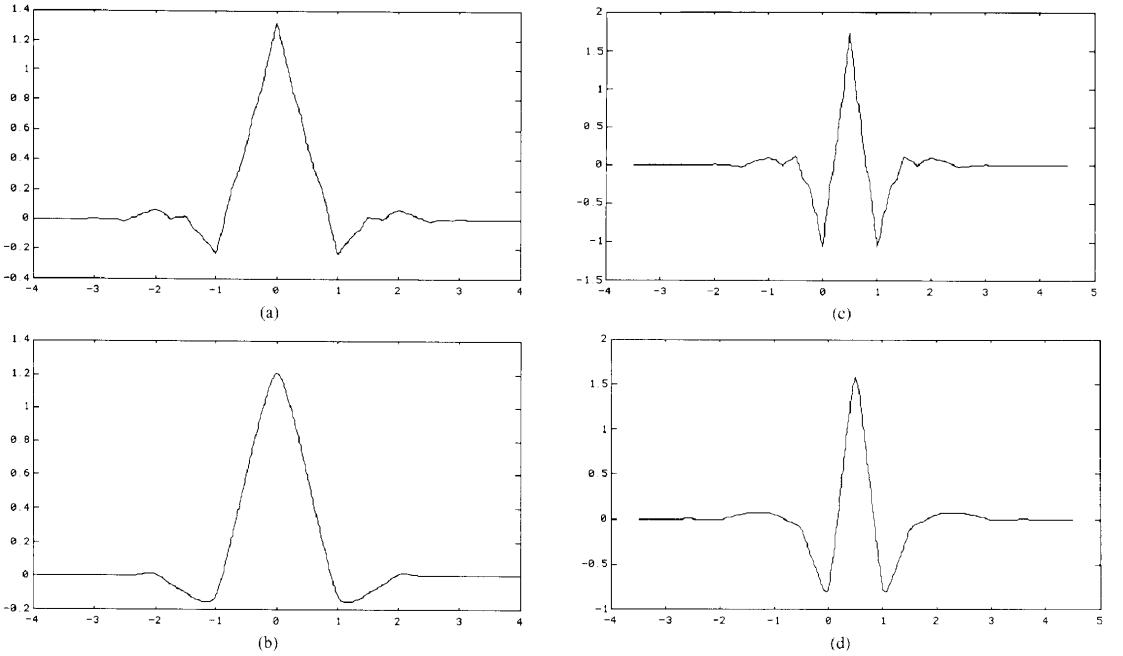


Fig. 3. Scaling functions ϕ, $\tilde{\phi}$ and wavelets ψ, $\tilde{\psi}$ for example 2 (spline variant with less dissimilar lengths; $l = 4 = k$, $\tilde{k} = 4$). (a) Scaling function ϕ. (b) Scaling function $\tilde{\phi}$. (c) Wavelet ψ. (d) Wavelet $\tilde{\psi}$.

TABLE III
FILTER COEFFICIENTS FOR EXAMPLE 3. THE ENTRIES ARE RATIONAL, AND THE TWO FILTERS ARE VERY CLOSE. THE h-FILTER COINCIDES WITH A LAPLACIAN PYRAMID FILTER PROPOSED IN [9]. IN THIS CASE $l = 2 = k$, $\tilde{k} = 2$

n	0	± 1	± 2	± 3	± 4
$2^{-1/2}h_n$	0.6	0.25	−0.05	0	0
$2^{-1/2}\tilde{h}_n$	17/28	73/280	−3/56	−3/280	0

The two biorthogonal filters in this example are both close to an orthonormal wavelet filter of length 6 constructed in [17], where it was called a "coiflet." Being an orthonormal wavelet filter, the coiflet is nonsymmetric. The filters in this example are shorter than in examples 1 and 2, but k is also smaller. The next example in this family corresponds to $k = 4$ (and $l = 4$); the filters h and $\tilde{h}$ then have length 9 and 15; they are both close to a coiflet of length 12.

5) Extension to the Two-Dimensional Case: There exist various extensions of the one-dimensional wavelet transform to higher dimensions. We follow Mallat [27] and use a two-dimensional wavelet transform in which horizontal and vertical orientations are considered preferential.

In two-dimensional wavelet analysis one introduces, like in the one-dimensional case, a scaling function $\phi(x, y)$ such that:

$$\phi(x, y) = \phi(x)\,\phi(y) \tag{11}$$

where $\phi(x)$ is a one-dimensional scaling function.

Let $\psi(x)$ be the one-dimensional wavelet associated with the scaling function $\phi(x)$. Then, the three two-dimensional wavelets are defined as:

$$\begin{aligned} \psi^H(x, y) &= \phi(x)\,\psi(y) \\ \psi^V(x, y) &= \psi(x)\,\phi(y) \\ \psi^D(x, y) &= \psi(x)\,\psi(y). \end{aligned} \tag{12}$$

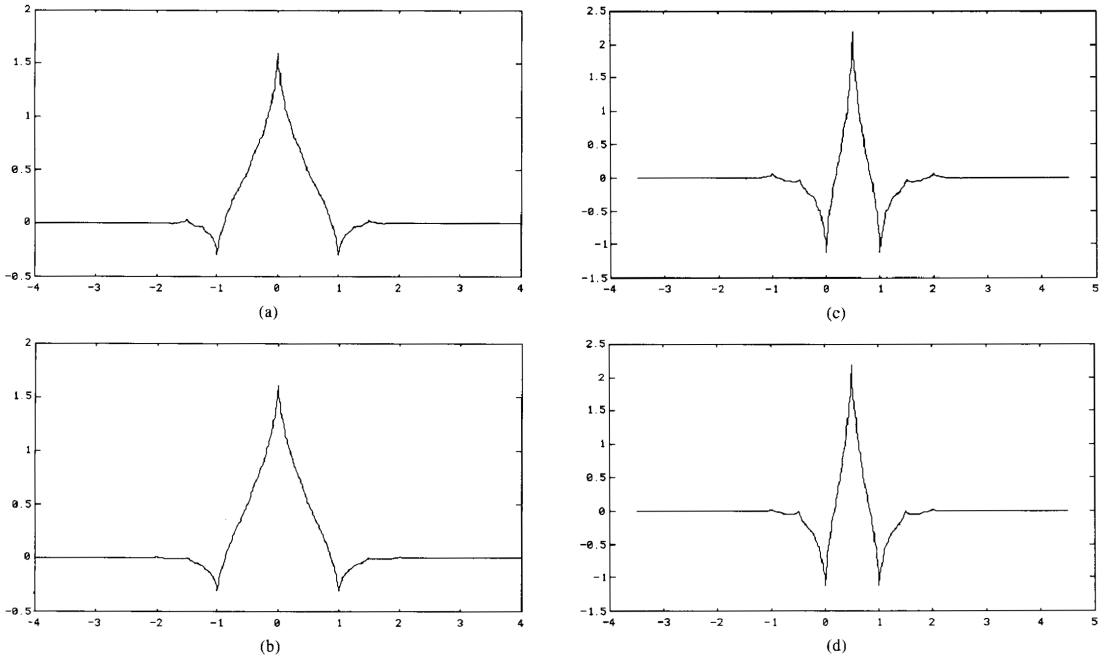


Fig. 4. Scaling functions ϕ, $\tilde{\phi}$ and wavelets ψ, $\tilde{\psi}$ for example 3 (biorthogonal filters close to an orthonormal wavelet filter, $l = 2 = k$, $\tilde{k} = 2$). (a) Scaling function ϕ. (b) Scaling function $\tilde{\phi}$. (c) Wavelet ψ. (d) Wavelet $\tilde{\psi}$.

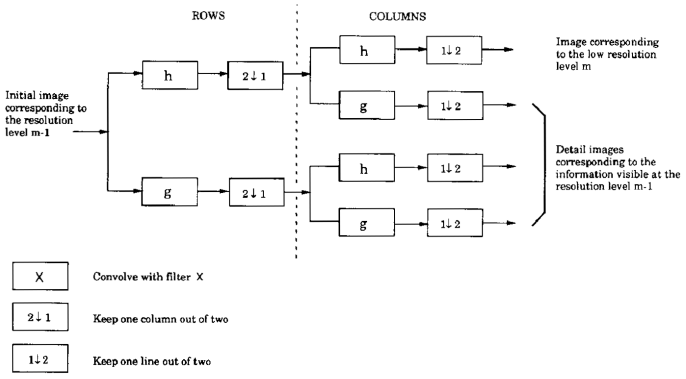


Fig. 5. One stage in a multiscale image decomposition.

Fig. 5 represents one stage in a *multiscale pyramidal* decomposition of an image: wavelet coefficients of the image are computed, as in the one-dimensional case (Sections II-A and II-B.1), using a subband coding algorithm. The filters h and g are one-dimensional filters. This decomposition provides subimages corresponding to different resolution levels and orientations (see Fig. 6). The reconstruction scheme of the image is presented Fig. 7.

To compare the three different filters presented in this paper, we have decomposed the image Lena (Fig. 16) with each of these filters. The results are presented in Fig. 8. In Fig. 8(a) we can see the normalized detail subimages at different resolution levels $m = 1$, $m = 2$, and $m = 3$ (wavelet coefficients) and in Fig. 8(b) the low resolution level subimages.

III. Image Coding Application

A. Statistical Properties of Wavelet Coefficients

The performance of a coder used for a given resolution and direction can be determined by the statistics of the corresponding subimage, i.e., its probability density function (PDF).

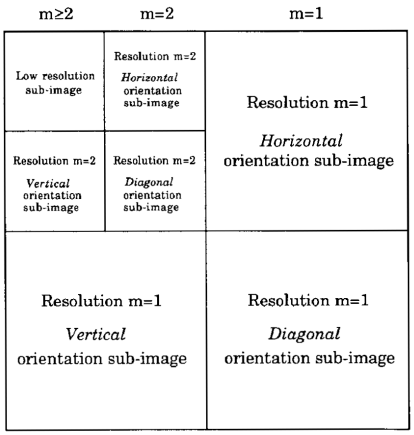


Fig. 6. Image decomposition.

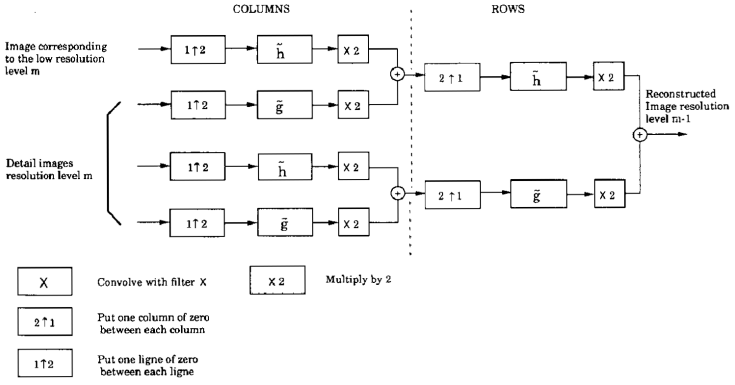


Fig. 7. One stage in a multiscale image reconstruction.

A typical PDF and different approximations are given in Fig. 9, where we plot the true PDF for resolution level $m = 1$ and direction d = vertical together with three model functions: a Gaussian, a Laplacian, and an intermediate function, the so-called generalized Gaussian [2].

This generalized Gaussian law is given explicitly by

$$p_{m,d}(x) = a_{m,d} \exp\left(-|b_{m,d}x|^{r_{m,d}}\right)$$

with

$$a_{md} = \frac{b_{m,d} r_{m,d}}{2\Gamma\left(\frac{1}{r_{m,d}}\right)} \quad \text{and} \quad b_{m,d} = \frac{1}{\sigma_{m,d}} \frac{\Gamma\left(\frac{3}{r_{m,d}}\right)^{1/2}}{\Gamma\left(\frac{1}{r_{m,d}}\right)^{1/2}} \tag{13}$$

where $\sigma_{m,d}$ is the standard deviation of the subimage (m, d), and $\Gamma(\cdot)$ is the usual Gamma function.

The general formula (13) contains the other two examples as particular cases:

- $r_{m,d} = 2$ leads to the well-known Gaussian PDF;
- $r_{m,d} = 1$ leads to a Laplacian PDF.

The variance of this approximation model is set equal to the variance of the corresponding subimage. Thus the parameter $r_{m,d}$ is computed in order to match the real PDF using the well-known chi-squared test. In this case the optimum parameter was 0.7. Other experiments for other resolutions (except the lowest resolution) lead to very similar results.

We can see in Fig. 9 that the real PDF (scale $m = 1$ and vertical orientation) is closely approximated by a generalized Gaussian law with parameter $r_{1,v} = 0.7$.

B. Encoding of Wavelet Coefficients Using Vector Quantization

Different techniques involving vector or scalar quantization can be used to encode wavelet coefficients.

According to Shannon's rate distortion theory, better results are always obtained when vectors rather than sca-

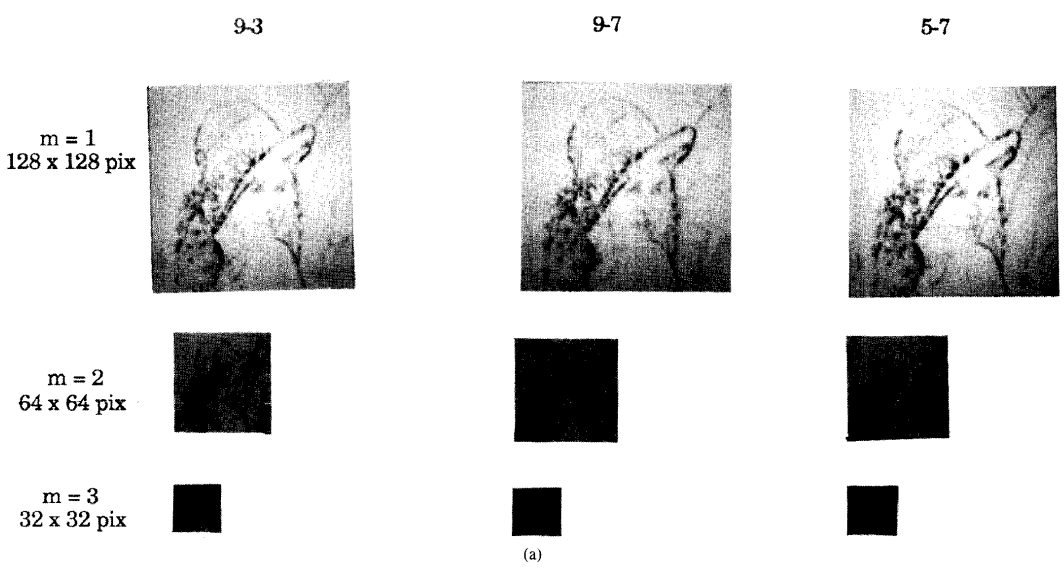


Fig. 8. Comparison among the different subimages. (a) Comparison among the normalized detail subimages. (b) Comparison among the low resolution level subimages.

lars are encoded. Therefore, the present application uses vector quantization.

1. Principle of Vector Quantization: Developed recently by Gersho and Gray (1980) [20], [21], vector quantization has proven to be a powerful tool for digital image compression [4], [29], [30], [32], [39]. The principle involves encoding a sequence of samples (vector) rather than encoding each sample individually. Encoding is performed by approximating the sequence to be coded by a vector belonging to a catalogue of shapes, usually known as a codebook.

The codebook is created and optimized using the well-known Linde–Buzo–Gray (LBG) [26] classification algorithm with a mean squared error (MSE) criterion. This algorithm is designed to perform a classification based on a training set comprised of vectors belonging to different images; it converges iteratively toward a locally optimal codebook.

Each of the vectors in the codebook is indexed. At the encoding stage, the index of the vector in the codebook most closely describing (in terms of MSE criterion) the sample set to be encoded is selected to represent this set. Of course, in order to reconstruct the sample set, the decoder must have the same codebook as the coder.

The encoding/decoding scheme depicted in Fig. 10 was proposed in [29] and [30] for orthonormal wavelets.

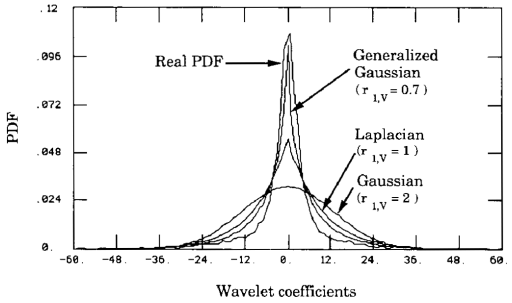


Fig. 9. Real PDF of subimage at scale $m = 1$ for vertical orientation, and its different approximations.

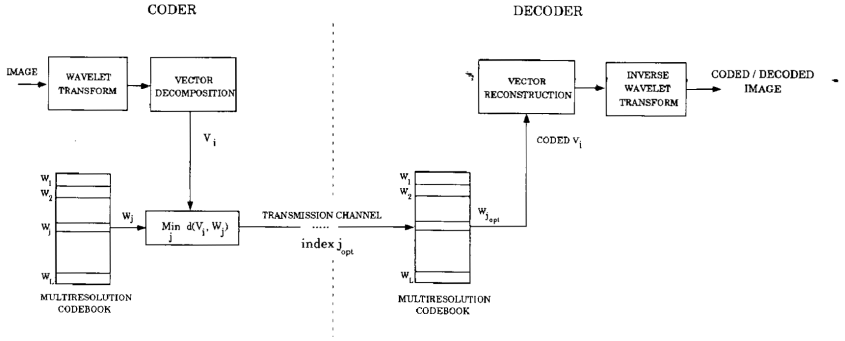


Fig. 10. Encoding/decoding scheme.

2) Comparative Performances of Vector Quantization (VQ) and Scalar Quantization (SQ): According to [3], [13], [19], [43], [30] the asymptotic lower bound distortion gain obtained when VQ, rather than SQ, is applied to a subimage is expressed as:

$$G_{m,d}^{VQ} \geq \frac{2^{-c}}{(c+1)A(k_{m,d}, c)} \times \frac{\left[\int [p_{m,d}(x)]^{1/(c+1)}\, dx\right]^{(c+1)}}{\left[\int [p_{m,d}(x)]^{k_{m,d}/(c+k_{m,d})}\, dx\right]^{(c+k_{m,d})}} \tag{14}$$

for a subimage corresponding to resolution m and direction d. $p_{m,d}(x)$ is the PDF of wavelet coefficients of the subimage with resolution m and direction d.

Here, the MSE criterion is used as a distortion measure ($c = 2$). The values of $A(k_{m,d}, 2)$ used are the upper bounds of the MSE computed and tabulated by Conway and Sloane for vector size $k_{m,d}$ [13]. This formula gives an indication of the minimum theoretical gain that can be obtained.

However, this approximation is valid only for small quantization errors, i.e., for a high bit rate $R_{m,d}$. Thus the gain $G_{m,d}^{VQ}$ only gives here an asymptotic indication.

In Fig. 11, the curves of $G_{m,d}^{VQ}$ are plotted as a function of the vector dimension $k_{m,d}$ for the Laplacian, Gaussian,

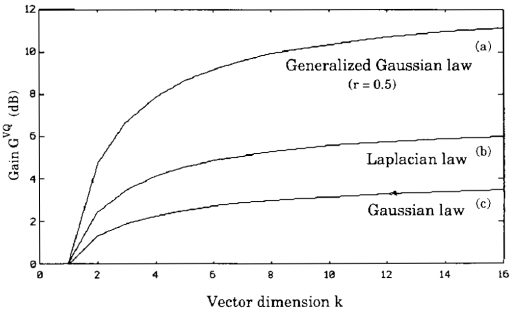


Fig. 11. Asymptotic lower bound distortion gain $G_{m,d}^{VQ}$ = function $(k_{m,d})$.

and generalized Gaussian approximation laws, and for a subimage at scale $m = 1$ and vertical orientation. Experimental results are closely matched by the theoretical results for a generalized Gaussian law with $r_{m,d} = 0.7$ except for the lower subband. Therefore, all computations based on this approximation law show that, in each subband, VQ outperforms SQ (see Fig. 11).

In summary VQ performs better for coding wavelet coefficients.

3) Generation of a Multiresolution Codebook: The preceding paragraph explained why VQ outperforms other methods. Nonetheless, major problems are encountered in the VQ of images.

- It is impossible to create a universal codebook (efficient for each image to be encoded).

- The LBG algorithm smooths high frequencies (loss of resolution).
- There is a trade-off between low distortion and high compression rate (computational cost).
- It is not easy to take into account the properties of the human visual system [28], [33].

The use of the wavelet transform (i.e., multiresolution) is one way of overcoming these different problems.

The wavelet decomposition of an image enables the generation of a codebook containing two-dimensional vectors for *each resolution level and preferential direction* (horizontal, vertical, and diagonal). Each of these subcodebooks (see Fig. 12) is generated using the LBG algorithm.

- The training set is comprised of vectors belonging to different images corresponding to the resolution and orientation under consideration.
- The initial codebook is generated by splitting the centroid (center of gravity) of this training set [21].

A multiresolution codebook can thus be obtained by assembling all of these resulting subcodebooks. Each subcodebook has a low distortion level and contains few words, which clearly facilitates the search for the best coding vector; the coding computational load is reduced, because only the appropriate subcodebook (resolution direction) of the multiresolution codebook is checked for each input vector. In addition, the quality of the coded image is better. The multiresolution codebook is depicted in Fig. 12.

Global codebook design has drawbacks in that it results in edge smoothing while the proposed method preserves edges. In fact, each subcodebook contains the shape of the wavelet coefficients which are most highly representative in terms of the MSE criterion.

Since the spatial and frequency aspects of the image are taken into account in the wavelet decomposition, the classification and search during the encoding of a subimage vector can be achieved using a simple criterion such as least mean squares. This frees us from using distortion measurements such as weighted least mean squares or other measurements involving perceptual factors. These algorithms are indeed costly in computation time.

C. Optimal Bit Allocation

Multiresolution exploits the eye's masking effects, and therefore, enables us to refine and select the type of coding according to the resolution level and the contour orientation. Although a flat noise shape minimizes the MSE criterion, it is generally not optimal for a subjective quality of image. To apply *noise shaping* across the VQ subimages, we define a total weighted MSE distortion $D_T^*(R_T)$ ((17)) for a total bit rate R_T ((18)).

Let us define $D_{m,d}(R_{m,d})$ the average distortion in the coding of the subimage (m, d) for $R_{m,d}$ bits per pixel:

$$D_{m,d}(R_{m,d}) = E(|x - q(x)|^c) = d(x, q(x)) \qquad c \geq 1 \tag{15}$$

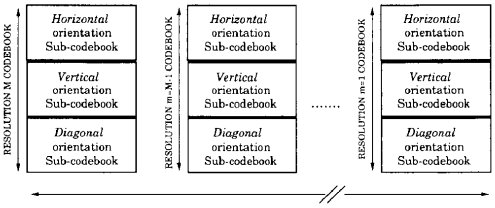


Fig. 12. Multiresolution codebook.

for all coefficients x belonging to the subimage, $q(x)$ being the quantization of x.

Total distortion of the image for a total rate of R_T bits per pixel is then given by:

$$D_T(R_T) = \frac{1}{2^{2M}} D_M^{SQ}(R_M^{SQ}) + \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{d=1}^{3} D_{m,d}(R_{m,d}) \tag{16}$$

where $D_M^{SQ}(R_M^{SQ})$ corresponds to the distortion in the subimage of lowest resolution M (texture subimage).

The problem of finding an optimal bit assignment (in bits per pixel) for each subimage vector quantizer is then formulated as:

$$\min_{R_{m,d}} \left[D_T^*(R_T) = \frac{1}{2^{2M}} D_M^{SQ}(R_M^{SQ}) + \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{d=1}^{3} D_{m,d}(R_{m,d}) \times B_{m,d} \right] \tag{17}$$

$$\textbf{subject to: } R_T = \frac{1}{2^{2M}} R_M^{SQ} + \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{d=1}^{3} R_{m,d} \tag{18}$$

where R_M^{SQ} corresponds to the bit allocation, in bits per pixel, of lowest resolution M subimage.

Assignment of the weights is based on the fact that the human eye is not equally sensitive to signals at all spatial frequencies. On the basis of contrast sensitivity data collected by Campbell and Robson [10], and to obtain a controlled degree of noise shaping across the subimages, we consider a function $B_{m,d}$ such that:

$$B_{m,d} = \gamma^m \log (\sigma_{m,d}^{2\beta_{m,d}}) \tag{19}$$

where $\sigma_{m,d}$ is the standard deviation corresponding to subimage (m, d) and the values of γ and $\beta_{m,d}$ are chosen experimentally in order to match human vision.

$D_T^*(R_T)$ is the total weighted encoding distortion function, and M is the lowest resolution considered.

The expression of $D_{m,d}(R_{m,d})$ is given by [19]

$$D_{m,d}(R_{m,d}) = 2^{-cR_{m,d}} \times \alpha_{m,d}(p, c), \qquad c \geq 1$$

with

$$\alpha_{m,d}(p, c) = A(k_{m,d}, c) \times \left[\int [p_{m,d}(x)]^{k_{m,d}/(c + k_{m,d})} \, dx \right]^{(c + k_{m,d})}. \tag{20}$$

This minimization problem can be solved by using Lagrangian multipliers. Using this technique, we must solve the following equation:

$$\frac{\partial}{\partial R_{m,d}}\left[D_T^*(R_T) - \lambda\left(R_T - \frac{1}{2^{2M}} R_M^{SQ} - \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{d=1}^{3} R_{m,d}\right)\right] = 0 \tag{21}$$

where λ is a Lagrangian multiplier.

Using (17) and (20), this equation becomes:

$$\frac{\partial}{\partial R_{m,d}}\left[\frac{1}{2^{2m}} D_M^{SQ}(R_M^{SQ}) + \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{d=1}^{3} (2^{-cR_{m,d}} \alpha_{m,d}(p, c) B_{m,d}) - \lambda\left(R_T - \frac{1}{2^{2M}} R_M^{SQ} - \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{d=1}^{3} R_{m,d}\right)\right] = 0. \tag{22}$$

Taking the partial derivative with respect to $R_{m,d}$ yields an expression for $R_{m,d}$ in terms of λ:

$$R_{m,d} = \frac{1}{c} \log_2 \left[\frac{(c \ln 2)\alpha_{m,d}(p, c) B_{m,d}}{\lambda}\right]. \tag{23}$$

By substituting (23) into the constraint (18) of the minimization problem we obtain an expression of the Lagrangian multiplier λ

$$\lambda = c \ln 2 \left[2^{-c(R_T - (1/4^M) R_M^{SQ})} \prod_{m=1}^{M} \prod_{d=1}^{3} \cdot [\alpha_{m,d}(p, c) B_{m,d}]^{1/4^m}\right]^{4^M/4^M - 1}. \tag{24}$$

Finally, substituting λ into (23) results in an expression of the optimal bit assignment $R_{m,d_{opt}}$ (in bits per pixel (bpp)) to the vector quantizer of subimage (m, d):

$$R_{m,d_{opt}} = \frac{4^M R_T - R_M^{SQ}}{4^M - 1} + \frac{1}{c} \log_2 \left[\frac{\alpha_{m,d}(p, c) B_{m,d}}{\left[\prod_{m'=1}^{M} \prod_{d'=1}^{3} [\alpha_{m',d'}(p, c) B_{m',d'}]^{1/4^{m'}}\right]^{4^M/4^M - 1}}\right]. \tag{25}$$

This expression requires the knowledge of the subimage's PDF's.

The optimal distortion of the quantizer, $D^*_{T_{opt}}(R_T)$, is then computed by combining (25) and (17). We find:

$$D^*_{T_{opt}}(R_T) = \frac{1}{2^{2M}} D_M^{SQ}(R_M^{SQ}) + \frac{4^M - 1}{4^M} 2^{-c(4^M R_T - R_M^{SQ})/4^M - 1} \cdot \left[\prod_{m=1}^{M} \sum_{d=1}^{3} [\alpha_{m,d}(p, c) B_{m,d}]^{1/4^m}\right]^{4^M/4^M - 1}. \tag{26}$$

Finally, bit allocation which is a function of the image will be transmitted as side information requiring only a few bits.

IV. Experimental Results

The images used are sampled 256 by 256 black and white images. The intensity of each pixel is coded on 256 grey levels (8 bpp).

The numerical evaluation of the coder's performance is achieved by computing the peak signal-to-noise ratio (PSNR) between the original image and the coded image.

For each coded image, we can use a variable length code. We also give the corresponding $\mathcal{R}_T$ if an optimal entropy coding was performed, defined as follows.

To the L codewords $w_j; j = 1, 2, \cdots, L$ of the vector quantizer corresponds to L regions (clusters) of $\mathbb{R}^k$, $\mathcal{P}_j; j = 1, 2, \cdots, L$. The jth region is defined by

$$\mathcal{P}_j = \{x \in \mathbb{R}^k / Q(x) = w_j\}$$

and represents the subset of vectors of $\mathbb{R}^k$ which are well matched by the codeword w_j of the codebook.

Thus for each resolution and direction, we can introduce the average information of the codebook, called the entropy measure:

$$\mathcal{R}_{m,d} = -\frac{1}{k_{m,d}} \times \sum_{j=1}^{L} p(w_j) \log_2 p(w_j) \text{ bpp}$$

where $p(w_j)$ is the probability of selecting the source vector w_j, belonging to the codebook at scale m and corresponding to the orientation d, during the coding of the image (m, d).

Then, as in (18), $\mathcal{R}_T$ is the sum of the estimated entropy in each subimage as follows:

$$\mathcal{R}_T = \frac{1}{2^{2M}} \mathcal{R}_M^{SQ} + \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{d=1}^{3} \mathcal{R}_{m,d} \text{ bpp}.$$

The vector quantizer used is a *full search* quantizer, i.e., during the coding, all of the vectors in the subcodebook corresponding to the resolution and direction to be encoded are searched. The selection criterion used is the MSE criterion.

A. Comparison Between the Different Wavelets

In the following, we present results obtained with the Lena image (image within the training set) for a real bit rate of 1 bpp and using the three different filters proposed in Section II-B. (Fig. 13 corresponds to filters 9–3 presented in example 1, Fig. 14 corresponds to filters 9–7 presented in example 2, and Fig. 15 corresponds to filters 5–7 presented in example 3.) Here, the Lena image is taken as part of the training set in order to minimize the effects of quantization noise: this enables the influence of the filters to be taken into account.

For a given set of filters, separate codebooks are trained for each resolution–orientation subimage, and bit alloca-

Fig. 13. Filters no. 1 9-3, PSNR = 31.82 dB, $\mathcal{R}_T = 0.80$ bpp.

Fig. 15. Filters no. 3, 5-7, PSNR = 31.46 dB, $\mathcal{R}_T = 0.80$ bpp.

Fig. 14. Filters no. 2, 9-7, PSNR = 32.10 dB, $\mathcal{R}_T = 0.78$ bpp.

Fig. 16. Original 256 by 256 Lena, 8 bpp.

tion is carried out according to (25). For the Lena image, the bit assignment is represented in Fig. 17. Resolution 1 (diagonal orientation) is discarded. Resolution 1 (horizontal and vertical orientations) and resolution 2 (diagonal orientation) are coded using 256-vector codebooks (codeword size 4 by 4) resulting in a 0.5-b/pixel rate, while resolution 2 (horizontal and vertical orientations) is coded at a 2-b/pixel rate using 256-vector codebooks (codeword size 2 by 2). Finally, the lowest resolution is coded at 8 b/pixel.

B. Results as a Function of Regularity and Vanishing Moments

In Section II-B, we mentioned our belief that both the regularity of the reconstruction wavelet $\tilde{\psi}$ and the number

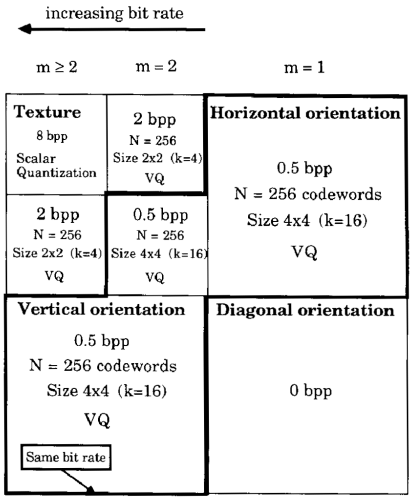


Fig. 17. Subimages bit rate allocation: example of a bit allocation for a total bit rate of 1 bpp and for the 256 by 256 Lena image.

of vanishing moments of the analyzing wavelet ψ are important in applications. To illustrate this we carried out the following experiments. For a given pair, h, $\tilde{h}$, we analyzed the same image twice: once as described above, and a second time after exchanging the roles of the filters h and $\tilde{h}$.

The filter pairs in example 2 both have the same number of vanishing moments, $k = \tilde{k} = 4$. However, $\tilde{\psi}$ is considerably more regular than ψ (see Fig. 3). With this filter pair, our experiment on the Lena image led to a PSNR of 32.10 dB in the first case, and to a PSNR of 31.51 dB if the roles of h and $\tilde{h}$ are inverted. The case where the reconstruction wavelet has the highest regularity, therefore, performs best.

In example 1 the functions ψ and $\tilde{\psi}$ have comparable regularity: both are continuous and neither has a continuous derivative. In fact $\tilde{\psi}$ is a bit more regular than ψ: $\tilde{\psi}$ is differentiable almost everywhere, and is Hölder continuous with exponent 1, while ψ is Hölder continuous with the exponent only at 0.83. On the other hand, ψ has 2 vanishing moments, while $\tilde{\psi}$ has 4 ($k = 4$, $\tilde{k} = 2$). The same experiment, again with the Lena image, now leads to a PSNR of 31.82 dB if h, $\tilde{h}$ are taken as in Table I, and to a PSNR of 31.13 dB when the roles of h and $\tilde{h}$ are reversed. The situation where $\tilde{\psi}$ is most regular but ψ has fewer vanishing moments, therefore, performs better (gain of 0.69 dB) than the case where ψ has more vanishing moments but $\tilde{\psi}$ is less regular. This seems to suggest that the regularity of $\tilde{\psi}$ has a larger effect than the number of vanishing moments of ψ. However, in this example the difference in overall regularity, as measured by the differences between Hölder exponents, is much smaller here than in example 2 (0.17 as compared to 0.63 in example 2), and it seems hard to explain how this smaller difference in Hölder exponent could account for a comparable gain in PSNR. In fact, the Hölder exponent is not a very good measure for the regularity of $\tilde{\psi}$ in this case: it is completely determined by the discontinuity of the derivative of $\tilde{\psi}$ in only a few points, and it is insensitive to the fact that $\tilde{\psi}$ is infinitely differentiable in all other points. If this is taken into account, then $\tilde{\psi}$ looks much more regular than ψ (the Hölder exponent of which is determined by its behavior near a dense set of points), which might explain the gain in PSNR.

We conclude from all this that: 1) for the same number of vanishing moments for ψ, the scheme with most regular $\tilde{\psi}$ is likely to perform best; and 2) increasing the regularity of $\tilde{\psi}$, even at the expense of the number of vanishing moments for ψ, may lead to better results.

Based on theoretical arguments (Taylor expansions) and results from numerical analysis [8], we also expect: 3) for comparable regularity of $\tilde{\psi}$, the scheme with largest vanishing moments for ψ is likely to perform best.

C. Comparison with Other Coders

If the PSNR is chosen as a criterion of comparison, these results are close to those obtained by Woods and O'Neil [42] and Westerink *et al.* [40]. However, in their subband coding algorithm, they use 32-taps Johnston filters, while only 9 or 7 taps are necessary for our method. According to Westerink's results in [41], the PSNR decreases by about 2 dB when using 8-taps Johnston filters. However, some others new QMF designs can also lead to good results with about 9 taps for image coding [1].

In this section, we present both numerical and qualitative comparison between our coding scheme and other previously published results. Since the most popular image in the recent literature has been the 512 by 512 Lena image, the comparison is made using this image taken outside the training set.

Among the different methods published, we consider the three following well-known methods: Ho and Gersho obtained a 30.93-dB PSNR at 0.36 bpp, result using "variable-rate multi stage VQ" [23]. Riskin and Gray improved on the full search VQ (PSNR = 29.29 dB, 0.32 bpp) using pruned tree structured VQ (PSNR = 30.92 dB, 0.32 bpp) [34]. High PSNR values were obtained by Woods and Cohen using entropy coded and predictive VQ (PSNR = 32.5 dB, 0.45 bpp) [11].

Our aim is not to optimize the PSNR but rather a weighted function of the MSE in order to match human vision. We give two examples at low bit rate using wavelet VQ.

Our initial result at 0.37 bpp presented Fig. 18 with a 30.85-dB PSNR is very close to those of Ho and Gersho [23] and Riskin *et al.* [34]. The perceptual quality of our coded images is better than indicated by the PSNR value

Fig. 18. 512 by 512 Lena image. Filters no. 2 9-7, PSNR = 30.85 dB, $\mathcal{R}_T = 0.37$ bpp.

Fig. 19. 512 by 512 Lena image. Filters no. 2 9-7, PSNR = 29.11 dB, $\mathcal{R}_T = 0.21$ bpp.

mainly due to the regularity of the wavelet and the bit allocation. These images do not suffer from the blocking effects obtained when using VQ in the spatial domain. No ringing effects can be observed.

The second result at 0.21 bpp presented in Fig. 19 with a 29.11-dB PSNR shows that a very low bit rate can be achieved with our method, without severe degradation.

Our method using a new class of filters derived from wavelet theory using full search VQ can be improved by any of the three above-mentioned methods.

In fact the LBG clustering algorithm is a very simple algorithm but not optimal for variable length code. The PSNR of the method could be improved by about 3 dB, for example, using ECVQ [34] but CPU time becomes prohibitively expensive.

D. Progressive Transmission Scheme

The main objective of progressive transmission is to allow the receiver to recognize a picture as quickly as possible at minimum cost, by sending a low resolution level picture first. Then, it can be decided to receiver further picture details or to abort the transmission. Further details of the picture are obtained by sequentially receiving the encoded wavelet coefficients at different resolution levels and directions.

Following the example of [40], we will display each picture level during the progressive transmission with a size that matches the resolution of that particular level.

To test the efficiency of the vector quantizer, the image to be coded is taken outside the training set.

Fig. 20 represents 5 stages in the progressive transmission of a 256 by 256 image using filters 9-7 given in example 2. According to the bit allocation procedure (Section III-C) with a generalized Gaussian PDF approximation law, only the wavelet coefficients corresponding to the $m = 1$ and $m = 2$ high resolution levels are vector quantized, while the low level subimages ($m \geq 2$) are scalar quantized.

V. Conclusion

This paper describes a new image coding scheme combining the wavelet transform and VQ.

A new family of filters has been derived from the wavelet theory. We have shown the importance of regularity and vanishing moments for image coding. Furthermore, these filters require few taps, unlike standard QMF methods.

The wavelet transform used here attempts to exploit the masking effect of the human eye, yielding encouraging results. Indeed, the proposed method enables high compression bit rates while maintaining good visual quality through the use of bit allocation in the subimages. The blocking effects seen when spatial VQ is performed are avoided.

This method is well adapted to progressive transmission as well as very low bit rate compression. Furthermore, using a simple full-search VQ provides good results, comparable to the best results published currently.

Further research should include some new derivation such as entropy constraint and predictive VQ. We would improve this coding scheme, if we accept a heavier computational load.

Resolution	*m=4*	*m=3*	*m=2*	*m=1*
Size	**16×16 pix**	**32 × 32 pix**	**64 × 64 pix**	**128 × 128 pix**
R_T	0.031 bpp	0.125 bpp	0.5 bpp	0.781 bpp
$\mathcal{R}_T$	0.0264 bpp	0.0919 bpp	0.3354 bpp	0.5039 bpp

Resolution m=0

256 × 256 pix

R_T = 1 bpp $\mathcal{R}_T$ = 0.6297 bpp

PSNR = 31.28 dB

Fig. 20. Progressive transmission–filters no. 2 9–7.

REFERENCES

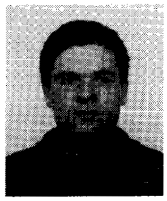
[1] E. H. Adelson and E. Simoncelli, "Non-separable extensions of quadrature mirror filters to multiple dimensions," *Proc. IEEE*, vol. 78, Apr. 1990.

[2] M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1965.

[3] V. R. Algazi, "Useful approximation to optimum quantization," *IEEE Trans. Commun.*, vol. COM-14, pp. 297–301, June 1966.

[4] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image

coding using vector quantization in the wavelet transform domain," in *Proc. IEEE ICASSP*, April 1990, pp. 2297-2300.
[5] M. Barlaud, L. Blanc-Féraud, P. Mathieu, J. Menez, and M. Antonini, "2D linear predictive image coding with vector quantization," in *Proc. EUSIPCO*, Grenoble, France, Sept. 5-8, 1988, pp. 1637-1640.
[6] M. Barlaud, P. Mathieu, and M. Antonini, "Wavelet transform image coding using vector quantization," presented at 6th Workshop on MDSP, Monterey, CA, Sept. 1989.
[7] G. Battle, "A block spin construction of wavelets. Part I Lemarié functions," *Comm. Math. Phys.*, vol. 110, pp. 601-615, 1987.
[8] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical analysis. I," to be published.
[9] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 482-540, 1983.
[10] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Phys.*, vol. 197, pp. 551-566, 1968.
[11] R. A. Cohen and J. W. Woods, "Sliding block entropy coding of images," in *Proc. IEEE ICASSP*, Glasgow, Scotland, May 23-26, 1989, pp. 1731-1733.
[12] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," AT&T Bell Lab., Tech. Rep., TM 11217-900529-07, 1990.
[13] J. H. Conway and N. J. A. Sloane, "A lower bound on the average error of vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 106-109, Jan. 1985.
[14] I. Daubechies, A. Grossman, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math. Phys.*, vol. 27, pp. 1271-1283, 1986.
[15] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," to be published.
[16] —, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 41, pp. 909-996, 1988.
[17] —, "Orthonormal bases of compactly supported wavelets. II. Variations on a theme," AT&T Bell Lab., Tech. Rep. TM 11217-891116-17, 1990.
[18] J. C. Feauveau, "Analyse multirésolution par ondelettes non orthogonales et bancs de filtres numériques," Ph.D. dissertation, Univ. Paris Sud, France, Jan. 1990.
[19] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, July 1979.
[20] —, "On the structure of vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-28, Mar. 1982.
[21] R. M. Gray, "Vector quantization," *IEEE ASSP Mag.*, pp. 4-29, Apr. 1984.
[22] A. Grossman and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM J. Math Anal.*, vol. 15, pp. 723-736, 1984.
[23] Y. Ho and A. Gersho, "Variable-rate multi-stage vector quantization for image coding," in *Proc. IEEE ICASSP*, New York, Apr. 1988.
[24] P. G. Lemarié, "Une nouvelle base d'ondelettes de $L^2(\mathbb{R})$," *J. Math. Pures et Appl.*, vol. 67, pp. 227-238, 1988.
[25] P. G. Lemarié and Y. Meyer, "Ondelettes et bases hilbertiennes," *Rev. Mat. Iberoamericana*, vol. 2, pp. 1-18, 1986.
[26] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84-95, Jan. 1980.
[27] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 11, July 89.
[28] D. Marr, *Vision*. New York: Freeman, 1982.
[29] P. Mathieu, M. Barlaud, and M. Antonini, "Compression d'Images par transformée en ondelette," *12ième colloque GRETSI, Juan les Pins*, June 12-16, 1989.
[30] P. Mathieu, M. Barlaud, and M. Antonini, "Compression d'Image par transformée en ondelette et quantification vectorielle," *Traitement du Signal*, vol. 7, no. 2, 1990.
[31] Y. Meyer, "Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs," *Seminaire Bourbaki*, no. 662, 1985-1986.
[32] N. M. Nasrabadi and R. A. King, "Image coding using vector quantization: A review," *IEEE Trans. Commun.*, vol. 36, Aug. 1988.
[33] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.
[34] E. Riskin, E. M. Daly, and R. M. Gray, "Pruned tree-structured vector quantization in image coding," in *Proc. IEEE ICASSP*, Glasgow, Scotland, May 1989, pp. 1735-1738.
[35] M. J. Smith and D. P. Barnwell, "Exact reconstruction for tree-structured subband coders," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-34, pp. 434-441, 1986.
[36] J. O. Stromberg, "A modified haar system and higher order spline systems," in *Conf. in Harmonic Analysis in Honor of Antoni Zygmund*, Vol. II, pp. 475-493.
[37] M. Vetterli, "Splitting a signal into subsampled channels allowing perfect reconstruction," in *Proc. IASTED Conf. Appl. Signal Processing Digital Filtering*, Paris, France, June 1985.
[38] M. Vetterli and C. Herley, "Wavelets and filter banks: Relationships and new results," in *Proc. IEEE ICASSP*, Albuquerque, Apr. 1990.
[39] P. H. Westerink, D. E. Boekee, J. Biemond, and J. W. Woods, "Subband coding of image using vector quantization," *IEEE Trans. Commun.*, vo. 36, pp. 713-719, 1988.
[40] P. H. Westerink, J. Biemond, and D. E. Boekee, "Progressive transmission of images using subband coding," in *Proc. IEEE ICASSP*, 1989, pp. 1811-1814.
[41] P. H. Westerink, "Subband coding of images," Ph.D. dissertation Delft Univ., 1989.
[42] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-34, Oct. 1986.
[43] P. Zador, "Asymptotic quantization error of continuous signals and their quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 139-149, 1982.



Marc Antonini was born in France on August 29, 1965. He received the DEA degree in signal processing in 1988 from the University of Nice-Sophia Antipolis, France, and the Ph.D. degree from the Laboratory of Signaux et Systèmes, URA I3S, CNRS and the University of Nice-Sophia Antipolis in 1991.

His research interests include multidimensional image processing, wavelet analysis, and image coding.

Michel Barlaud (M'88) was born in France on November 24, 1945. He received the "Doctorat d'Etat" degree from University of Paris XII.

He is currently a Professor and a member of the Laboratory of Signaux et Systèmes, URA I3S both from CNRS and University of Nice-Sophia Antipolis. After some work on non-stationary signal processing, his research interests move towards multidimensional image processing, wavelet analysis, image coding, inverse problems, image restoration, and edge detection.

Dr. Barlaud is member of the IEEE-ASSP MDSP committee.

Pierre Mathieu was born in Alger on May 10, 1956. He received the Ingenieur ENSEEIHT and Ph.D. degrees from INP Toulouse.

He is currently Maître de Conférences in the Laboratory of Signaux et Systèmes, URA I3S both from CNRS and University of Nice-Sophia Antipolis. His research interests include multidimensional image processing, wavelet analysis, image coding, and image restoration.

Ingrid Daubechies (M'89) received the B.S. and Ph.D. degrees from the Vrje Universiteit Brussel, Belgium in 1975 and 1980, both in physics.

She is currently a Member of Technical Staff in the Mathematics Center of AT&T Bell Laboratories, Murray Hill, NJ. Her current research interests include mathematical problems in connection with signal analysis, in particular applications of time-frequency representations.

Annexe B

3D Scan-based Wavelet Transform and Quality Control for Video Coding

EURASIP Journal on Applied Signal Processing vol.2003, no.2, pp. 56-65, janvier 2003

J'ai co-écrit cet article avec Christophe Parisot et Michel Barlaud.

Résumé La transformée en ondelettes s'est avérée être plus performante que la DCT en terme de compression sur des images fixes et permet en outre la scalabilité des données. Pour le codage des vidéos, il est donc intéressant d'étendre la transformée en ondelettes bidimensionnelle au cas tridimensionnel (2D+t). Cependant, la transformée 2D+t nécessite beaucoup d'espace mémoire pour coder des grands blocs spatio-temporels transformés et éviter ainsi un effet de saccades lors de la visualisation. Dans cet article, nous avons proposé une méthode qui permet de réaliser une transformée en ondelettes 2D+t "au fil de l'eau" scalable et à faibles coûts mémoire et CPU, permettant d'éviter l'effet de saccades. Nous avons combiné cette approche avec une allocation des ressources binaires qui favorise la conservation de la "qualité" image au cours du temps.

EURASIP Journal on Applied Signal Processing 2003:1, 56–65

3D Scan-Based Wavelet Transform and Quality Control for Video Coding

Christophe Parisot
Laboratoire I3S, UMR 6070 (CNRS, Université de Nice-Sophia Antipolis) Bât. Algorithmes/Euclide 2000, route des Lucioles, BP 121 F-06903 Sophia Antipolis Cedex, France
Email: parisot@i3s.unice.fr

Marc Antonini
Laboratoire I3S, UMR 6070 (CNRS, Université de Nice-Sophia Antipolis) Bât. Algorithmes/Euclide 2000, route des Lucioles, BP 121 F-06903 Sophia Antipolis Cedex, France
Email: am@i3s.unice.fr

Michel Barlaud
Laboratoire I3S, UMR 6070 (CNRS, Université de Nice-Sophia Antipolis) Bât. Algorithmes/Euclide 2000, route des Lucioles, BP 121 F-06903 Sophia Antipolis Cedex, France
Email: barlaud@i3s.unice.fr

Received 4 March 2002 and in revised form 17 October 2002

Wavelet coding has been shown to achieve better compression than DCT coding and moreover allows scalability. 2D DWT can be easily extended to 3D and thus applied to video coding. However, 3D subband coding of video suffers from two drawbacks. The first is the amount of memory required for coding large 3D blocks; the second is the lack of temporal quality due to the sequence temporal splitting. In fact, 3D block-based video coders produce jerks. They appear at blocks temporal borders during video playback. In this paper, we propose a new temporal scan-based wavelet transform method for video coding combining the advantages of wavelet coding (performance, scalability) with acceptable reduced memory requirements, no additional CPU complexity, and avoiding jerks. We also propose an efficient quality allocation procedure to ensure a constant quality over time.

Keywords and phrases: scan-based DWT, 3D subband coding, quality control, video coding.

1. INTRODUCTION

Although 3D subband coding of video [1, 2, 3, 4, 5] provides encouraging results compared to MPEG [6, 7, 8, 9], its generalization suffers from significant memory requirements. One way to reduce memory requirements is to apply the temporal discrete wavelet transform (DWT) on 3D blocks coming from a temporal splitting of the sequence. But this block-based DWT method introduces temporal blocking artifacts which result in undesirable jerks during video playback. In this paper, we propose new tools for 3D subband codecs to guarantee the output frames a constant quality over time.

Scan-based 2D wavelet transforms were first suggested for on-board satellite compression in [10, 11] and by Chrysafis and Ortega in [12].

In Section 2, we propose a 3D scan-based DWT method and a 3D scan-based motion-compensated lifting DWT for video coding. The method allows the computation of the temporal wavelet decomposition of a sequence with infinite length using little memory and no extra CPU. Furthermore, the proposed wavelet transform provides higher quality control than 3D block-based video compression schemes (avoiding jerks).

In Section 3, we propose an efficient model-based quality control procedure. This bit-allocation procedure controls the output frames quality over time. This new quality-control procedure takes advantage of the model-based rate allocation methods described in [13].

Finally, Section 4 presents experimental results obtained by our method.

2. 3D VIDEO WAVELET TRANSFORM

2.1. Principle

The method generally used to reduce memory requirements for large image coding is to split the image and then perform the transform on tiles such as JPEG with 8×8 DCT blocks or JPEG2000 [14]. Unfortunately, the coefficients are computed from periodic or symmetrical extensions of the signal.

This results in undesirable blocking artifacts. For video coding, the same blocking artifacts in the temporal direction (introduced by temporal splitting) result in jerks.

In this section, we propose a 3D wavelet transform framework for video coding that requires storing a minimum amount of data without any additional CPU complexity [15]. The frames of the sequence are acquired and processed on the fly.

Definitions of the temporal coherence and the buffer names

We consider a temporal interval (set of input frames). We define the set of its *temporally coherent wavelet coefficients* as the set of all coefficients, in all subbands, obtained by a filter (or convolution of filters) centered on any one of the frames of this temporal interval. In this paper, we assume that encoding is allowed only when we have a temporally coherent set of wavelet coefficients. Temporal coherence improves the encoder performance since it allows optimal bit allocation for wavelet coefficients of the same temporal interval.

The set of buffers used to perform the temporal wavelet transform will be called *filtering buffers*. These buffers produce low- and high-frequency temporal wavelet coefficients. In the same way, we call *synchronization buffers*, the set of buffers used to store output coefficients before their encoding.

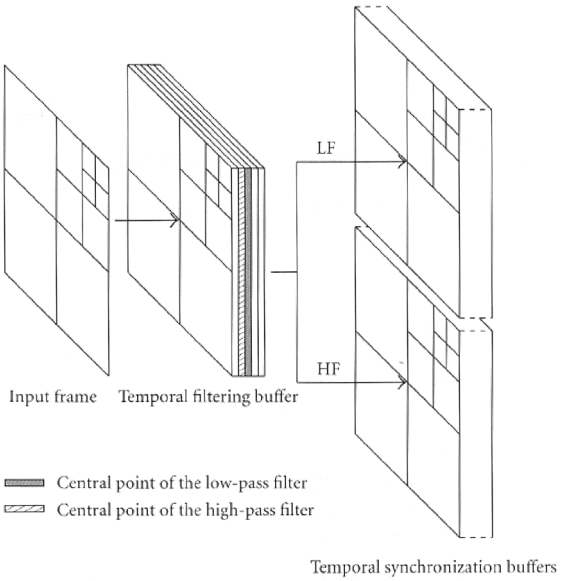


FIGURE 1: One-level temporal scan-based wavelet decomposition for the 5/3 filter bank.

2.2. Temporal scan-based video DWT and delay

Consider the case of a 3D wavelet transform which can be split into a 2D DWT on each frame and an additional 1D DWT in the time direction [16]. In this paper, we focus on an efficient implementation of the temporal wavelet transform and we propose a method independent of the choice of the spatial wavelet transform.

Each time a frame is received, we perform its 2D wavelet transform and send it into our scan-based temporal wavelet transform system. We consider symmetrical filters with odd length since they are the most widely used in image compression algorithms [14, 17]. To simplify, we also suppose that the low-pass filter is longer than the high-pass one. Let $L = 2S + 1$ be the length of the low-pass filter with $S \geq 2$. We want to design components that can be easily reused for any wavelet decomposition tree. Therefore, the memory used for the filtering buffers is supposed to be internal and cannot be shared with other filtering buffers nor with the synchronization buffers for wavelet coefficients storage. We propose a method that minimizes the total memory requirements for FIR filtering.

2.2.1 Single-stage DWT

We first consider a single stage of the temporal wavelet transform.

The length of the low-pass filter is L. Therefore, we need L frames of 2D wavelet coefficients in memory to compute one frame of low-frequency temporal wavelet coefficients. The high-pass filter is shorter. Thus, our filtering buffer must contain exactly L frames of 2D wavelet coefficients. Consequently, filtering buffers are FIFO with length L. Figure 1 shows the scheme for a single stage of a 5/3 temporal wavelet decomposition. The filtering buffer contains five frames of 2D wavelet coefficients. The synchronization buffers are used to store output 3D wavelet coefficients until we get a temporally coherent set of 3D wavelet coefficients.

When the $(S+1)$st 2D transformed frame is received, the filtering buffer is symmetrically filled up in order to avoid side effects. The central frame is the 2D wavelet transform of the first image of the sequence. We can compute the first low-frequency temporal coefficients applying the low-pass filter to the central frame of the filtering buffer (gray frame in Figure 1). The first high-frequency temporal coefficients must be computed on the second 2D transformed frame. This frame (hatched frame in Figure 1) and all its necessary neighbours are already present in the filtering buffer since the high-pass filter is shorter than the low-pass one. Therefore, the high-frequency temporal wavelet coefficients can also be computed without additional input frame.

Finally, we have to wait for only $S + 1$ input frames to get one low-frequency and one high-frequency temporal frames of wavelet coefficients. Then, for each pair of input frames, we can compute both low-frequency and high-frequency coefficients. Each pair of low- and high-frequency frames is a set of temporally coherent wavelet coefficients. Therefore, we need $S+1$ input frames to get the first set of temporally coherent wavelet coefficients and $S+1+2(n-1) = S+2n-1$ input frames to get a set of n low-frequency and n high-frequency output frames.

When the input sequence is finished, input frames are replaced by a symmetrical extension using the frames present in the filtering buffer in order to flush it.

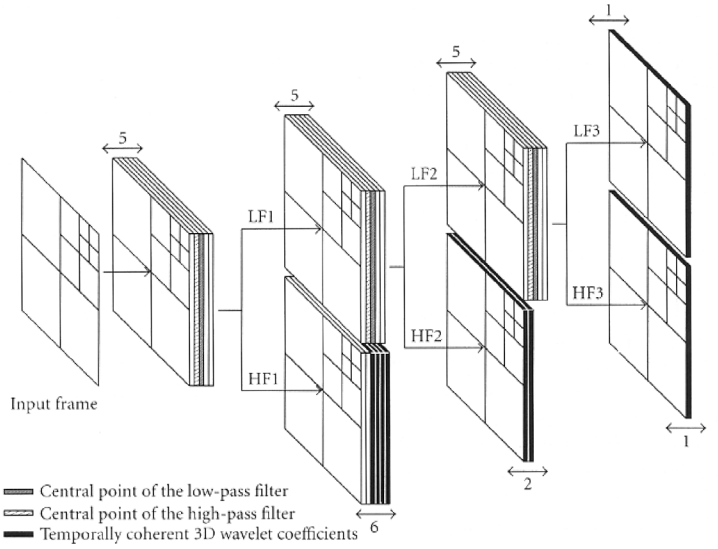


FIGURE 2: Three-level temporal scan-based wavelet decomposition for the 5/3 filter bank.

2.2.2 Multistage DWT

We now consider the general scheme of an N-level temporal wavelet decomposition. We focus only on the usual dyadic decomposition without additional high-frequency subband decomposition. We assume that decomposition levels are indexed from 1 to N, where level j corresponds to the coefficients produced by the jth wavelet decomposition (level 0 is the sequence of all 2D wavelet transformed frames).

We compute the encoding delay for a two-level wavelet decomposition. The first stage has to compute $S+1$ low-frequency temporal frames to get coefficients in both low-frequency and high-frequency subbands of the second level. In the same time, the first stage has also computed $S+1$ high-frequency temporal frames. But, from Section 2.2.1, we know that these $S+1$ low-frequency and $S+1$ high-frequency output frames of 3D wavelet coefficients can only be computed after the delay of $S+2(S+1)-1$ frames. Thus, we have to wait for $3S+1$ frames to get one frame of 3D coefficients in all subbands of the second decomposition level and $S+1$ frames of 3D coefficients in the first level. Notice that for temporal coherence, we need only the first two frames among the $S+1$ of the first level.

To compute the delay for an N-level temporal wavelet decomposition, we define d_j as the number of frames required at the input of the jth filtering buffer to get temporally coherent coefficients in all subbands. The processing of the first set of 3D subbands of temporally coherent wavelet coefficients will be possible after $D = d_1$ frames have been received. From Section 2.2.1, we know that $d_j = S + 2d_{j+1} - 1$ for $j \in \{1, \ldots, N-1\}$ and $d_N = S+1$. Solving these equations, we find that the number of input frames required at level j before the first wavelet coefficients are available for processing is $d_j = (2^{N+1-j} - 1)S + 1$. Therefore, for an N-level temporal wavelet decomposition, the number of input frames needed to get the first set of temporally coherent wavelet coefficients is

$$D = (2^N - 1)S + 1. \tag{1}$$

TABLE 1: Number of input frames necessary to get the first set of temporally coherent wavelet coefficients (1).

Number of levels (N)	9/7 DWT	5/3 DWT
1	5	3
2	13	7
3	29	15

Thus, the number of frames needed for the synchronization of the multistage decomposition increases exponentially with the number of decomposition levels. Figure 2 shows the scheme of a three-level wavelet decomposition for $S = 2$. Dark frames in the synchronization buffers are the set of coefficients which will be processeded together (quantized and encoded) as soon as we have coefficients in all temporal frequency bands. This set of coefficients is temporally coherent. At the beginning of the sequence, we have to wait for D input frames. Then, sets of temporally coherent coefficients will be available each 2^N input frames. Table 1 shows the number of input frames needed to get the first set of temporally coherent wavelet coefficients for two widely used filter banks. This table shows that a three-level decomposition introduces an encoding delay of less than one second with the 9/7 filter

bank and only half a second with the 5/3 filter bank. In 3D block-based video coders, the delay is equal to the size of the temporal block. As blocks are larger in order to minimize the number of jerks, the delay is more important for 3D block-based wavelet transform video coders.

2.3. Memory requirements

Memory requirements are given by the sum of the number of frames in the N filtering buffers and the number of frames in the synchronization buffers.

The memory requirements for the filtering buffers are equal to $(2S+1)N$ frames.

The synchronization buffers of the last decomposition level must contain one frame of 3D wavelet coefficients for both the low-frequency and high-frequency subbands. For the jth decomposition level ($j < N$), d_{j+1}, low-frequency outputs need to be computed and, in the same time, d_{j+1} high-frequency outputs can be computed. As we know that temporal coherence requires less than d_{j+1} 3D frames of wavelet coefficients at level j, we can decide to delay the computation of the last computable high-frequency coefficients until the new set of temporally coherent 3D wavelet coefficients has been encoded. Once the set of temporally coherent coefficients has been encoded, we compute all the high-frequency coefficients for levels 1 to $N-1$ and send them into the synchronization buffers. Then, the on-the-fly wavelet transform can resume normally. This trick allows to spare one frame in the memory requirements of each synchronization buffer for levels 1 to $N-1$. Thus, the memory requirements for the synchronization buffers are limited to $2+\sum_{j=1}^{N-1}(d_{j+1}-1)$.

We need to store $M_S = (2^N - N - 1)S + 2$ frames of coefficients for all the synchronization buffers. Therefore, the total memory requirements of this method are

$$M = (2^N + N - 1)S + N + 2 \tag{2}$$

frames, for an N-level temporal wavelet transform with filter length $L = 2S+1$. When memory can be shared between filtering buffers and synchronization buffers, the total memory requirements are limited to

$$M = (2^N + N - 1)S + 1 \tag{3}$$

frames. See [18] for complete memory requirements formulae.

Tables 2 and 3 show the total memory requirements for the 9/7 and 5/3 filter banks, respectively, for independent and shared buffers.

Memory requirements increase as an exponential function of the resolution N and as a linear function of the filter length.

Note that, for the same memory requirements (e.g., 48 frames) and three levels of the 9/7 DWT decomposition with a frame rate of 30 fps, the encoding delay for temporal block-based video coders is equal to 1.6 second while it is 0.97 second in our case (from Table 1). Furthermore, block-based video coders have jerks for each group of 48 frames while our method avoids these annoying artifacts.

Table 2: Memory requirements (2), in terms of frames, of the scan-based DWT system including both filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	11	7
2	24	14
3	45	25

Table 3: Memory requirements (3), in terms of frames, of the scan-based DWT system including both filtering and synchronization buffers when memory can be shared between filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	9	5
2	21	11
3	41	21

The CPU complexity of our temporal scan-based DWT is exactly the same as to perform the regular 1D DWT in the temporal direction on the entire sequence.

2.4. Scan-based motion compensated lifting

The main drawback of the 3D scan-based DWT is that it does not take motion compensation into account. 3D motion compensated lifting is an efficient tool to take account of motion in video [4, 6, 9, 19, 20, 21].

Thus, we propose a new 3D scan-based motion compensated lifting scheme [18, 22]. This method combines the benefits of scan-based filtering, block-based coding, and quality control [22].

When filtering and synchronization buffers are independent, the total memory requirements become

$$M = (2^N - N - 1)S + \beta N + 2, \tag{4}$$

where β is a parameter depending on the filter, $\beta = 6$ for the 9/7 Daubechies DWT [23], and $\beta = 4$ for the 5/3 DWT. When memory can be shared between filtering and synchronization buffers, the total memory requirements are limited to

$$M = (2^N - N - 1)S + (\beta - 1)N + 1. \tag{5}$$

Complete memory requirements computation can be found in [18]. The scan-based motion compensated lifting scheme saves memory compared to the regular filter banks implementation. Furthermore, our method does not increase the CPU complexity compared to the usual lifting implementation.

Tables 4 and 5 show the memory requirements for scan-based motion compensated lifting video coders, respectively, for independant and shared buffers.

Thus, the scan-based motion compensated lifting scheme saves 12 to 33% memory (Tables 2 and 4 or Tables 3 and 5) and takes account motion compensation.

Table 4: Memory requirements (4), in terms of frames, of the scan-based motion compensated lifting DWT system including both filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	8	6
2	18	12
3	36	22

Table 5: Memory requirements (5), in terms of frames, of the scan-based motion compensated lifting DWT system including both filtering and synchronization buffers when memory can be shared between filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	6	4
2	15	9
3	32	18

A 32-frames memory (which is a reasonable GOP memory) allows to implement a 3D scan-based motion compensated lifting with efficient filters (9/7) and three-level decomposition.

The scan-based motion compensated lifting also removes jerks with quality control.

3. MODEL-BASED TEMPORAL QUALITY CONTROL

The bit allocation for the successive sets of temporally coherent coefficients can be performed with respect to either rate or quality constraints. In both cases, the goal is to find a set of quantizers to apply in each subband, which performances lie on the convex hull of the global rate-distorsion curve [24, 25, 26, 27].

Three different methods can be used to model the rate and distortion.

(i) The first one—used in JPEG2000 [14]—consists in prequantizing the wavelet coefficients with a small predetermined quantization step and encodes their bitplanes until the rate or distortion constraint (depending on the application) is verified. In this method, the quantization step of each wavelet coefficient can only be a product of the chosen quantization step multiplied by an integer power of two. The distortion and bitrate functions are exact but they are computed during the encoding process.

(ii) The second method uses asymptotic models for both the distortion and the bitrate. As the asymptotic rate and distortion functions are simple, the minimum of the rate or distortion allocation criterion can be computed analytically. This method is therefore the simplest one to get the quantization steps to apply in each subband. However, the asymptotic assumption is only true for high bitrate subbands.

(iii) We have proposed to use nonasymptotic theoretical models for both rate and distortion [13]. The rate and the distortion depend on the quantization step but also on the probability density function of the wavelet coefficients. Assuming that the probability density model is accurate, this method provides optimal rate-distortion performances.

In this section, we propose a new nonasymptotic temporal *quality* control procedure to ensure constant quality over time. The quality measure is based on the mean square error (MSE) between the compressed signal and the original one.

3.1. Principle of the model-based MSE allocation

The purpose of MSE allocation is to determine the optimal quantizers in each subband which minimize the total bitrate for a given output MSE. Since the 9/7 biorthogonal filter bank is nearly orthogonal, the MSE between the original image and the decoded one can be computed by a weighted sum of the mean squared quantization errors of each subband. We have

$$\mathrm{MSE}_{\mathrm{output}} = \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_{Q_i}^2, \tag{6}$$

with $\#SB$ the number of 3D subbands, $\sigma_{Q_i}^2$ the mean squared quantization error for subband i, and $\{\pi_i\}$ the weights used to take account of the nonorthogonality of the filter bank [28]. The weights Δ_i are optional and can be used for frequency selection or distortion measures. The output bitrate can be expressed as the following weighted sum:

$$R_{\mathrm{output}} = \sum_{i=1}^{\#SB} a_i R_i, \tag{7}$$

with R_i the output bitrate for subband i and a_i the weight of subband i in the total bitrate (a_i is the ratio of the size of subband i divided by the size of the sequence).

The subband quantizers are uniform scalar quantizers. They are defined by their quantization bins q_i. The solution of our constrained problem is obtained thanks to Lagrangian operators by minimizing the following criterion:

$$J(\{q_i\}, \lambda) = \sum_{i=1}^{\#SB} a_i R_i(q_i) + \lambda \left(\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_{Q_i}^2(q_i) - D_T \right), \tag{8}$$

where D_T denotes the target output MSE and both R_i and $\sigma_{Q_i}^2$ depend on the quantization steps q_i. The models used for the bitrate and distortion functions are described in the next subsection.

3.2. Rate and distortion models

In each 3D subband, the probability density function of the wavelet coefficients is unimodal with zero mean and can be approximated with generalized Gaussian [23, 29]. Therefore, we have

$$p_{\alpha,\sigma}(x) = a e^{-|bx|^{\alpha}}, \tag{9}$$

with $b = (1/\sigma)\sqrt{\Gamma(3/\alpha)/\Gamma(1/\alpha)}$ and $a = b\alpha/2\Gamma(1/\alpha)$. We also assume that wavelet coefficients are independent and identically distributed (i.i.d.) [13] in each subband.

Let $\Pr(m)$ be the probability of the quantization level m so that

$$\Pr(m) = \int_{(|m|-1/2)q}^{(|m|+1/2)q} p_{\alpha,\sigma}(x)dx, \tag{10}$$

for $m \neq 0$ and

$$\Pr(0) = \int_{-q/2}^{+q/2} p_{\alpha,\sigma}(x)dx. \tag{11}$$

From (10) and (11), we can approximate the bitrate R by the entropy of the output quantization levels

$$R = -\sum_{m=-\infty}^{+\infty} \Pr(m)\log_2 \Pr(m). \tag{12}$$

The best coding value for the quantization level m [30] is the centroid of its quantization bin

$$\hat{x}_m = \operatorname{sign}(m) \times \frac{\int_{(|m|-1/2)q}^{(|m|+1/2)q} x p_{\alpha,\sigma}(x)dx}{\Pr(m)}, \tag{13}$$

for $m \neq 0$ and $\hat{x}_0 = 0$.

The mean squared quantization error is given by

$$\sigma_Q^2 = \int_{-q/2}^{+q/2} x^2 p_{\alpha,\sigma}(x)dx + 2\sum_{m=1}^{+\infty} \int_{(m-1/2)q}^{(m+1/2)q} (x-\hat{x}_m)^2 p_{\alpha,\sigma}(x)dx. \tag{14}$$

Inserting the value of $\hat{x}_m$ into (14), we get

$$\sigma_Q^2 = \sigma^2 - 2\sum_{m=1}^{+\infty} \frac{\left(\int_{(m-1/2)q}^{(m+1/2)q} x p_{\alpha,\sigma}(x)dx\right)^2}{\int_{(m-1/2)q}^{(m+1/2)q} p_{\alpha,\sigma}(x)dx}. \tag{15}$$

Proposition 1. *When $p_{\alpha,\sigma}$ is a generalized Gaussian distribution with standard deviation σ and shape parameter α, there is a family of functions $f_{n,m}$ which verifies*

$$\begin{aligned} \int_{-q/2}^{+q/2} x^n p_{\alpha,\sigma}(x)dx &= \sigma^n f_{n,0}\left(\alpha, \frac{q}{\sigma}\right), \\ \int_{(m-1/2)q}^{(m+1/2)q} x^n p_{\alpha,\sigma}(x)dx &= \sigma^n f_{n,m}\left(\alpha, \frac{q}{\sigma}\right) \quad \forall m > 0 \end{aligned} \tag{16}$$

with

$$\begin{aligned} f_{n,0}\left(\alpha, \frac{q}{\sigma}\right) &= \int_{-(1/2)(q/\sigma)}^{+(1/2)(q/\sigma)} x^n p_{\alpha,1}(x)dx, \\ f_{n,m}\left(\alpha, \frac{q}{\sigma}\right) &= \int_{(m-1/2)(q/\sigma)}^{(m+1/2)(q/\sigma)} x^n p_{\alpha,1}(x)dx. \end{aligned} \tag{17}$$

Proof of Proposition 1 is given in [18].

Therefore, the bitrate R and the quantization distortion σ_Q^2 depend only on the shape parameter α and the ratio q/σ,

$$R = R\left(\alpha, \frac{q}{\sigma}\right), \qquad \sigma_Q^2 = \sigma^2 D\left(\alpha, \frac{q}{\sigma}\right) \tag{18}$$

with

$$\begin{aligned} R\left(\alpha, \frac{q}{\sigma}\right) = &-f_{0,0}\left(\alpha, \frac{q}{\sigma}\right)\log_2 f_{0,0}\left(\alpha, \frac{q}{\sigma}\right) \\ &- 2\sum_{m=1}^{+\infty} f_{0,m}\left(\alpha, \frac{q}{\sigma}\right)\log_2 f_{0,m}\left(\alpha, \frac{q}{\sigma}\right), \end{aligned} \tag{19}$$

$$D\left(\alpha, \frac{q}{\sigma}\right) = 1 - 2\sum_{m=1}^{+\infty} \frac{f_{1,m}(\alpha, q/\sigma)^2}{f_{0,m}(\alpha, q/\sigma)}. \tag{20}$$

3.3. Optimal model-based quantization for MSE control

Therefore, the goal is to find the quantization steps $\{q_i\}$ and λ which minimize

$$J(\{q_i\}, \lambda) = \sum_{i=1}^{\#SB} a_i R\left(\alpha_i, \frac{q_i}{\sigma_i}\right) + \lambda\left(\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D\left(\alpha_i, \frac{q_i}{\sigma_i}\right) - D_T\right). \tag{21}$$

We differentiate the criterion with respect to q_i and λ. This provides the following equations:

$$\begin{aligned} a_i \frac{\partial R}{\partial \widetilde{q}_i}(\alpha_i, \widetilde{q}_i) + \lambda \Delta_i \pi_i \sigma_i^2 \frac{\partial D}{\partial \widetilde{q}_i}(\alpha_i, \widetilde{q}_i) = 0, \quad \forall i, \\ \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D(\alpha_i, \widetilde{q}_i) - D_T = 0, \end{aligned} \tag{22}$$

where $\widetilde{q}_i = q_i/\sigma_i$.

Thus, the quantizers parameters $\{q_i\}$ must verify the following system of $\#SB + 1$ equations and $\#SB + 1$ unknowns:

$$\begin{aligned} \frac{(\partial D/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)}{(\partial R/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)} = -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2}, \quad \forall i, \\ \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D(\alpha_i, \widetilde{q}_i) = D_T. \end{aligned} \tag{23}$$

In order to simplify the notation, write

$$h_{\alpha_i}(\widetilde{q}_i) = \frac{(\partial D/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)}{(\partial R/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)}, \tag{24}$$

where

$$h_\alpha(\widetilde{q}) = \frac{A}{B}\ln 2, \tag{25}$$

where $A = \sum_{m=1}^{+\infty}(2(\partial f_{1,m}/\partial \widetilde{q})(\alpha, \widetilde{q}) f_{1,m}(\alpha, \widetilde{q}) f_{0,m}(\alpha, \widetilde{q}) - f_{1,m}(\alpha, \widetilde{q})^2(\partial f_{0,m}/\partial \widetilde{q})(\alpha, \widetilde{q}))/f_{0,m}(\alpha, \widetilde{q})^2$, $B = (p_{\alpha,1}(\widetilde{q}/2)/2) \times [\ln f_{0,0}(\alpha, \widetilde{q}) + 1] + \sum_{m=1}^{+\infty}(\partial f_{0,m}/\partial \widetilde{q})(\alpha, \widetilde{q})[\ln f_{0,m}(\alpha, \widetilde{q}) + 1]$

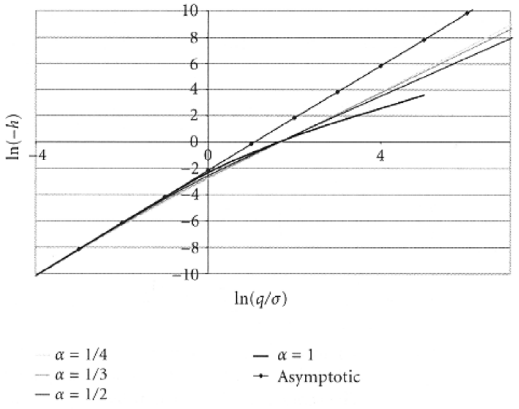


FIGURE 3: Tables of $\ln(-h(q/\sigma))$ for different shape parameters α of the generalized Gaussian distribution.

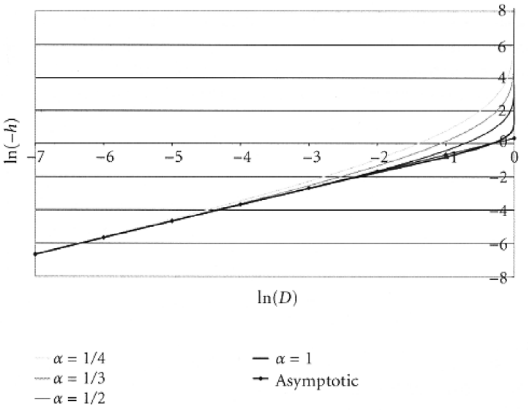


FIGURE 4: Tables of $\ln(-h) = \ln(a_i/\lambda\pi_i\sigma_i^2)$ versus $\ln D$ for different shape parameters α of the generalized Gaussian distribution.

with

$$\frac{\partial f_{n,m}}{\partial \tilde{q}}(\alpha,\tilde{q}) = \left[\left(m+\frac{1}{2}\right)^{n+1} p_{\alpha,1}\left(m\tilde{q}+\frac{\tilde{q}}{2}\right) - \left(m-\frac{1}{2}\right)^{n+1} p_{\alpha,1}\left(m\tilde{q}-\frac{\tilde{q}}{2}\right)\right]\tilde{q}^n. \tag{26}$$

Equations (23) become

$$h_{\alpha_i}(\tilde{q}_i) = -\frac{a_i}{\lambda\Delta_i\pi_i\sigma_i^2}, \quad \forall i,$$
$$\sum_{i=1}^{\#SB} \Delta_i\pi_i\sigma_i^2 D(\alpha_i,\tilde{q}_i) = D_T. \tag{27}$$

The solution of the MSE allocation problem can be obtained with the following equations:

$$\sum_{i=1}^{\#SB} \Delta_i\pi_i\sigma_i^2 D\left(\alpha_i, h_{\alpha_i}^{-1}\left(-\frac{a_i}{\lambda\Delta_i\pi_i\sigma_i^2}\right)\right) = D_T, \tag{28}$$

$$\tilde{q}_i = h_{\alpha_i}^{-1}\left(-\frac{a_i}{\lambda\pi_i\sigma_i^2}\right), \quad \forall i, \tag{29}$$

where h^{-1} is the inverse function of h. The parameter λ can be found from (28), and then (29) provides the optimal quantization steps q_i. Unfortunately, as there is no analytical formula for h^{-1}, the MSE allocation problem will be solved using a parametric approach described below.

3.4. Parametric approach

Equation (29) gives the values of the quantization steps using tables of the function h for different shape parameters α. Figure 3 shows the tables of $\ln(-h_\alpha(\tilde{q}))$ for $\alpha = 1, 1/2, 1/3$, and $1/4$ and the asymptotic curve of equation

$$\ln(-h) = 2\ln\frac{q}{\sigma} + \ln\frac{\ln 2}{6}. \tag{30}$$

To solve (28), we need tables linking D and λ. Using (20) and (25), we plot the parametric curve (with parameter $\tilde{q}$)

$$[\ln D(\alpha,\tilde{q}); \ln(-h_\alpha(\tilde{q}))], \tag{31}$$

for a given α. Using (29), this parametric curve is equivalent in each subband to the following parametric curve:

$$\left[\ln D; \ln\left(\frac{a_i}{\lambda\pi_i\sigma_i^2}\right)\right]. \tag{32}$$

Figure 4 shows these tables for $\alpha = 1, 1/2, 1/3$, and $1/4$ and the asymptotic curve of equation

$$\ln(-h) = \ln D + \ln(2\ln 2). \tag{33}$$

Thus, we have a relation between D and λ in each subband. The optimal λ is found using the constraint (28). Then, we have a relation between λ and the quantization step q_i in each subband.

3.5. Algorithm of the model-based MSE allocation

The proposed MSE allocation procedure is the following.

(1) Set the initial value of λ to its asymptotic optimum value $\lambda = 1/2D_T \ln 2$.

(2) For each 3D subband i, compute $\ln(a_i/\lambda\Delta_i\pi_i\sigma_i^2) = \ln(-h)$ and read the corresponding normalized MSE D_i using the tables shown in Figure 4.

(3) Compute $|\sum_{i=1}^{\#SB} \Delta_i\pi_i\sigma_i^2 D_i - D_T|$. If it is lower than a given threshold, the constraint (28) is verified and the current λ is optimal. Otherwise, compute[1] a new value of λ and go back to step (1).

[1] Several methods (such as dichotomy, bisection, secant method, golden section search) can be used.

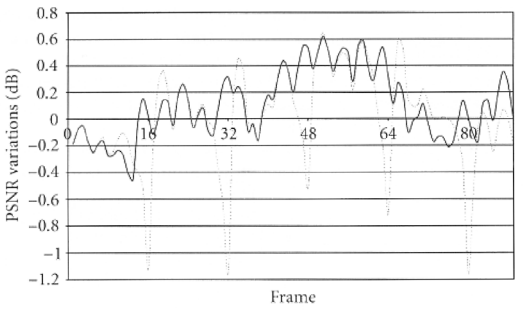


FIGURE 5: PSNR variations for the 3D scan-based temporal DWT (continuous) and the 3D temporal tiling approach (dashed) on the 89 first luminance frames of the sequence Akiyo at 80 kbps (25 fps). 9/7 DWT with two levels of decomposition; bitrate control for groups of 16 frames.

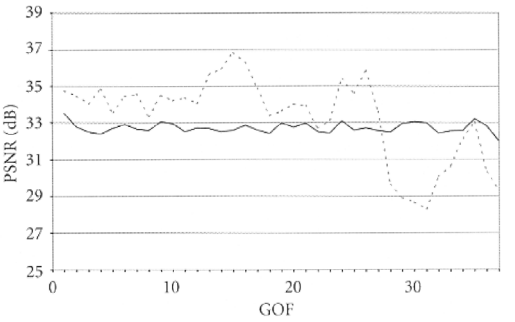


FIGURE 6: PSNR of each group of height frames (GOF) for the proposed quality control procedure (continuous) and a bitrate control procedure (dashed). The sequence is Foreman at 890 kbps (30 fps) in both cases.

(4) For each 3D subband i, compute $\ln(a_i/\lambda\Delta_i\pi_i\sigma_i^2) = \ln(-h)$ with the optimal λ and read q_i/σ_i using the tables shown in Figure 3. This q_i is the optimal quantization step for subband i.

The tables shown in Figures 3 and 4 are stored for several shape parameters α. They are valid for any video sequence.

4. EXPERIMENTAL RESULTS

To show the efficiency of our 3D scan-based wavelet transform method in removing the temporal blocking artifacts (jerks), we first extended EBWIC [13] to 3D data. The quantized wavelet coefficients have been encoded using JPEG2000's bit-plane context-based arithmetic coder [14]. We first encoded a sequence with the proposed 3D scan-based temporal wavelet transform and a bitrate regulation for the temporally coherent coefficients of each group of 16 frames. Then, we encoded the same sequence with the block-based approach, where the temporal wavelet coefficients and their encoding were performed on independent temporal blocks of 16 frames. Figure 5 shows a global PSNR improvement of mean 0.11 dB with our approach. Furthermore, we have reduced the PSNR variance from 0.13 to 0.06. The peaks of the block-based approach fit with the artifacts produced at temporal tiles borders (jerks). Regarding the visual quality, the proposed method is also better since the annoying jerks are cancelled out.

Then, we replaced the bitrate regulation by our new MSE allocation procedure. Figure 6 shows that the quality of successive groups of 8 frames is well controlled. The PSNR variations are less than 1 dB with our method while they were up to 9 dB with a bitrate control procedure. The global sequence PSNR is 32.7 dB in both cases. Therefore, our method provides the same global rate-distortion performance but ensures constant quality output frames. This results in a better visual quality.

5. CONCLUSION

In this paper, we have proposed methods for efficient quality control in video-coding applications.

In Section 2, we have proposed a 3D scan-based DWT method which allows the computation of the temporal wavelet decomposition of a sequence with infinite length using few memory and no extra CPU. Compared to temporal tiling approaches often used to reduce memory requirements, our method avoids temporal tiles artefacts. We have also shown in Section 2.3 that, for the same memory requirements, our method reduces the encoding delay. We have proposed the scan-based motion compensated lifting which results in both saving memory and temporal quality control.

In Section 3, we have proposed a new efficient model-based quality control procedure. This bit allocation procedure controls the output frames quality over time. The extension to scalar quantizers with a deadzone [31, 32, 33] is straightforward.

These methods combine the advantages of wavelet coding (performance, scalability) with minimum memory requirements and low CPU complexity.

REFERENCES

[1] G. Karlsson and M. Vetterli, "Three-dimensional subband coding of video," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1100–1103, New York, NY, USA, April 1988.

[2] C. I. Podilchuk, N. S. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Processing*, vol. 4, no. 2, pp. 125–139, 1995.

[3] B. Felts and B. Pesquet-Popescu, "Efficient context modeling in scalable 3D wavelet-based video compression," in *Proc. IEEE International Conference on Image Processing*, Vancouver, BC, Canada, September 2000.

[4] A. Wang, Z. Xiong, P. A. Chou, and S. Mehrotra, "Three-dimensional wavelet coding of video with global motion compensation," in *Proc. IEEE Data Compression Conference*, pp. 404–414, Snowbird, Utah, USA, March 1999.

[5] J. Xu, S. Li, Y.-Q. Zhang, and Z. Xiong, "A wavelet video coder

using three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," in *Proc. IEEE Pacific-Rim Conf. on Multimedia*, Sydney, Australia, December 2000.

[6] S. J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 155–167, 1999.

[7] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 572–588, 1994.

[8] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. IEEE Data Compression Conference*, pp. 251–260, Snowbird, Utah, USA, March 1997.

[9] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559–571, 1994.

[10] P. Charbonnier, M. Antonini, and M. Barlaud, "Implantation d'une transformée en ondelettes 2D dyadique au fil de l'eau," CNES contract report 896/95/CNES/1379/00, CNES, October 1995.

[11] C. Parisot, M. Antonini, M. Barlaud, C. Lambert-Nebout, C. Latry, and G. Moury, "On board stripe-based wavelet image coding for future space remote sensing missions," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, pp. 2651–2653, Honolulu, Hawaii, July 2000.

[12] C. Chrysafis and A. Ortega, "Line based, reduced memory, wavelet image compression," *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 378–389, 2000.

[13] C. Parisot, M. Antonini, and M. Barlaud, "EBWIC: A low complexity and efficient rate constrained wavelet image coder," in *Proc. IEEE International Conference on Image Processing*, Vancouver, BC, Canada, September 2000.

[14] ISO/IEC 15444-1:2000, "Information technology—JPEG 2000 image coding system," 2000.

[15] C. Parisot, M. Antonini, and M. Barlaud, "3D scan-based wavelet transform for video coding," in *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 403–408, Cannes, France, October 2001.

[16] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.

[17] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Processing*, vol. 4, no. 8, pp. 1053–1060, 1995.

[18] C. Parisot, *Allocations basées modèles et transformée en ondelettes au fil de l'eau pour le codage des images et des vidéos*, Ph.D. thesis, University of Nice-Sophia Antipolis, Nice, France, January 2003.

[19] J.-R. Ohm, "Motion-compensated wavelet lifting filters with flexible adaptation," in *Proc. Tyrrhenian International Workshop on Digital Communications*, Palazzo dei Congressi, Capri, Italy, September 2002.

[20] J. Viéron, C. Guillemot, and S. Pateux, "Motion compensated 2D+t wavelet analysis for low rate FGS video compression," in *Proc. Tyrrhenian International Workshop on Digital Communications*, Palazzo dei Congressi, Capri, Italy, September 2002.

[21] T. Wiegand and B. Girod, *Multi-frame Motion-Compensated Prediction for Video Transmission*, Kluwer Academic, Boston, Mass, USA, 2001.

[22] C. Parisot, M. Antonini, and M. Barlaud, "Motion-compensated scan based wavelet transform for video coding," in *Proc. Tyrrhenian International Workshop on Digital Communications*, Palazzo dei Congressi, Capri, Italy, September 2002.

[23] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[24] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.

[25] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distorsion sense," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 160–176, 1993.

[26] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic, Boston, Mass, USA, 1992.

[27] A. Ortega, "Variable bit-rate video coding," in *Compressed Video Over Networks*, M.-T. Sun and A. R. Reibman, Eds., pp. 343–382, Marcel Dekker, New York, NY, USA, 2000.

[28] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proc. IEEE Data Compression Conference*, pp. 387–395, Snowbird, Utah, USA, April 1996.

[29] M. Barlaud, *Wavelets in Image Communication*, Elsevier, Amsterdam, Netherlands, 1994.

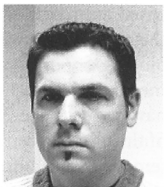
[30] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[31] C. Parisot, M. Antonini, and M. Barlaud, "Optimal nearly uniform scalar quantizer design for wavelet coding," in *Visual Communications and Image Processing*, vol. 4671 of *SPIE Proceedings*, San Jose, Calif, USA, January 2002.

[32] C. Parisot, M. Antonini, and M. Barlaud, "Stripe-based MSE control in image coding," in *Proc. IEEE International Conference on Image Processing*, Rochester, NY, USA, September 2002.

[33] P. Raffy, M. Antonini, and M. Barlaud, "Non-asymptotical distortion-rate models for entropy coded lattice vector quantization," *IEEE Trans. Image Processing*, vol. 9, no. 12, pp. 2006–2017, 2000.

Christophe Parisot graduated and received the M.S. degree in computer vision from the École Supérieure en Sciences Informatiques (ESSI), Sophia Antipolis, France, in 1998. He will receive the Ph.D. degree in image processing from the University of Nice-Sophia Antipolis, France, in 2003. His research interests include image and video compression, quantization, and bit allocation problems.



Marc Antonini received the Ph.D. degree in electrical engineering from the University of Nice-Sophia Antipolis, France, in 1991. He was a Postdoctoral Fellow at the Centre National d'Etudes Spatiales, Toulouse, France, in 1991 and 1992. Since 1993, he has been working with the CNRS at the I3S laboratory both from the CNRS and the University of Nice-Sophia Antipolis. He is a regular reviewer for several journals (IEEE Transactions on Image Processing, Information Theory and Signal Processing, IEE Electronics Letters) and participated in the organization of the IEEE Workshop Multimedia and Signal Processing 2001 in Cannes, France. He also participates in several national research and development projects with French industries, and in several international academic collaborations. His research interests include multidimensional image processing, wavelet analysis, lattice vector quantization, information theory, still image and video coding, joint source/channel coding, inverse problem for decoding, multispectral image coding, and multiresolution 3D mesh coding.

Michel Barlaud received his Thèse d'Etat from the University of Paris XII. He is currently a Professor of image processing at the University of Nice-Sophia Antipolis, and the leader of the Image Processing Group of I3S. His research topics are image and video coding using scan-based wavelet transform, inverse problem using half-quadratic regularization, and image and video segmentation using region-based active contours and PDEs. He is a regular reviewer for several journals, a member of the technical committees of several scientific conferences. He leads several national research and development projects with French industries and participates in several international academic collaborations (Universities of Maryland, Stanford, Boston, Louvain La Neuve). He is the author of a large number of publications in the area of image and video processing and the editor of the book *Wavelets and Image Communication* Elsevier, 1994.

Annexe C

Pyramidal Lattice Vector Quantization for Multiscale Image Coding

IEEE Transactions on Image Processing vol.3, no.4, pp. 367-381, juillet 1994

J'ai co-écrit cet article avec Michel Barlaud, Pierre Solé, Thierry Gaidon et Pierre Mathieu.

Résumé Le propos de cet article était d'introduire un nouveau schéma de codage d'images multirésolutions qui utilise la quantification vectorielle algébrique (QVA). Notre objectif était d'adapter la QVA à la statistique des images à quantifier. En effet, la construction d'un QVA est tributaire de la norme choisie pour l'indexation des vecteurs du réseau. Nous nous sommes intéressé par la modélisation Laplacienne de la distribution des coefficients d'ondelettes pour laquelle les surfaces de probabilité constante sont des hyper-sphères pour la métrique L_1 (hyper-pyramides). Nous avons donné une expression explicite des fonctions génératrices Nu pour les réseaux les plus courant tels que les réseaux $\mathbb{Z}^n$, D_n, E_8, et Λ_{16}.

Pyramidal Lattice Vector Quantization for Multiscale Image Coding

Michel Barlaud, *Member, IEEE*, Patrick Solé, Thierry Gaidon, Marc Antonini, and Pierre Mathieu

***Abstract*—The purpose of this paper is to introduce a new image coding scheme using lattice vector quantization. The proposed method involves two steps:**

- **biorthogonal wavelet transform of the image**
- **lattice vector quantization of wavelet coefficients**

In order to obtain a compromise between minimum distortion and bit rate, we must truncate and scale the lattice suitably. To meet this goal, we need to know how many lattice points lie within the truncated area.

In this paper, we investigate the case of Laplacian sources where surfaces of equal probability are spheres for the L^1 metric (pyramids) for arbitrary lattices. We give explicit generating functions for the codebook sizes for the most useful lattices like $\mathbf{Z}^n, D_n, E_8, \Lambda_{16}$.

I. Introduction

A NEW image coding scheme was introduced by the authors in [3]. First, a biorthogonal wavelet transform is applied to the image, and then, wavelet coefficients are vector quantized (see Fig. 1).

In the first step, a biorthogonal wavelet transform with multiresolution scale factor of $\sqrt{2}$ is preferred to the usual dyadic wavelet transform [27] since it is more isotropic and yields fewer artifacts [17].

In the second step, we avoid the well-known LBG method [26], which is computationally expensive and results in blur artifacts at low bit rate. Lattice point quantization is a nice alternative [8] since we can use fast encoding algorithms (see ch. 20 of [11] and [12]). For a bit rate of R b/sample, the number of codebook vectors, or equivalently of lattice points used, is 2^{nR}. For high bit rate and high spatial dimension n, the number 2^{nR} is not achievable by the LBG method (high complexity) but is achievable by, and easily implemented with, lattice quantizers. Thus, the so-called "codebook" is a particular subset of a regular arrangement of points in an n-dimensional space centered in zero (lattice).

In order to obtain the best trade-off distortion rate, we must scale and truncate the lattice suitably. To do this, we need to know how many lattice points lie within the truncated area. Hence, we need to know the shape of the truncated area. When the signal to be compressed has an i.i.d. multivariate Gaussian distribution, the surfaces of equal probability are ordinary spheres. The truncated area is then spherical. In these applications the size of the codebook was evaluated by use of the theta function of the lattice, which had been computed by generations of number theorists [11].

Motivated by image coding applications, Fischer [19] investigated the case of Laplacian sources (for cubic lattices) where surfaces of equal probability are spheres for the L^1 metric, which are sometimes called *pyramids*. Due to nonexistence of theta functions for the L^1 metric, he had to restrain himself to cubic lattices. In his approach, the radius and codebook index of vectors lying on the pyramid are coded.

In this paper, we generalize Fischer's approach to arbitrary lattices and give explicit generating functions for the codebook sizes for the most useful lattices like $\mathbf{Z}^n, D_n, E_8$, etc ... Moreover, we use as a codebook the lattice points inside a *pyramid* of a given height. Beyond this height, vectors are normalized to be on this largest pyramid. Within this pyramid, vectors to be encoded are scaled down to the closest concentric pyramid and approximated by the lattice point that is the closest in the sense of the mean square error criterion. This enables us to employ the fast algorithms of ch. 20 of [11].

The paper is organized as follows. Section II deals with wavelet transform. A brief review on vector quantization is presented in Section III. In Section IV, we present a new scheme involving pyramidal lattice vector quantization design. Then, usual lattice vector quantizers and their associated fast decoding algorithms are investigated. We then present a new theory for counting lattice points in the pyramidal case in Section V and labeling in Section VI. Finally, Section VII presents experimental results at low bit rate.

II. Wavelet Transform

A. First Principles

Wavelets are functions generated from one single function Ψ by dilations and translations

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \quad a \in \mathbf{R}^{*+}, b \in \mathbf{R} \qquad (1)$$

where a is the scaling parameter, whereas b is the shift parameter.

The mother wavelet must satisfy the admissibility condition

$$\int_{-\infty}^{+\infty} \frac{|\hat{\Psi}(\omega)|^2}{|\omega|} d\omega < +\infty$$

where $\hat{\Psi}$ is the Fourier transform of Ψ.

Manuscript received April 24, 1992; revised July 16, 1993. This work was supported by DRET under the convention no. 90/206. The associate editor coordinating the review of this paper and approving it for publication was Prof. William A. Pearlman.

The authors are with CNRS—University of Nice Sophia Antipolis, Valbonne, France.

IEEE Log number 9400311.

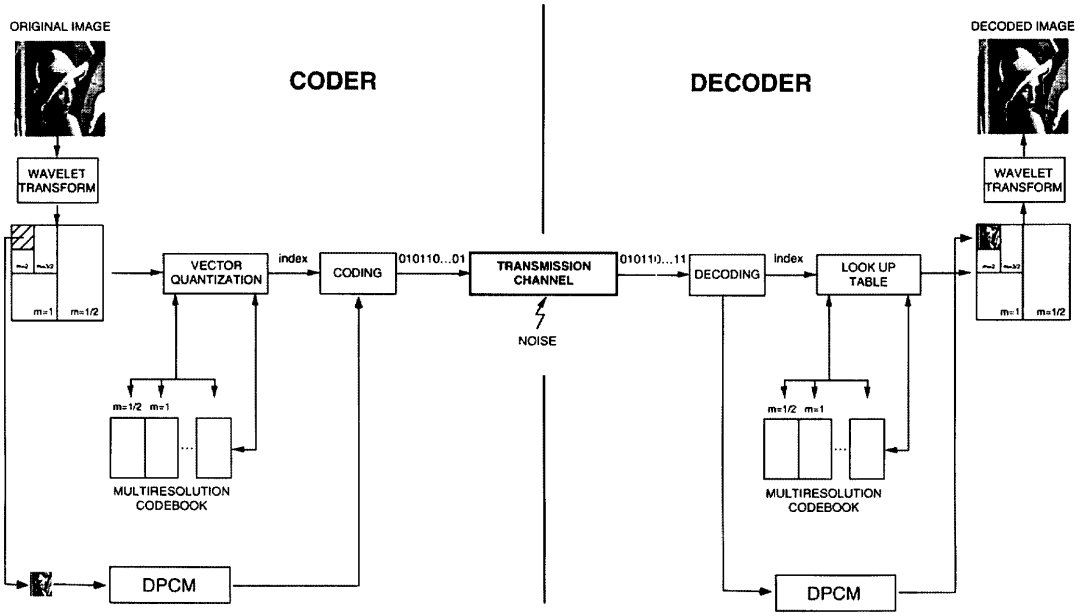


Fig. 1. Encoding scheme.

The basic idea of the wavelet transform is to represent any arbitrary function f as a superposition of wavelets

$$f = \sum_{m,n} c_{m,n}(f)\Psi_{m,n} \text{ with } m, n \in \mathbf{Z} \tag{2}$$

where

$$c_{m,n}(f) = \langle \Psi_{m,n}, f \rangle = \int_{-\infty}^{+\infty} \Psi_{m,n}(x) f(x) dx. \tag{3}$$

For $a = 2^m, b = n2^m$, there exist special choices of Ψ such that the $\Psi_{m,n}$ constitute an orthonormal basis [14]. The wavelet is therefore

$$\Psi_{m,n}(x) = 2^{-m/2}\Psi(2^{-m}x - n) \tag{4}$$

where m is the scaling parameter, whereas n is the shift parameter.

Orthonormal compact wavelets were designed by Daubechies. In this case, wavelet coefficients are computed using a multiresolution analysis as follows [16].

To introduce the multiresolution notion, we must define a scaling function

$$\phi_{m,n}(x) = 2^{-m/2}\phi(2^{-m}x - n). \tag{5}$$

The projection on this family of functions $\phi_{m,n}$ gives an approximation of a signal f with resolution 2^{-m}. The wavelet coefficients $c_{m,n}$ describe the information lost when going from an approximation of f with resolution 2^{-m+1} to a coarser approximation with resolution 2^{-m}; see Fig. 2.

These functions can be constructed using an infinite product from a function $H(\omega)$. This function is connected to digital

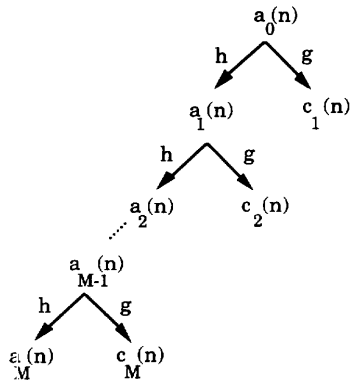


Fig. 2. Multiresolution decomposition of a signal a_0.

filters. Daubechies showed that an appropriate choice of a function $H(\omega)$ generates regular wavelets. Let $H(\omega)$ be a function of $L^2([0, 2\pi[)$ such that $H(0) = 1, H(\pi) = 0$, and $|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1$.

Moreover, if the function satisfy the additional condition that the infinite product $\prod_{n=0}^{+\infty} H\left(\frac{\omega}{2^n}\right)$ decays faster than $C(1 + |\omega|)^{-\nu-0.5}$ with $\nu > 0$ for large $|\omega|$, then we define the scaling function ϕ by its Fourier transform

$$\hat{\phi}(2\omega) = H(\omega)\hat{\phi}(\omega) = \prod_{n=0}^{\infty} H\left(\frac{\omega}{2^n}\right).$$

TABLE I
COEFFICIENTS OF FILTERS h AND h^* (NINE AND SEVEN TAPS)

n	0	± 1	± 2	± 3	± 4
$(1/\sqrt{2})h(n)$	0,602949	0,266864	-0,078223	-0,016864	0,026749
$(1/\sqrt{2})h^*(n)$	0,557543	0,295636	-0,028772	-0,045636	0

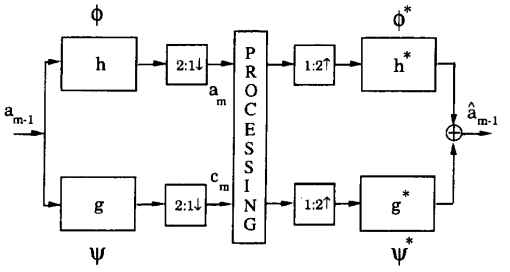


Fig. 3. Basic structure for biorthogonal wavelet transform.

The wavelet is therefore given by

$$\hat{\Psi}(2\omega) = H(\omega + \pi)e^{-i\omega}\hat{\phi}(\omega) \tag{6}$$

B. Biorthogonal Wavelet Basis

For image coding, it would be nice to have an *exact reconstruction sub-band coding scheme (orthogonal bases).* In order to have fast computation, the *filters should be short.* Since filters with a nonlinear phase give poor results in image coding they should be also *symmetric.* Unfortunately, these requirements conflict with each other [16].

Nevertheless, it is possible to preserve some of the properties by relaxing the orthonormality requirement using biorthogonal bases. Biorthogonal wavelet bases were introduced by Cohen *et al.* [9] and were extensively used by the authors [3], [7] in image coding. The basic idea of the biorthogonal wavelet transform is to represent any arbitrary function f as a superposition of wavelets of the type:

$$\Psi_{m,n}(x) = 2^{-m/2}\Psi(2^{-m}x - n) \tag{7}$$

(scaling factor m and shift parameter n). Thus

$$f = \sum_{m,n} c_{m,n}(f)\Psi^*_{m,n} \tag{8}$$

and

$$c_{m,n}(f) = \langle \Psi_{m,n}, f \rangle = \int_{-\infty}^{+\infty} \Psi_{m,n}(x)f(x)dx \tag{9}$$

where $\Psi^*_{m,n}$ is the dual basis of $\Psi_{m,n}$ [9].

As the orthonormal case, we introduce multiresolution analysis. We must define two dual scaling functions: $\phi_{m,n}(x) = 2^{-m/2}\phi(2^{-m}x - n)$ and the dual scaling function $\phi^*_{m,n}$.

The projection on this family of functions $\phi_{m,n}$ gives an approximation of a signal f with resolution 2^{-m}. The wavelet coefficients $c_{m,n}$ describe the information lost when going from an approximation to f with resolution 2^{-m+1} to a coarser approximation with resolution 2^{-m}. The signal can be reconstructed using the dual basis functions; see Figs. 2 and 3.

C. Application to Sampled Signals

This theory is translated into the following sub-band decomposition algorithm for the computation of the wavelet coefficients $c_{m,n}$ (for more details, see [27])

$$c_{m,n} = \sum_k g_{2n-k}\, a_{m-1,k}$$
$$a_{m,n} = \sum_k h_{2n-k}\, a_{m-1,k} \tag{10}$$

where h is a low-pass filter and g is a high-pass filter. Basically, there are two pairs of filters: one pair h and g for decomposition and another pair h^* and g^* for reconstruction. The relations between the different filters are given by the system [9]

$$g^*_n = (-1)^n h_{-n+1} \text{ and } g_n = (-1)^n h^*_{-n+1} \tag{11}$$
$$\sum_n h_n h^*_{n+2k} = \delta_{k,0} \tag{12}$$

which ensures exact reconstruction

$$a_{m-1,n} = \sum_k (h^*_{2n-k}a_{m,n} + g^*_{2n-k}c_{m,n}). \tag{13}$$

The condition on h and h^* reduces, for symmetric filters, to the construction of two trigonometric polynomials $H(\omega)$ and $H^*(\omega)$ such that $H(\omega)H^*(\omega) + H(\omega + \pi)H^*(\omega + \pi) = 1$. ($h$ and h^* are related to $H(\omega)$ and $H^*(\omega)$ by the Fourier transform.)

Regularity for Ψ and Ψ^* implies that $H(\omega)$ and $H^*(\omega)$ must be divisible, respectively, by $(1 + e^{-j\omega})^\alpha$ and $(1 + e^{-j\omega})^{\alpha^*}$. Thus, Ψ and Ψ^* are, respectively, $(\alpha - 1)$ and $(\alpha^* - 1)$ continuously differentiable [9]

For our simulations, we have chosen a spline-variant family of filters with less dissimilar lengths to make the lengths of h and h^* as close as possible. This family provides an analysis filter h with nine taps and a synthesis filter h^* with seven taps [7] (see Table I). This choice gives the best tradeoff between lengths of filters and the regularity of the associated wavelets.

D. Extension to the 2-D Case

Various extensions of the 1-D wavelet transform to higher dimensions exist. The most straightforward way to generate 2-D wavelet transforms is to apply two 1-D wavelet transforms separately [27], [37]–[39] (separable pyramid or dyadic multiresolution analysis). This produces a multiresolution scale factor of 2 and privileges horizontal and vertical orientations. Such scaling transforms are characterized by scaling functions that can be written

$$\phi_{m,n}(x, y) = \phi_{m,n}(x)\phi_{m,n}(y)$$

and by three 2-D wavelets.

TABLE II
COEFFICIENTS OF BIDIMENSIONAL FILTERS h AND h^* (9×9 AND 7×7 TAPS)

	h	h^*
a	0.001671	-
b	-0.002108	-0.005704
c	-0.019555	-0.007192
d	0.139756	0.164931
e	0.687859	0.586315
f	0.006687	-
g	-0.006324	-0.017113
i	-0.052486	-0.014385
j	0.010030	-

Such a multiresolution analysis provides, at each resolution level, three wavelet coefficient sub-images having different orientations and one low-resolution sub-image. This scheme is well known in sub-band coding [32].

We prefer to use the quincunx wavelet transform. This algorithm uses nonseparable and nonoriented filters. It decomposes the original image with a multiresolution scale factor of $\sqrt{2}$. This means that this new analysis will be twice as fine as the dyadic multiresolution analysis [17], [1]. The scaling function is then defined by the following formula:

$$\phi_{m,n}(x,y)=2^{-m}\phi(\mathcal{L}^{-2m}(x,y)-n)\ n\in\mathbf{Z}^2, n=(n_x,n_y)$$

where $\mathcal{L}(x,y) = (x+y, x-y)$ is a linear transform with $m \in \frac{1}{2}\mathbf{Z}$ and $\mathcal{L}^n = \mathcal{L}\circ\mathcal{L}\circ..\circ\mathcal{L}$ n times. This isotropic multiresolution analysis provides, at each resolution level, only one wavelet coefficients sub-image and one low-resolution sub-image. The 1-D filters defined in Section II-C can be extended to the quincunx 2-D case by applying the transformation

$$\cos(\omega) \rightarrow \frac{1}{2}(\cos(\omega_x)+\cos(\omega_y))$$

to the 1-D trigonometric polynomials.

This transformation ensures that all the properties of the 1-D filters are satisfied by the 2-D filters. Note that this transform was also used by other authors [21], [40].

Table II presents coefficients of the 2-D filters for the quincunx 2-D case (see Fig. 4 for the associated scaling functions and Fig. 5 for the quincunx decomposition of Barbara image).

E. Statistics of Wavelets Coefficients

Wavelet transform is made to organize the information of the original image into several resolutions. The image is "split" into M images of wavelet coefficients (resolution 0 to $M-1$) and an image at the lowest resolution. [17].

Only the wavelets coefficients of the quincunx biorthogonal wavelet transform are vector quantized. In [7], statistics of coefficient's subimages were modelized by the generalized Gaussian. A particular case of this function is the Laplacian function. We use this Laplacian approximation to study the performances of the quantizer.

Statistics of the subimage at the lowest resolution are similar to the statistics of the original image. Thus, a DPCM technique is used to encode it.

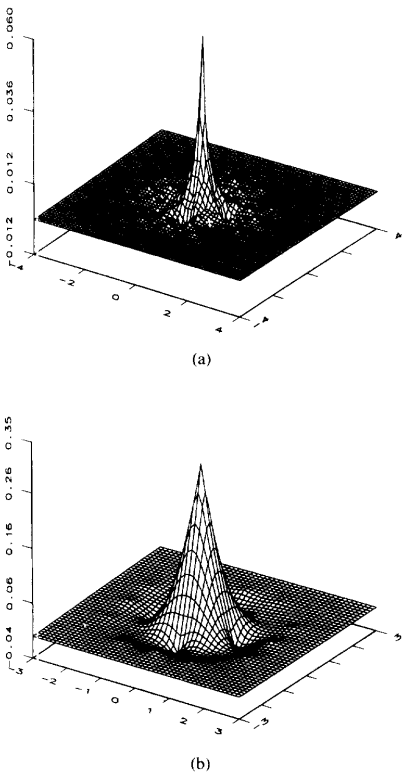


Fig. 4. (a) Scaling function associated with monodimensional nine taps filters; (b) seven taps filters.

III. VECTOR QUANTIZATION

Vector quantizers have been investigated extensively by many researchers [29], [36] in recent years.

Let X be a n-component source vector with joint probability density function (pdf) $f_X(x) = f_X(x_1,\ldots,x_n)$.

A vector quantizer Q of dimension n and size L is defined as a function that maps an arbitrary vector $X \in \mathbf{R}^n$ into one of L output or reproduction vectors $Y_1, Y_2, \ldots, Y_L$ called codewords belonging to $\mathbf{R}^n$. Thus, we have the chart

$$\begin{aligned} Q &: \mathbf{R}^n \rightarrow \mathcal{Y} \\ X &\mapsto Q(X) \end{aligned} \tag{14}$$

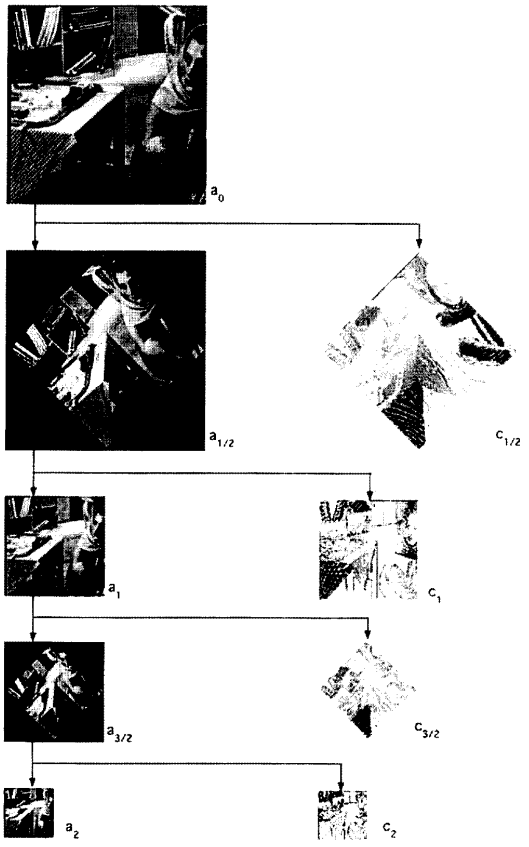


Fig. 5. Quincunx biorthogonal wavelet decomposition of Barbara image.

where $\mathcal{Y} = \{Y_1, \ldots, Y_L\}$ is the set of reproduction vectors called the codebook and $Q(X) = Y_i$ if $X \in \mathcal{C}_i$.

The vector quantizer is completely specified by listing the L output vectors Y_i and their corresponding nonoverlapping partitions $\mathcal{C}_i (i = 1, 2, \ldots, L)$ of $\mathbf{R}^n$ called Voronoi regions. A Voronoi region $\mathcal{C}_i$ is defined by the equation

$$\mathcal{C}_i = \{X \in \mathbf{R}^n / \|X - Y_i\| \leq \|X - Y_j\| \; i \neq j\}$$

and represents the subset of vectors of $\mathbf{R}^n$, which are well matched by the codeword Y_i of the codebook. $||.||$ denotes the usual L^2 norm. The total distortion per dimension of this quantizer is then given by

$$d_n = \frac{1}{n} E\{\|X - Q(X)\|^2\} = \frac{1}{n} \sum_{i=1}^{L} \int_{X \in \mathcal{C}_i} \|X - Y_i\|^2 f_X(x) dx.$$

For the purposes of transmission or storage, a binary word c_i of length b_i bits and called the index of the codeword is assigned to each output vector Y_i. Thus, vector quantization can also be seen as a combination of two functions: an encoder that views the input vector X and generates the index of the reproduction vector specified by $Q(X)$ and a decoder that uses this index to generate the reproduction vector Y_i. The average binary word length is given by the formula

$$H(Y) = -\sum_{i=1}^{L} p(Y_i) \log_2 p(Y_i) \text{ bits/vector}$$

which is the so-called entropy measure of the codebook that specifies the minimum bit rate necessary to achieve a distortion d_n with the chosen quantizer. This value is supposed achieved using an adaptive entropy coder.

IV. Pyramidal Lattice Vector Quantization

For a fixed vector length of practical size and, as we have seen from the previous paragraph, the scheme that entropy-codes its index sequence [10] is the one that achieves the same distortion d_n, with a lower average transmission rate. In fact, the ECVQ algorithm, which is a generalization of the LBG algorithm [26], is used to design vector quantizers having minimum distortion subject to an entropy constraint.

For high bit rate R and high dimension space n, the number 2^{nR} is not achievable by vector quantizers generated by the LBG or ECVQ algorithms. In fact, this optimal algorithm is computationally expensive.

In 1979, Gersho [23] conjectured that the optimal high-resolution ECVQ should have the form of a *lattice*. Thus, lattice quantizer methods have been investigated in image coding; see [19], [24], and [15]. Lattice coding has been used successfully in the past in vector quantization applications such as speech coding [2].

A. Lattices

A lattice in $\mathbf{R}^n$ is composed of all integral combinations of a set of linearly independent vectors. Thus, an n-dimensional lattice Λ_n is defined as a set of vectors

$$\Lambda_n = \{Y \in \mathbf{R}^m | Y = u_1 a_1 + \ldots + u_n a_n\} \qquad (15)$$

where $a_1, \ldots, a_n$ are linearly independent vectors in m-dimensional real Euclidean space $\mathbf{R}^m$ with $m \geq n$, and $u_1, \ldots, u_n$ are in $\mathbf{Z}$.

It was shown by Zador [31] that the quadratic error depended crucially on the geometry of the lattice by a term called the second moment of the lattice G_n [11]. If G_n denotes the average mean squared error per dimension for the best quantizing lattice in n dimensions, then Zador showed that $\lim(G_n) = \frac{1}{2\pi e} < \frac{1}{12}$ for large n. This demonstrates the interest of using multidimensional lattices. As Conway and Sloane put it: "it pays to procrastinate." In general, finding the best known quantizing lattices in n dimensions is a difficult task. They are only known for some integers $n \leq 24$.

B. Classical Lattices and Fast Decoding Algorithms

Investigation of the properties of lattices suggests that certain lattices should perform better than others. Conway and Sloane [11], [12] have determined the best known lattices for several dimensions as well as fast quantizing and decoding algorithms.

Classical Lattices Used: Some important n-dimensional lattices are the root lattices $A_n(n \geq 1), D_n(n \geq 3)$, and $E_n(n = 6, 7, 8)$ as well as the Barnes–Wall lattice Λ_{16} in dimension 16. These lattices give the best sphere packings and coverings [11] in their respective dimension. For our application, we have used the D_4, E_8, and Λ_{16}.

⇒ For $n \geq 3$, the D_n lattice is defined as follows:

$$D_n = \left\{ (y_1, y_2, \ldots, y_n) \in \mathbf{Z}^n \Big/ \sum_{i=1}^{n} y_i \text{ even} \right\} \quad (16)$$

and consists of those points of the rectangular lattice $\mathbf{Z}^n$ whose sum of coordinates is even.

The lattices E_8 and Λ_{16} are constructed by using, respectively, the lattices D_8 and D_{16}.

⇒ The lattice E_8 is defined by

$$E_8 = D_8 \cup \left(\frac{1}{2}\mathbf{1} + D_8 \right) \quad (17)$$

where $\mathbf{1}$ stands for the all-one vector.

⇒ Finally, the Barnes–Wall (Λ_{16}) lattice is constructed as a union of cosets of D_{16}

$$\Lambda_{16} = \cup_{i=0}^{31} (r_i + 2D_{16}) \quad (18)$$

where the translation vectors r_i correspond to the codewords of the first Reed–Muller code of length 16.

Quantizing Algorithm: If the lattice is used as a quantizer, all the points in the Voronoi region around the lattice point Y are represented by Y in terms of mean squared error (MSE) distortion. In fact, a lattice quantizer may be defined as a quantizer whose output set $\mathcal{Y}$ is a subset of a lattice.

Conway and Sloane have developed fast quantization algorithms for lattices that are unions of cosets of the root lattice D_n. Their results are explained in [12].

Unlike LGB-type algorithms, there is practically no need to compute a norm to find (among all of the vectors in the codebook) the best reproduction vector.

C. Encoding and Decoding Schemes

The source is a subimage of the wavelet transform. This subimage is split into patterns. The values of a pattern (pixel values) are the components of a vector X.

In order to optimize the tradeoff between minimum distortion and bit rate using a lattice quantizer, we truncate and scale the lattice. To encode the vectors, we project them on or within surfaces of constant energy using a scaling factor or "quantization step." In the case of an i.i.d. Gaussian distribution, these surfaces are spheres. In the case of an i.i.d. Laplacian distribution, these are the so-called pyramids [18] or hyperoctahedra of $\mathbf{R}^n$. They are defined by the relation (for a radius m):

$$S(n, m) = \left\{ Y \in \mathbf{R}^n \Big/ \sum_{i=1}^{n} |Y_i| = m \right\}. \quad (19)$$

Different lattice encoding/decoding schemes have been investigated in order to obtain the best tradeoff between distortion and entropy.

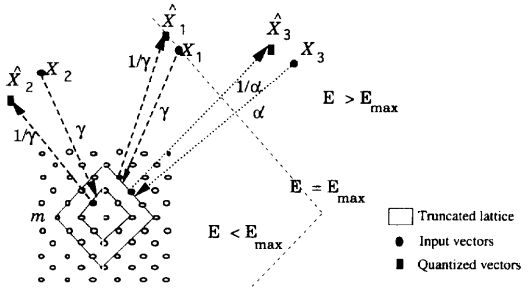


Fig. 6. Pyramidal lattice vector quantization scheme.

Encoding Scheme: Let m be the radius of the largest pyramid or sphere. Instead of scaling the lattice codebook, we scale the source vector in order to use fast encoding algorithm described in [12]; see Fig. 6.

The following algorithm is used for quantization:

— Select a dictionary size (determine a lattice truncation energy m). The truncated shape is a multidimensional sphere in the case of a source with gaussian pdf and a multidimensional pyramid for a laplacian pdf;

— Select a maximum energy $E_{\max}$ for the source to be encoded (which is not necessarily the maximum energy of the source);

— Norm the source by applying the scaling factor γ, defined hereafter, to all of the source vectors ($\gamma = \frac{E_{\max}}{m}$ for pyramids and $\gamma = \sqrt{E_{\max}/m}$ for spheres). This scaling factor enables quantization of the source vectors with energy $E_{\max}$ by vectors from the outermost shell of the dictionary. To modify the value of γ, we modify the value of $E_{\max}$. Source vectors with energy greater than $E_{\max}$ are treated separately.

— Afterwards, quantization is performed.

More precisely for each source vector, do the following:

① Scale the source vector by the scaling factor γ;

② For this γ, compute the radius of the sphere or pyramid associated to the scaled source vector. If this radius is greater than m, modify the scaling factor and project the source vector according one of the following schemes i), ii), and iii);

③ Quantize the scaled source vector using fast algorithm in Section IV-B-2 (these vectors are belonging to the truncated lattice);

④ Encode the index of the nearest lattice vector using entropy coding;

⑤ End.

i) In a first approach also used by the authors [6] and by Gibson [25], this vector is scaled on the largest shell using a different scaling factor. Then, goto ③.

ii) This method looks like the previous one. Compute a suitable scaling factor for this "high energy" vector in order to project it on or within the largest sphere or pyramid. Quantize Round(α) and send it on the transmission channel. Then goto ③.

This results in an increase of entropy of $\log_2(\text{Round}(\alpha))$ but also improves distortion.

iii) We code each component of the source vector on its basis instead of scaling. The problem now is to compute the encoding cost of each component x_i of the vector. We use the method of Conway and Sloane described in [13]. The method is briefly described in Section VI.

Decoding Scheme: For each source vector, do the following:

① Decode the index;
② Find the lattice vector associated with this index;
③ Rescale the lattice vector using $1/\gamma$ or one of the three following schemes i), ii), or iii), according to the previous subsection.
④ End.

i) Rescale the lattice vector by the inverse of the scaling factor γ. This scheme gives high distortion for the greatest wavelet coefficients, which results in blurring artifacts. Moreover, remember that these coefficients represent edges (pdf tails) that are of most interest for image understanding. Then, goto ④.
ii) Rescale the lattice vector by the inverse of Round(α). Then, goto ④.
iii) Do nothing. Goto ④.

Now, the problem is to compute the number of vectors lying in pyramid of given radius for different lattices. This determines the size of the codebook.

V. Counting Lattice Points

To know the number of output lattice points, we must be able to calculate this number on or within surfaces with the same probability for the source pdf.

A classical approach to enumeration is the use of suitable generating functions. In a lattice Λ, the number N_m of points with energy m^2 (squared L^2 distance m^2 from the origin) is given by the theta series of the lattice [11]

$$\theta_\Lambda(q) = \sum_{Y\in\Lambda} q^{\|Y\|} = \sum_{m=0}^{+\infty} N_m q^m. \tag{20}$$

The total number of points contained in the sphere of radius m (surface energy m^2) is then $\Sigma_{i=0}^{m} N_i$. Equivalently, it is the coefficient of q^m in $\frac{\theta_\Lambda(q)}{(1-q)}$. This function is very useful for computing the number of points in spheres but useless for computing points in pyramids. Furthermore, the theta function of classical lattices has been studied and computed by number theorists for other purposes [11], and no such function is available for the L^1 metric. We thus need to develop an entirely new theory.

We introduce the Nu functions and summarize new results about them. We give the formulas for some classical lattices. Then, we compare the number of points belonging to a codebook for the L^1 and L^2 norms.

Notation

— To denote the coefficient of z^m in $f(z)$, we shall use the symbol $[z^m]f(z)$.
— For $Y \in \mathbf{R}^n$, let $||Y||_1 = \Sigma_{i=1}^{n}|y_i|$ (norm L^1) and $||Y||_2 = \sqrt{\Sigma_{i=1}^{n}|y_i|^2}$ (norm L^2).

A. Nu Functions of Lattices

For a lattice Λ, we define its *Nu* function by the relation

$$\nu_\Lambda(z) = \sum_{Y\in\Lambda} z^{\|Y\|_1} = \sum_{m=0}^{+\infty} [z^m]\{Y \in \Lambda / \|Y\|_1 = m\}. \tag{21}$$

Note the analogies and differences with the definition of the theta function [11].

$$\theta_\Lambda(q) = \sum_{Y\in\Lambda} q^{\|Y\|_2^2} = \sum_{m=0}^{+\infty} [q^m]\{Y \in \Lambda / \|Y\| = m\}. \tag{22}$$

The coefficient of z^m denotes the number of points lying on the "pyramid" of energy m.

Caution: Unlike the Euclidean norm, the norm $\|.\|_1$ depends on the chosen orthogonal basis. Different bases may yield different functions ν_Λ for the same lattice Λ. An example of this situation will be given for $\Lambda = E_8$, where construction A and construction B yield different *Nu* functions.

B. Easy Examples of Nu Functions: $\mathbf{Z}$, $\mathbf{Z}^n$, and $\mathbf{D}_n$

We start with dimension 1, where the well-known uniform quantizer (lattice $\mathbf{Z}$) has Nu function

$$\nu_Z(z) = 1 + 2\sum_{n=1}^{+\infty} z^n = 1 + \frac{2z}{1-z} = \frac{1+z}{1-z} \tag{23}$$

which is a geometric serie.

The lattice of even integers has *Nu* function (by scaling)

$$\nu_{2Z}(z) = \sum_{n=1}^{+\infty} z^{2\|Y\|_1} = \nu_Z(z^2) = \frac{1+z^2}{1-z^2}. \tag{24}$$

The *Nu* function of the set of odd integers, which is a coset of the preceding into $\mathbf{Z}$ (by difference)

$$\nu_{1-2\mathbf{Z}}(z) = \nu_{\mathbf{Z}}(z) - \nu_{2\mathbf{z}}(z) = \frac{2z}{1-z^2}. \tag{25}$$

Since the cubic lattice is a cartesian product of n 1-D lattices, we obtain

$$\nu_{\mathbf{Z}^n}(z) = (\nu_{\mathbf{Z}}(z))^n = \left(\frac{1+z}{1-z}\right)^n. \tag{26}$$

As D_n consists of those vectors of $\mathbf{Z}^n$ whose L^1 norm is even, we see that its *Nu* function is the even part of $\nu_{\mathbf{Z}^n}(z)$, that is

$$\nu_{D_n}(z) = \frac{1}{2}(\nu_{\mathbf{Z}^n}(z) + \nu_{\mathbf{Z}^n}(-z)). \tag{27}$$

Nu functions for denser lattices are slightly more complicated to define. We use hereafter some theory using blockcodes to define the *Nu* function for E_8, Λ_{16}.

C. Connection with Block Codes

Important lattices (root lattices, Leech lattice) can be constructed from algebraic codes. Thus, we remember hereafter some useful notions to define the *Nu* functions. For more details and complete proofs, see [30]–[34]. Let $F_2 = \{0, 1\}$ denote the finite field with two elements. For us, a *binary linear code* $\mathcal{C}$ will be a linear subspace of F_2^n. $\mathcal{C}$ is a set of n-dimensional binary vectors or *codewords*. The *weight* of a binary vector, henceforth denoted by $|c|$, is the number of its nonzero coordinates. The *Hamming weight enumerator* of a code $\mathcal{C}$ is defined as the formal polynomial with indeterminates x and y.

$$W_c(x, y) = \sum_{c \in \mathcal{C}} x^{n-|c|} y^{|c|} = \sum_{i=0}^{n} A_i \, x^{n-i} y^i$$

where A_i denotes the number of codewords at Hamming distance i from a codeword $c \in \mathcal{C}$ (weight i). The Hamming weight enumerator classifies codewords according to the number of nonzero coordinates. Variable x counts the number of zeros in a codeword (y counts the 1).

We denote by $|\mathcal{C}|$ the cardinality of $\mathcal{C}$. For more details on block codes, we refer to [11].

A *Nu* function is defined using weight enumerator. The weight enumerator is associated to the lattice according to the type of the construction. Constructions A and B are two ways to construct lattices using block codes. We give theorems and corollaries for each construction and determine which of the two is better.

Construction A: The simplest way to associate a lattice with a code $\mathcal{C}$ is construction A.

$$A(\mathcal{C}) := \{Y = (y_1, \ldots, y_n) \in \mathbf{Z}^n / \exists c \in \mathcal{C}, Y \equiv c(\bmod 2)\}.$$

The following result characterizes the *Nu* function of lattices constructed from a binary code by the construction A. The proofs of the propositions of this section can be found in [34].

Theorem 1: Let $\mathcal{C}$ be a binary linear block code with weight enumerator $W(x, y)$. Let $A(\mathcal{C})$ be the sublattice of $\mathbf{Z}^n$ of vectors congruent to $\mathcal{C}$ mod 2. Then

$$\nu_{A(c)}(z) = W_c(\nu_{2\mathbf{Z}}(z), \nu_{1+2\mathbf{Z}}(z)) = \frac{W_c(1 + z^2, 2z)}{(1 - z^2)^n}.$$

Applying the complex variable techniques of [20] yields the following asymptotic estimate of the number of points lying in a pyramid of radius m.

Corollary 1: For a given code and m large, we have

$$[z^m] \frac{\nu_{A(c)}(z)}{1 - z} \approx |\mathcal{C}| \frac{m^n}{n!}.$$

When $\mathcal{C}$ is the universe code of size 2^n, the lattice $A(\mathcal{C})$ is $\mathbf{Z}^n$, and the RHS is the volume of the pyramid of radius m due to the Gauss counting principle [28], which says that the number of integer points in a convex body is roughly equal to its volume.

⇒ Denoting by U_n the universe code F_2^n, the associated weight enumerator is $W_{U_n}(x, y) = (x + y)^n$, and we also see that $A(U_n) = \mathbf{Z}^n$.

Thus, replacing the value of the w.e. into the formula of Theorem 1 gives the *Nu* function of lattice $\mathbf{Z}^n$. $V_{\mathbf{Z}^n}(z) = \frac{((1+z^2)+(2z))^n}{(1-z^2)^n} = \frac{(1+z)^n}{(1-z)^n}$, which is the same result of previous paragraph.

⇒ Root lattice D_n. We can say that $D_n = \mathrm{A}(\mathrm{EW}n)$, where EW_n denotes the code of even weight vectors, which is better known as the parity-check code. Its w.e. collects the even powers of y in the w.e. of U_n.

$$W_{EW_n}(x, y) = \frac{1}{2}\left((x + y)^n + (x - y)^n\right).$$

From Theorem 1, replace W_{EW_n} in the formula. The *Nu* function becomes

$$\nu_{EW_n} = \frac{\frac{1}{2}\left[\left(\underbrace{1 + z^2}_{x} + \underbrace{2z}_{y}\right)^n + \left(\underbrace{1 + z^2}_{x} - \underbrace{2z}_{y}\right)^n\right]}{(1 - z^2)^n}.$$

Noticing that $\frac{(1+z)^2}{1-z^2} = \frac{1+z}{1-z}$ as well as $\frac{(1-z)^2}{1-z^2} = \frac{1-z}{1+z}$, we get $\nu_{D_n} = \frac{1}{2}(\nu_{Z^n}(z) + \nu_{Z^n}(-z))$ as in the preceding subsection.

⇒ From the Hamming code H_8 of length 8, with weight enumerator $W_8(x, y) = x^8 + 14x^4y^4 + y^8$, (see [11] or any standard textbook on codes), we get

$$\nu_{E_8}(z) = \frac{(1 + z^2)^8 + 224z^4(1 + z^2)^4 + 256z^8}{(1 - z^2)^8}.$$

Construction B: Construction B of [11] is defined as follows. Let $\mathcal{C}$ denote a binary linear code whose weights are multiples of 4 (such a code is usually called "doubly even"). With this code, construction B associates a lattice $B(\mathcal{C})$ with the formula

$$B(\mathcal{C}) := \left\{Y = (y_1, \ldots, y_n) \in \mathbf{Z}^n / \exists c \in \mathcal{C}, \right. \\ \left. X \equiv \mathcal{C} \bmod (2); \sum_{i=0}^{n} y_i \equiv 0 \bmod (4)\right\}.$$

Theorem 2: Let $\mathcal{C}$ be a doubly even binary code of length n and weight enumerator W. The *Nu* function of $B(\mathcal{C})$ is $\nu_{B(\mathcal{C})}(z) = \frac{1}{2} W_c\left(\frac{1+z^2}{1-z^2}, \frac{2z}{1-z^2}\right) + \frac{1}{2}\left(\frac{1-z^2}{1+z^2}\right)^n$.

Corollary 2 gives the asymptotic estimate of the number of points lying in a pyramid of radius m in construction B.

Corollary 2: For a given code and m large

$$[z^m] \frac{\nu_{B(\mathcal{C})}(z)}{1 - z} \approx \frac{|C|}{2} \frac{m^n}{n!}.$$

This shows that if a lattice can be constructed both by construction A and construction B, they will yield different orientations, and in addition, orientation B will have asymptotically fewer points than construction A (for the same lattice, the code used for construction B is necessarily smaller than the one used in construction A).

⇒ Applied to the code $R_8 = \{0^8, 1^8\}$ with weight enumerator $W = x^8 + y^8$, this theorem yields another expression for the *Nu* function of E_8

$$\nu_{E_8}(z) = \frac{1}{2}\frac{(1 + z^2)^8 + 256z^8}{(1 - z^2)^8} + \frac{1}{2}\frac{(1 - z^2)^8}{(1 + z^2)^8} \\ = 1 + 128z^4 + 2944z^8 + 1024z^{10} + O(z^{12}).$$

TABLE III
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *SPHERES* OF ENERGY m FOR THE PLANE CUBIC LATTICE $\mathbf{Z}^2.[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9
0	1	4	4	0	4	8	0	0	4	4
	1	5	9	9	13	21	21	21	25	29
1	8	0	0	8	0	0	4	8	4	0
	37	37	37	45	45	45	49	57	61	61
2	8	0	0	0	0	12	8	0	0	8
	69	69	69	69	69	81	89	89	89	97

TABLE IV
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *PYRAMIDS* OF ENERGY m FOR THE PLANE CUBIC LATTICE $\mathbf{Z}^2.[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9
0	1	4	8	12	16	20	24	28	32	36
	1	5	13	25	41	61	85	113	145	181
1	40	44	48	52	56	60	64	68	72	76
	221	265	313	365	421	481	545	613	685	761
2	80	84	88	92	96	100	104	108	112	116
	841	925	1013	1105	1201	1301	1405	1513	1625	1741

$\Rightarrow$ Barnes-Wall lattice. This lattice is obtained by construction B applied to the first-order Reed–Muller code $R(1,4)$ of length 16, with 32 codewords, and w.e.

$$W_{16}(x,y) = x^{16} + 30x^8y^8 + y^{16}.$$

By Theorem 2, we know that its *Nu* function is

$$\nu_{\Lambda_{16}}(z) = \frac{1}{2}\frac{W_{16}(1+z^2, 2z)}{(1-z^2)^{16}} + \frac{1}{2}\left(\frac{1-z^2}{1+z^2}\right)^{16}.$$

D. Comparison with Spherical Codebooks

Tables III to XVI were obtained by computing the Taylor expansion of the generating series of the preceding section in MAPLE. In these tables, the white lines indicate the number of lattice points *on* a shell of height m, and the shaded lines indicate the number of lattice points *within* a shell of height m.

In all examples, the pyramids of sufficiently large radius have many fewer lattice points than the spheres. In the case of cubic lattices, this can be paralleled with the fact that the volume of the n-sphere of radius m is $\frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}m^n$, whereas the volume of the n-pyramid of radius m is $\frac{2^n}{n!}m^n$, which is much smaller for large n.

As noticed earlier, the number of points $\mathbf{Z}^n$ in the pyramid of radius m is $[z^m]\frac{\nu_{z(c)}(z)}{1-z} \approx \frac{2^n m^n}{n!}$ for large m, whereas it is $\frac{\pi^{n/2}}{n/2!}m^n$ (n even) in the sphere of radius m. This result shows that there are increasingly more points in a sphere than in a pyramid when the dimension n grows.

Previous approximations can be used for high radii (m large) and, therefore, the number of points in a certain pyramid or sphere. This number is obtained dividing the volume of the pyramid or sphere by the volume of the considered Voronoi. However, this not possible for low radii like in our experiments. It is then necessary to use the theta series for spheres or the *Nu* series for pyramids.

TABLE V
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *SPHERES* OF ENERGY m FOR THE LATTICE $\mathbf{Z}^4.[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9
0	1	8	24	32	24	48	96	64	24	104
	1	9	33	65	89	137	233	297	321	425
1	144	96	96	112	192	192	24	144	312	160
	569	665	761	873	1065	1257	1281	1425	1737	1897
2	144	256	288	192	96	248	336	320	192	240
	2041	2297	2585	2777	2873	3121	3457	3777	3969	4209

TABLE VI
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *PYRAMIDS* OF ENERGY m FOR THE LATTICE $\mathbf{Z}^4.[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9
0	1	8	32	88	192	360	608	952	1408	1992
	1	9	41	129	321	681	1289	2241	3649	5641
1	2720	3608	4672	5928	7392	9080	11008	13192	15648	18392
	8361	11969	16641	22569	29961	39041	50049	63241	78889	97281
2	21440	24808	28512	32568	36992	41800	47008	52632	58688	65192
	118721	143529	172041	204609	241601	283401	330409	383041	441729	506921

TABLE VII
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *SPHERES* OF ENERGY m FOR THE LATTICE $D_4.[m = 10t + u]$. Even values only.

t\u	0	2	4	6	8	10	12	14	16	18
0	1	24	24	96	24	144	96	192	24	312
	1	25	49	145	169	313	409	601	625	937
2	144	288	96	336	192	576	24	432	312	480
	1081	1369	1465	1801	1993	2569	2593	3025	3937	3817
4	144	768	288	576	96	744	336	960	192	720
	3961	4729	5017	5593	5689	6433	6769	7729	7921	8641
6	576	768	24	1152	432	1152	312	912	480	1344
	9217	9985	10009	11161	11593	12745	13057	13969	14449	15793

Dimension 2

Plane Cubic Lattice $\mathbf{Z}^2$*:* See Tables III and IV.

Dimension 4

Cubic Lattice in Dimension 4 $\mathbf{Z}^4$*:* See Tables V and VI. Note that the larger number of spheres as compared with pyramids starts earlier ($m = 2$) than in the preceding example.

Lattice D_4*:* See Tables VII and VIII. This code is obtained by Construction A applied to the parity check of length 4: the code EW_4. Corollary 1 yields $[z^m]\frac{\nu_{Z(C)}(z)}{1-z} \approx \frac{2^4 m^4}{2!4!}$.

TABLE VIII
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *PYRAMIDS* OF ENERGY m FOR THE LATTICE D_4.$[m = 10t + u]$. Even values only.

t\u	0	2	4	6	8
0	1	32	192	608	1408
	1	33	225	833	2241
1	2720	4672	7392	11008	15648
	4961	9633	17025	28033	43681
2	21440	28512	36992	47008	58688
	65121	93633	130625	177633	236321
3	72160	87552	104992	124608	146528
	308481	396033	501025	625633	772161

TABLE IX
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *SPHERES* OF ENERGY m FOR THE LATTICE $\mathbf{Z}^8$.$[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9
0	1	16	112	448	1136	2016	3136	5504	9328	12112
	1	17	129	577	1713	3729	6865	12369	21697	33809

TABLE X
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *PYRAMIDS* OF ENERGY m FOR THE LATTICE $\mathbf{Z}^8$.$[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9
0	1	16	128	688	2816	9424	27008	68464	157184	332688
	1	17	145	833	3649	13073	40081	108545	265729	598417

TABLE XI
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *SPHERES* OF ENERGY m FOR THE LATTICE E_8.$[m = 10t + u]$.

t\u	0	2	4	6	8
0	1	240	2160	6720	17520
	1	241	2401	9121	26641
1	30240	60480	82560	140400	181680
	30481	117361	170625	310321	502001

Dimension 8

Cubic Lattice in Dimension 8 Z^8: See Tables IX and X.

Lattice E_8: See Tables XI and XII. The version of E_8 we implemented according to the definition of Section V-C-1 is in fact 0.5 times the one obtained by construction B applied to R_8. Hence, the new *Nu* function is obtained by replacing z by $\sqrt{z}$ in the expression obtained in the preceding section. For large m, Corollary 2 (with $|C| = 2$) yields

$$[z^m]\frac{\nu_{z(c)}(z)}{1-z} \approx \frac{2^8 m^8}{8!}.$$

TABLE XII
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *PYRAMIDS* OF ENERGY m FOR THE LATTICE E_8.$[m = 10t + u]$.

t\u	0	2	4	6	8	10	12	14	16
0	1	0	128	0	2944	1024	31616	15360	199424
	1	1	129	129	3073	4097	35713	19457	250497

TABLE XIII
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *SPHERES* OF ENERGY m FOR THE LATTICE $\mathbf{Z}^{16}$.$[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9
0	1	32	480	4480	29152	140736	525952	1580800	3994080	8945824
	1	33	513	4993	34145	174881	700833	2271633	6275713	15221537

TABLE XIV
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *PYRAMIDS* OF ENERGY m FOR THE LATTICE $\mathbf{Z}^{16}$.$[m = 10t + u]$.

t\u	0	1	2	3	4	5	6
0	1	32	512	5472	44032	285088	1549824
	1	33	545	6017	50049	335137	1884961

TABLE XV
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *SPHERES* OF ENERGY m FOR THE LATTICE Λ_{16}.$[m = 10t + u]$.

t\u	0	2	4	6	8	10	12	14	16	18
0	1	0	4320	61440	522720	2211840	8960640	23224320	67154400	135168000
	1	0	4321	65761	588481	2800321	11760961	34985281	102139681	237308161

TABLE XVI
NUMBER OF POINTS LYING ON (WHITE CELLS) OR WITHIN (GREY CELLS) *PYRAMIDS* OF ENERGY m FOR THE LATTICE Λ_{16}.$[m = 10t + u]$.

t\u	0	1	2	3	4	5	6	7	8	9	10
0	1	0	0	0	512	0	0	0	47872	0	92160
	1	1	1	1	513	513	513	513	48385	48385	140545

Dimension 16

Cubic Lattice in Dimension 16 $\mathbf{Z}^{16}$: See Tables XIII and XIV.

Barnes-Wall Lattice in Dimension 16 Λ_{16}: See Tables XV and XVI. For large m, Corollary 2 yields $[z^m]\frac{\nu(z)}{1-z} \approx 7.6\ 10^{-3} m^3$.

E. Conclusion

From numerical and analytical comparisons with the spherical case, it transpires that pyramids are far less crowded than spheres for large values of the radii, which is a fact that is consistent with the thumbtack aspect of the Laplacian distribution as opposed to the bell-shaped Gaussian. This results in smaller codebooks in the pyramid scheme for equivalent radii.

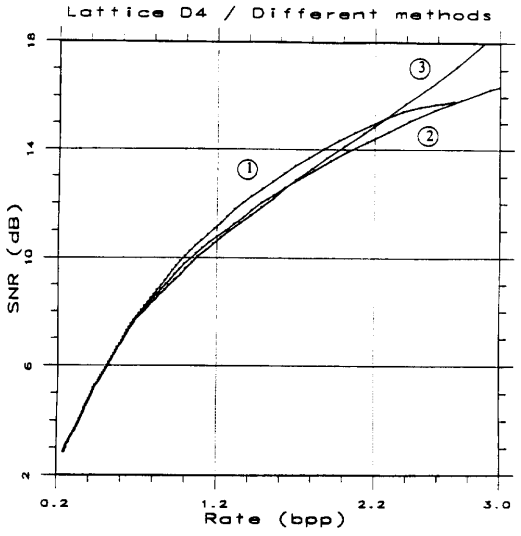


Fig. 7. Comparison between the three encoding schemes: 1) First scheme; 2) second scheme: encoding of scale factor γ; 3) third scheme: encoding each component of the vectors.

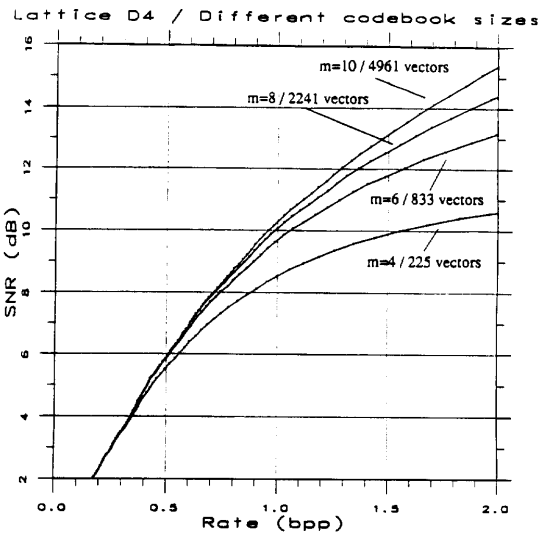


Fig. 8. Accuracy of lattice vector quantization versus codebook size. (For each radius m, the associated number of vectors is given.)

VI. Labeling Lattice Points

The Conway–Sloane algorithm [13] for labeling consists of associating with every point of the lattice Λ comprised in the Voronoi cell of $k\Lambda$(k integer) the index vector of its coefficients on the basis defining the lattice. It can be shown that there are exactly k^n such vectors and that these coefficients lie in the range $[-k, k]$. In particular, if needed, the index vector can be represented as the expansion to the base k of an integer (in the range $[0, \ldots, k^n - 1]$), which can be thought of as a scalar label.

In our scheme, k is chosen as the smallest integer large enough so that the Voronoi cell of $k\Lambda$ contains the dictionary. In the case of a L^1 sphere of radius r, one can take $k = \lceil r/\rho \rceil$, where ρ is the L^2 packing radius of the lattice.

The restitution process involves as a subroutine the fast quantizing algorithms of [13]. The encoding process involves using the inverse of the generator matrix for the lattice, which is well known for D_4, E_8, Λ_{16} etc. ... (no computation).

$$Y \underset{X = KG - kQ_\Lambda\left(\frac{KG}{k}\right)}{\overset{\text{index}(Y) = (k_1, \ldots, k_n) = \text{mod}_k\{YG^{-1}\}}{\longrightarrow}} K = \text{index}(Y)$$

With G the generator matrix of Λ

and Q_Λ is defined in Section IV-B.

Note that we need not use the full Voronoi cell of $k\Lambda$ (as in [13]). Any dictionary shape strictly contained in that cell can be addressed by this process.

We can also cite an other method to associate at each codebook vector an index. Lamblin, in her algorithm, determines a scalar index for each codebook vector. This index is the union of a sub-index determining the energy m of the vector and a sub-index giving its position in the class of vectors with energy m (see [42]).

Alternative labeling algorithms can be found in, e.g., Laroia and Farvardin [44], Lamblin [42], and our recent works [45].

VII. Experimental Results

In this section, we present both numerical and qualitative coding results. The simulations are made using the popular 512 by 512 black and white Lena and Barbara images. The intensity of each pixel is coded on 256 grey levels (8 bpp). The quantization scheme is based on lattice vector quantization of wavelet coefficients. Note that previous experiments showed that Laplacian pdf are more suitable for modeling the pdf of these wavelet coefficients [18], [25].

A. Lattice Coding Analysis

We perform our lattice coding evaluation on a given wavelet coefficients subimage (scale 2). Notice that these coefficients are well fitted by a Laplacian pdf. Hence, all the following experiments are based on pyramid encoding. Entropy is defined by $H(Y) = -\sum_{i=1}^{L} p(Y_i) \log_2 p(Y_i)$, where $p(Y_i)$ is the probability of using the ith vector of the codebook.

Comparison Between Three Encoding Schemes: First, we compare for a given lattice D_4 the three schemes described in Section IV-C for encoding vectors lying outside the codebook. Fig. 7 shows that scheme 2 ensures the best tradeoff between SNR and rate. In fact, scheme 1 privileges entropy, whereas scheme 3 privileges distortion.

Accuracy of Encoding Scheme versus Codebook Size: For lattice D_4, Fig. 8 plots SNR versus entropy. As is well known, increasing the size of the codebook results in better

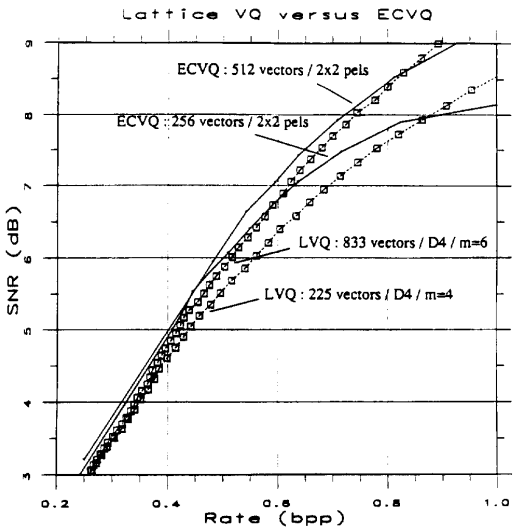


Fig. 9. Comparison between lattice VQ and ECVQ. Comparison between two sizes of codebooks: ECVQ 256 vectors and LVQ 225 vectors; ECVQ 512 vectors and LVQ 833 vectors.

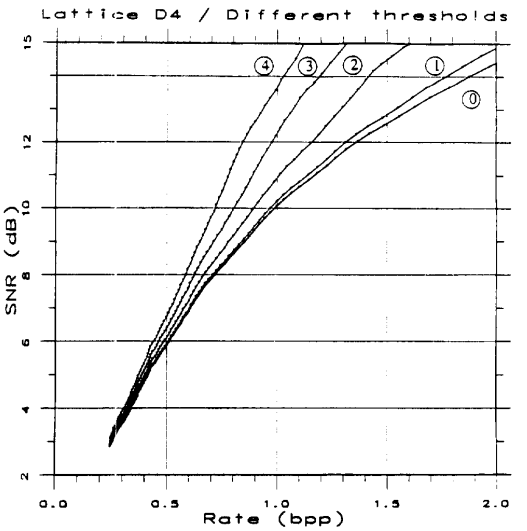


Fig. 10. Accuracy according threshold: 0) Without thresholding; 1) thresholds ± 1; 2) thresholds ± 2; 3) thresholds ± 3; 4) thresholds ± 4.

coding. It will be noticed that large codebooks can be used since no quadratic norm computation is required.

Slightly better SNR can be obtained using ECVQ for the same codebook size (256) (see Fig. 9). However, entropy lattice VQ outperforms ECVQ since in the lattice case, a large codebook can be easily implemented. Notice that in the lattice case, the asymptotic distortion is obtained for an achievable codebook size.

Accuracy According to a Threshold: It is well known that thresholding low-energy wavelet coefficients reduces the bit rate while image quality is preserved. Fig. 10 shows that thresholding improves entropy without decreasing SNR. Better thresholding methods based on Markovian modeling improve the tradeoff between image quality and bit rate [5].

Asymptotic Comparison Between the Best Usual Lattices: Fig. 11 shows comparison of optimal lattices at low bit rate. Notice that E_8 outperforms D_4 over 0.8 dB at low bit rate. However, this higher dimensional lattice requires a greater codebook size. We think that this will not be a problem since few quadratic norm computations are required for lattice encoding.

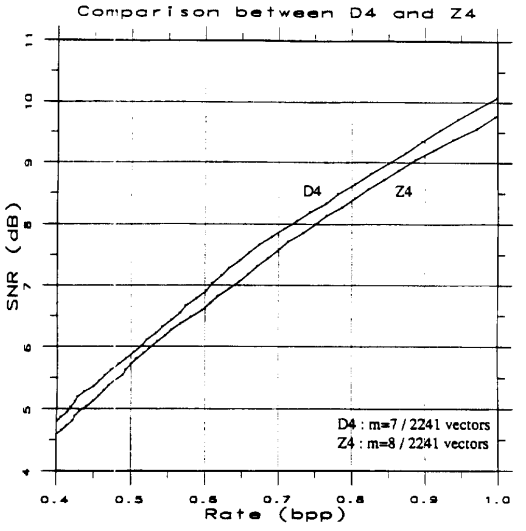


Fig. 11. Optimal lattice at low bit rate.

B. Image Coding Results

The numerical evaluation of the coder's performances is usually achieved by computing the peak signal-to-noise ratio (PicSNR) between the original image and the coded image. For each coded sub-image, we use a variable length code. Here, we give the bit rate R if an optimal entropy coding is performed corresponding to the lower bound bit rate.

Lena Image: The vector lengths and the bit rate are defined by a bit allocation procedure developed in [3], [7], and [43]. Fig. 12 shows the original image, and Fig. 13 shows coding results using a quincunx pyramid (four levels) and a lattice VQ at a bit rate $R = 0.174$ bpp (compression ratio of 40:1). Better results (0.167 bpp) (Fig. 14) are obtained using wavelet transform, first Markovian modeling [5], and lattice VQ.

Barbara Image: Fig. 15 shows the original image and Fig. 16 the coding results using a quincunx pyramid (four levels) and a lattice VQ at a bit rate $R = 0.2$ bpp (compression ratio of 40:1). The bit allocation algorithm is detailed in [43].

Comparison with other coders is a difficult task since, in the literature, authors used either sub-band coding or LVQ schemes.

Fig. 12. Original image LENA (8 bpp).

Fig. 13. Image LENA coded at 0.174 bpp (compression ratio 45.8:1) PicSNR = 30.3 dB.

Fig. 14. Image LENA coded at 0.167 bpp (compression ratio 47.9:1) PicSNR = 30.2 dB.

The main advantage of this new method is that few computational burdens ensue compared with classical LBG VQ methods. Furthermore, image coding results do not depend on the training sequence.

Fig. 15. Original image Barbara (8 bpp).

Fig. 16. Image Barbara coded at 0.2 bpp (compression ratio 40:1) PicSNR = 26.0 dB.

VIII. Conclusion

We have proposed a new scheme for image coding. The method is based on quincunx wavelet transform and lattice vector quantization.

Compared with the results obtained with a quincunx pyramid, the dyadic wavelet transform gives more ringing artifacts, and the separability of the filters seems to cause a staircase on the edges. We have proposed a new filter design based on wavelet analysis.

Second, we propose a new pyramidal scheme preserving the sharpness of the edges and avoiding smoothing artifacts. Wavelet coefficients are quantized using lattice codebooks with pyramidal boundaries. Furthermore, this pyramidal lattice vector quantization involves a fast encoding algorithm.

We present a new method to compute lattice points in a pyramid for nonorthogonal lattices such as D_n, E_8, Λ_{16}. Finally, the use of both optimal wavelet transform and optimal lattice with a new encoding scheme provides good coding results at low bit rates.

Acknowledgment

The authors wish to thank R. A. Calderbank, I. Daubechies, and N. J. A. Sloane for helpful discussions.

REFERENCES

[1] E. H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transform for image coding," *SPIE Visual Commun. Image Processing*, vol. 845, pp. 50–58, 1987.

[2] J. P. Adoul and M. Barth, "Nearest neighbor algorithm for spherical codes from the Leech lattice," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1188–1202, 1988.

[3] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using vector quantization in the wavelet transform domain," in *Proc. IEEE ICASSP* (Albuquerque, NM), Apr. 1990, pp. 2297–2300.

[4] M. Barlaud, P. Solé, M. Antonini, and P. Mathieu, "A pyramidal scheme for lattice vector quantization of wavelet transform coefficients applied to image coding," in *Proc. IEEE ICASSP* (San Francisco, CA), Mar. 23–26, 1992.

[5] M. Barlaud, L. Blanc-Féraud, and P. Charbonnier, "New image coding scheme using multiresolution markov random fields," in *Proc. SPIE/IS&T Conf.* (San Jose, CA), Feb. 9–14, 1992.

[6] M. Antonini, M. Barlaud, P. Mathieu, and P. Solé, "Entropy constrained lattice vector quantization for image coding using wavelet transform," in *Proc. ESA Workshop* (Noordwijk, The Netherlands), June 1991.

[7] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, no. 2, Apr. 1992.

[8] M. Antonini, M. Barlaud, and P. Mathieu, "Image coding using lattice vector quantization of wavelet coefficients," in *Proc. IEEE ICASSP* (Toronto, Canada), May 14–17, 1991, pp. 2273–2277.

[9] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," AT&T Bell Labs. Tech. Rep., No. TM 11217-900529-07.

[10] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy constrained vector quantization," *IEEE Trans. Acoust Speech Signal Processing*, vol. 37, no. 1, pp. 31–42, 1989.

[11] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer Verlag, 1988.

[12] ——, "Fast quantizing and decoding algorithms for lattice quantizers and codes," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 227–232, Mar. 1982.

[13] ——, "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 6, pp. 820–824, 1983.

[14] Y. Meyer, *Ondelettes et Operateurs*. Hermann, 1990

[15] T. Senoo and B. Girod, "Vector quantization for entropy coding of image subbands," *IEEE Trans. Image Processing*, vol. 1, no. 4, pp. 526–533, Oct. 1992.

[16] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.* vol. 41, pp. 909–996, 1988.

[17] J. C. Feauveau, "Analyse Multirésolution pour les Images avec un Facteur de Résolution $\sqrt{2}$," *Traitement du signal*, vol. 7, no. 2, pp. 117–128, 1990.

[18] T. R. Fischer, "Entropy constrained geometric vector quantization for transform image coding," in *Proc. IEEE ICASSP* (Toronto, Canada), May 14–17, 1991, pp. 2269–2271.

[19] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 568–583, 1986.

[20] P. Flajolet and A. Odlyzko, "Singularity analysis of generating functions," *Rapport de Recherche INRIA 826*, Apr. 1988.

[21] J. Kovacevic and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathbf{R}^n$," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 533–555, Mar. 1992.

[22] H. Tseng and T. R. Fischer, "Transform and hybrid transform /DPCM coding of images using pyramid vector quantization," *IEEE Trans. Commun.*, vol. COM-35, pp. 79–86, 1987.

[23] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 4, July 1979.

[24] D. G. Jeong and J. D. Gibson, "Lattice vector quantization for image coding," in *Proc. IEEE ICASSP*, 1989, pp. 1743–1746.

[25] ——, "Image coding with uniform and piecewise uniform quantizers," Preprint 1992.

[26] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, no. 1, pp. 84–95, Jan. 1980.

[27] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 7, July 1989.

[28] J. E. Mazo and A. M. Odlyzko, "Lattice points in high-dimensional spheres," *Mh. Math. 110*, pp. 47–61, 1990.

[29] N. M. Nasrabadi and R. A. King, "Image coding using vector quantization: A review," *IEEE Trans. Commun.*, vol. COM-36, no. 8, Aug. 1988.

[30] N. J. A. Sloane, letter to P. Solé, Aug. 1, 1991.

[31] P. Zador, "Asymptotic quantization error of continuous signals and their quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, 1982.

[32] J. Woods and O. Neil, "Subband coding of images," *IEEE Trans. Acoust Speech Signal Processing*, vol. ASSP-34, pp. 1278–1288, Oct. 1986.

[33] R. Rao and W. Pearlman, "Multirate vector quantization of image pyramids," in *Proc. IEEE ICASSP* (Toronto, Canada), May 14–17, 1991, pp. 2257–60.

[34] P. Solé, "Counting lattice points in pyramids," *Actes de congres Séries Formelles et Combinatoire Algébrique Montréal*, Publication du LACIM, no. 11, pp 343–355 June 1992.

[35] J. D. Gibson and K. Sayood, "Lattice quantization," *Adv. Electron. Electron. Phys.*, vol. 72, pp. 259–330, 1988.

[36] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer, 1992.

[37] R. Ansari and C. Lau, "Two-D IIR filters for exact reconstruction in a tree-structured subband decomposition," *Electron. Lett.*, 1987.

[38] A. E. Cetir, "A multiresolution non-rectangular wavelet representation for 2-D signals," in *Proc. 1990 Bilkent Conf. New Trends Commun. Signal Processing* Ed. (E. Arikan, Ed.). Elsevier, 1990.

[39] C. Guillemot, A. E. Cetin, and R. Ansari, "M-channel nonrectangular wavelet representation for 2-D signals," in *Proc. IEEE ICASSP* (Toronto, Canada), May 14–17, 1991.

[40] R. Ansari and C. Guillemot, "Exact reconstruction filter banks using diamond FIR filters," in*Proc. 1990 Bilkent Conf. New Trends Commun. Signal Processing* (E. Arikan, Ed.). Elsevier, 1990.

[41] Tanabe and Favardin, "Subband image coding using entropy—Coded quantization over noisy channels," *IEEE J. Sel. Areas Commun.*, June 1992.

[42] C. Lamblin, "Quantification Vectorielle Algébrique Sphérique par le réseau de Barnes-Wall: Application au Codage de Parole," *Thése de Doctorat*, Sherbrooke, Canada, Mar. 1988.

[43] M. Antonini, M. Barlaud, B. Rouge, and C. Lambert-Nebout, "Weighted optimum bit allocation for multiresolution satellite image coding," in *Proc. Quartorzieme colloque GRETSI* (Juan Les Pins, France), Sept. 13–16, 1993, pp. 455–458.

[44] R. Laroia and N. Farvardin, "A structured fixed rate vector quantizer deriver from a variable length scalar quantizer: Parts I and II," *IEEE Trans. Inform. Theory*, vol. IT-39, no. 3, pp. 851–876, May 1993.

[45] J. M. Moureaux, M. Antonini, and M. Barlaud, "Fast encoding algorithm for lattice vector quantization," submitted to *IEEE Trans. Image Processing*, Mar. 1994.

Michel Barlaud (M'86) was born in France on November 24, 1945. He received the "Thèse d'état" degree from the University of Paris XII.

He is currently a Professor at the I3S CNRS Laboratory of the University of Nice-Sophia Antipolis. His research interests include image processing, wavelet analysis, Markov random fields, and adaptive deterministic relaxation for nonlinear optimization. His fields of application include image coding, image restoration, computed tomography, motion estimation, and stereo vision.

Patrick Solé (M'88) received the ingénieur and docteur ingénieur degrees from the Ecolé Nationale Supérieure des Télécommunications in 1984 and 1987, respectively.

From 1987 to 1989, he was a visiting assistant professor with the School of Computer and Information Science, Syracuse University, Syracuse, NY. Since 1989, he has been with the CNRS Laboratory I3S, Sophia Antipolis, France. His research interests include coding theory (covering radius, cyclic codes), vector quantization (lattices), and interconnection networks (graph spectra).

Thierry Gaidon was born in France on December 5, 1964. He received the DEA degree in signal processing in 1990 from the University of Nice-Sophia Antipolis, France. Since 1990, he has been working toward the Ph.D. degree at the I3S Laboratory at CNRS and the University of Nice-Sophia Antipolis.

His research interests include multidimensional image processing, wavelet analysis, and still image and image sequence coding.

Marc Antonini was born in France on August 29, 1965. He received the Ph.D. degree from the University of Nice-Sophia Antipolis, France, in 1991.

He is currently working with CNRS at the I3S Laboratory in Sophia Antipolis. His research interests include multidimensional image processing, wavelet analysis, and image coding.

Pierre Mathieu was born in Alger on May 10, 1956. He received the Ingenieur ENSEEIHT and Ph.D. degrees from the Institut National Polytechnique de Toulouse.

He is currently Maître de Conférences at the I3S Laboratory of CNRS and the University of Nice-Sophia Antipolis. His research interests include multidimensional image processing, wavelet analysis, image coding, and image restoration.

Annexe D

Lattice Codebook Enumeration for Generalized Gaussian Source

IEEE Transactions on Information Theory vol.49, no.2, pp. 521-528, février 2003

J'ai co-écrit cet article avec Pierre Loyer et Jean-Marie Moureaux.

Résumé Le but de cet article est de proposer un algorithme de dénombrement et d'indexage de faible complexité pour les vecteurs d'un réseau régulier, et efficace pour différentes distributions de sources, c'est-à-dire pour différentes métriques L_p $(0 < p \leq 2)$. Dans un cas particulier nous obtenons la formule de dénombrement introduite par Fischer pour le réseau $\mathbb{Z}^n$. L'algorithme proposé présente de nombreux avantages et principalement, en faisant le lien avec les séries *théta*, il est possible de réduire sa complexité au calcul de produits de convolution pour $p = 1$ et $p = 2$. Ceci permet une implantation facile de la méthode sur DSP. L'algorithme que nous avons proposé pour le réseau $\mathbb{Z}^n$ peut être généralisé à d'autres réseaux comme le D_n.

$$+ C \sum_{k=L}^{\infty} \frac{(4k-1)!!}{(4k)!!} \frac{1}{4k+1}$$
$$\sim C \sum_{k=L}^{\infty} \frac{(4k-1)!!}{(4k)!!} \frac{1}{4k+1}$$
$$\sim C \sum_{k=L}^{\infty} \frac{1}{k^{3/2}}, \tag{31}$$

where we have used Lemma 4.3. The result follows immediately from (13) and (31). □

ACKNOWLEDGMENT

The author would like to thank Dawei Zheng and Dongming Zhu for providing excellent research assistance. In addition, the author would like to thank two anonymous referees for helpful comments.

REFERENCES

[1] J. Beran, *Statistics for Long-Memory Processes*. London, U.K.: Chapman & Hall, 1994.
[2] R. T. Baillie, "Long memory processes and fractional integration in econometrics," *J. Econ.*, vol. 73, pp. 5–59, 1996.
[3] P. C. B. Phillips, "Econometric analysis of Fisher's equation," in *Cowles Foundation for Research in Economics*. New Haven, CT: Yale Univ. Press, 1998.
[4] P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, Mar. 1992.
[5] E. Masry, "The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 39, pp. 260–264, Jan. 1993.
[6] G. W. Wornell, "Wavelet-based representations for the $1/f$ family of fractal processes," *Proc. IEEE*, vol. 81, pp. 1428–1450, Oct. 1993.
[7] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[8] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 904–909, Mar. 1992.
[9] E. J. McCoy and A. T. Walden, "Wavelet analysis and synthesis of stationary long memory processes," *J. Comput. Graph. Statist.*, vol. 5, pp. 26–56, 1996.
[10] G. W. Wornell, *Signal Processing With Fractals: A Wavelet-Based Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[11] B. J. Whitcher, "Assessing nonstationary time series using wavelets," Ph.D. dissertation, Dept. Statist., Univ. Washington, Seattle, 1998.
[12] M. J. Jensen, "Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter," *J. Forecasting*, vol. 18, pp. 17–32, 1999.
[13] ——, "An approximate wavelet MLE of short and long memory parameters," *Studies of Nonlinear Dynamics and Econometrics*, vol. 3, pp. 239–253, 1999.
[14] ——, "An alternative maximum likelihood estimator of long memory processes using compactly supported wavelets," *J. Econ. Dyn. and Contr.*, vol. 24, pp. 361–387, 2000.
[15] ——, "Bayesian inference of long-memory dependence in volatility via wavelets," Dept. Economics, Univ. Missouri at Columbia, Tech. Rep., 2000.
[16] P. F. Craigmile, D. B. Percival, and P. Guttorp, "Wavelet-based parameter estimation for trend contaminated fractionally differenced processes," J. Time Series Anal., 2000, submitted for publication.
[17] ——, "The impact of wavelet coefficient correlations on fractionally differenced process estimation," Nat. Res. Ctr. for Statistics and the Environment, Univ. Washington, Seattle, Tech. Rep., 2000.
[18] T. Kawata, *Fourier Analysis in Probability Theory*. New York: Academic, 1972.
[19] M.-J. Lai, "On the digital filter associated with Daubechies' wavelets," *IEEE Trans. Signal Processing*, vol. 43, pp. 2203–2205, Sept. 1995.

Lattice Codebook Enumeration for Generalized Gaussian Source

Pierre Loyer, Jean-Marie Moureaux, and Marc Antonini

***Abstract*—The goal of this correspondence is to propose a low-complexity enumeration algorithm for lattice vectors, based on a geometrical interpretation and valid for different source distributions, i.e., for different L_p-norms in the range $0 < p \leq 2$. As a particular case, we obtain the Laplacian enumeration formula of Fischer. This point of view offers various advantages and particularly it enables one to make the link with the generalized *theta*-series and to reduce the algorithm to the calculation of a few convolutional products in the special cases $p = 1$ and $p = 2$. Using a dedicated digital signal processing (DSP) architecture, convolutional products are easy to implement and require few arithmetic operations. Our algorithm, developed for the $\boldsymbol{Z^n}$ lattice, can be generalized to other lattices like the $\boldsymbol{D_n}$.**

***Index Terms*—Convolutional product, generalized Gaussian, indexing, lattice, product code, quantization.**

I. INTRODUCTION

In the field of data compression, for years considerable attention has been given to the quantization of Laplacian- or Gaussian-distributed vectors by means of pyramidal or spherical lattice vector quantizer [8], [6], [5]. Such quantization is performed to an integer lattice lying on a pyramidal or spherical shell [6], [5]. A lattice Λ in $\boldsymbol{R}^n$ is composed by all integer combinations of a set of linearly independent vectors $\boldsymbol{a_i}$ (the basis of the lattice) such that

$$\Lambda = \{\vec{y} | \vec{y} = u_1 \boldsymbol{a_1} + u_2 \boldsymbol{a_2} + \cdots + u_n \boldsymbol{a_n}\} \tag{1}$$

where the u_i are integers.

The fundamental advantage of lattice quantization is that no codebook needs to be generated or stored and quantization is very fast because it does not depend on the number of codewords used [5]. Furthermore, encoding can be done using a prefix code [6], [15], which is well suited to the regular structure of a lattice. On the other hand, the operation which consists in translating vectors of signal values into binary words, hereafter called enumerating (or coding or indexing), remains difficult.

A lot of work has been done for enumerating or indexing Laplacian or Gaussian distributions, as for example in [7], [5], [14], [15], [2], [13], but the work was not adapted to generalized Gaussian sources with parameter different from 1 and 2.[1] Indeed, many important sources of data including wavelet coefficients of images can be well modeled by generalized Gaussian distributions with shape parameter p in the range $0 < p \leq 2$ [1], [16].

Manuscript received March 19, 2001; revised October 15, 2002. The material in this correspondence was presented in part at the International Symposium on Information Theory, Sorrento, Italy, June 2000 [12].

P. Loyer is with ALCATEL Space Industries, 06156 Cannes-La-Bocca Cedex, France (e-mail: Pierre.Loyer@space.alcatel.fr).

J.-M. Moureaux is with CRAN Laboratory, CNRS, and with the University Henri Poincaré, Nancy 1, 54506 Vandoeuvre-lès-Nancy Cedex, France (e-mail: moureaux@cran.uhp-nancy.fr).

M. Antonini is with I3S Laboratory, CNRS and the University of Nice-Sophia Antipolis, 06903 Sophia Antipolis, France (e-mail: am@i3s.unice.fr).

Communicated by P. A. Chou, Associate Editor for Source Coding.
Digital Object Identifier 10.1109/TIT.2002.807306

[1]Note that the work of [14] and [15] can be generalized to any $\boldsymbol{p}$. However, these methods remain costly since they require storage of some lattice vectors called *leaders*.

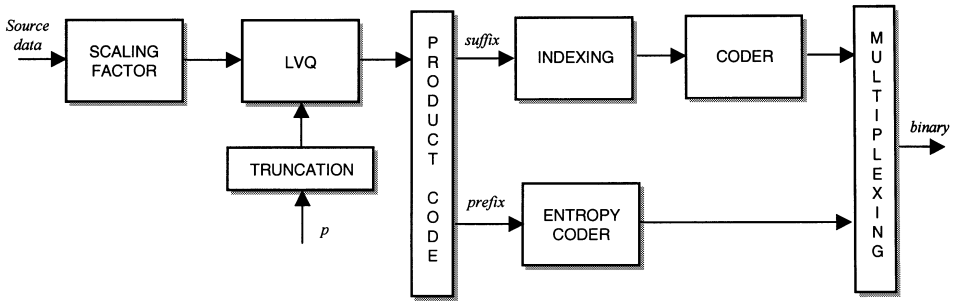


Fig. 1. General coding scheme.

Few works address the enumeration problem in the generalized Gaussian case. Let us cite, for instance, the papers of Laroia and Farvardin [10] and Chen and Villasenor [3]. Our main contribution here is to propose a lattice point enumeration algorithm for p in the range $0 < p \leq 2$ with a low complexity even for large codebook size; it is based on a geometrical interpretation and the use of *theta* series (and especially *nu*-series in the Laplacian case [17], [2]). This point of view offers various advantages and particularly it enables one to reduce the algorithm to the calculation of a few convolutional products in the cases $p = 1$ and $p = 2$. Using a dedicated digital signal processing (DSP) architecture, convolutional products are easy to implement and require few arithmetic operations. It is also important to note that our algorithm can be generalized to other lattices like the $\boldsymbol{D_n}$ lattice.

The correspondence is organized as follows. We begin by presenting in Section II the position of the problem and the general coding scheme on which is based the proposed method. Sections III and IV deal with the enumeration and coding algorithms, while Section V gives the decoding algorithm. An indexing example is provided in order to illustrate the proposed method, and an estimation of its complexity is given in Section VI. Finally, Section VII presents conclusions.

II. Statement of the Problem

A. Source Modeling

We consider vector quantization of a source where signal values have been assembled in blocks (or vectors) of size n. Coordinates of vectors are random processes which are assumed to be independent and identically distributed. Moreover, their distribution is assumed to be a generalized Gaussian, i.e., the probability density function (pdf) is of the kind $\alpha e^{-\beta \sum_{i=1}^{n} |x_i|^p}$. The most widespread values for p are $p = 2$ (Gaussian source) or $p = 1$ (Laplacian source, mainly in image coding). Values less than 1 have proven useful in some applications and their use motivates our study. In order to minimize the distortion, it is appropriate to consider codebooks constituted of vectors with integer coordinates distributed on surfaces of constant pdf

$$\sum_{i=1}^{n} |x_i|^p = R^p. \tag{2}$$

In mathematical words, our codebook is obtained by truncation of the cubic lattice $\boldsymbol{Z^n}$ by some sphere in the sense of the L_p-norm. Let us specify our vocabulary.

Definition: The L_p norm of the vector $\vec{x}$ is defined by

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}. \tag{3}$$

Strictly speaking, the word norm must be used if and only if $p \geq 1$. But, since it is not misleading in our context, we will use it also for $p < 1$. In the subspace of $\boldsymbol{R^n}$ spanned by the first k coordinates, we define the L_p-sphere radius as follows.

Definition: The L_p-sphere of radius R is

$$S_k(R) = \{\vec{x} / \|\vec{x}\|_p = R\}. \tag{4}$$

It is understood that the center of the sphere is the origin (the all 0 point). The number p is omitted for the sake of simplicity of notations. When the radius is 0, the L_p-sphere reduces to its center. The cardinality of S will be denoted by $\#S$. In particular, we have $\#S(0) = 1$.

B. General Coding Scheme

In Fig. 1 is presented the general coding scheme on which our indexing method is based. Source vectors are scaled inside a truncated lattice and then quantized using fast algorithms [4]. This process generates both granular and overload noises and the corresponding distortion is minimized during quantization by choosing an optimal scaling factor [9].

Then, the crucial problem in lattice encoding techniques is to represent any quantization symbol by a uniquely decodable binary word. To solve this problem, in our approach we use a *prefix code* such that a lattice vector $\vec{y}$ can be defined by an index pair (i, j) where

- i is the index of the radius R [see formula (2)] of $\vec{y}$, called *prefix* of $\vec{y}$;
- j is the index of the position of $\vec{y}$ on the shell of radius R, called *suffix* of $\vec{y}$.

The prefix code is well suited to the structure of a lattice [6]. The binary word will be constituted in two parts: a binary prefix code for the energy[2] and a binary suffix code for the position of the codevector on the shell corresponding to this energy [15]. The bit rate at the coder stage is then estimated by using this prefix code.

In this correspondence, we focus on the indexing of the *suffix* part of a lattice vector which is a difficult operation since we consider large codebooks. Indeed, for transmission or storage purposes, the *prefix* part, which represents the radius (or equivalently the energy) of the vector, can be easily indexed and encoded using a simple entropy coder. Furthermore, the maximum number of energy levels remains small for real-life sources.

Multiplexing the codes of the prefix and the suffix of a lattice vector before transmission or storage permits the decoder to recover the corresponding vector.

[2]In the remainder of the correspondence, the energy stands for the $\boldsymbol{L_p}$-norm at power $\boldsymbol{p}$, here $\boldsymbol{R^p}$.

III. Suffix Coding

A. Introduction

As we have seen in Section II-B, we focus in this correspondence on the suffix calculation at a given energy. The suffix, also called index in the remainder of the correspondence, is an integer ranging from 0 to the cardinality of the shell -1.

In the following, we present our approach for index calculation, the codevector energy being fixed. For the remainder of the section, one gives a lattice point $M(m_1, \ldots, m_n)$ such that $\|O\vec{M}\|_p = R$ to be indexed.

B. Generalized Theta Functions

Definition: The generalized *theta* function of the cubic lattice $\boldsymbol{Z}^n$ is the formal series θ_n, where the power of the variable takes every possible value of R^p and where the corresponding coefficient is the number of points on the sphere of radius R. Formally, we have

$$\theta_n(z) = \sum_R \#S_n(R) z^{R^p} \tag{5}$$

where $R^p = |r_1|^p + \cdots + |r_n|^p$, with $r_1, \ldots, r_n \in \boldsymbol{N}$.

Equivalently

$$\#S_n(R) = \left[z^{R^p}\right] \theta_n(z). \tag{6}$$

We omit p and $\boldsymbol{Z}$ in our notations for the sake of simplicity. As well, we will say "*theta* function" instead of "generalized *theta* function associated to $\boldsymbol{Z}^n$."

The theory of *theta* series was originally developed for $p = 2$ and is a whole chapter of mathematics. It is closely related to the theory of lattices and it is a useful tool in some communication applications when lattice points have to be counted. This is explained at length in the standard book by Conway and Sloane [5]. For example, for $p = 2$, we have

$$\theta_1(z) = 1 + 2z + 2z^4 + 2z^9 + \cdots \tag{7}$$

$$\theta_2(z) = \theta_1^2(z) = 1 + 4z + 4z^2 + 4z^4 + 8z^5 + \cdots \tag{8}$$

$$\theta_3(z) = \theta_1(z)\theta_2(z) = 1 + 12z^2 + 8z^3 + 14z^4 + \cdots. \tag{9}$$

In the case of Laplacian distributed sources ($p = 1$), an appropriate theory was developed, the theory of *nu* functions [17], [2]. It has been often used in image coding. Additional results were gathered by the first author in his Ph.D. dissertation [11]. In [17], Solé also mentioned possible extensions of his theory to other values of p.

Since coordinates are independent in $\boldsymbol{Z^n}$, we have the following property.

Property:

$$\theta_{n+1}(z) = \theta_n(z)\theta_1(z). \tag{10}$$

This recursive formula enables one to derive all *theta* functions from $\theta_1(z)$. In the Laplacian or the Gaussian case, $p = 1$ or 2, exponents of the *theta* series are integer values so that the following corollary can be derived.

Corollary: The sequence of coefficients of $\theta_{n+1}(z)$ is the convolutional product of the sequences of coefficients of $\theta_1(z)$ and $\theta_n(z)$.

This corollary will be essential in our algorithm since it enables us to reduce it to a few convolutional products.

C. Principle of the Method

Points (vectors) are ordered in the lexicographical order.

Definition: A point $M(m_1, \ldots, m_n)$ is said to be before (or "lower than") a point $M'(m_1', \ldots, m_n')$ if

- $m_1 < m_1'$ or
- ($m_1 = m_1'$ and $m_2 < m_2'$) or
- ($m_1 = m_1'$ and $m_2 = m_2'$ and $m_3 < m_3'$), etc.

The number assigned to M will be the number of integer points preceding it on $S_n(R)$. The very first point on this sphere is assigned the number 0.

D. Geometrical Interpretation

The notations are the same as above. We want to count points before M on $S_n(R)$. We remark that any hyperplane perpendicular to the first axis $x_1 = x$, $|x| \leq R$ cuts $S_n(R)$ in a sphere (in a space) of dimension $n-1$ and of radius $(R^p - |x|^p)^{1/p}$, centered at (0^{n-1}). Therefore

1) for points such that $x_1 = m_1$, we are brought to the initial problem: counting points before $M_2(m_2, \ldots, m_n)$ in dimension $n-1$, M_2 being located on $S_{n-1}((R^p - |m_1|^p)^{1/p})$;
2) for points such that $x_1 < m_1$, we have to add up the number of points on all the layers below M that is to compute

$$\sum_{x<m_1} \#S_{n-1}\left(\sqrt[p]{R^p - |x|^p}\right)$$

the $\#S_{n-1}(\sqrt[p]{R^p - |x|^p})$ being the coefficients of $\theta_{n-1}(z)$. Points before become points below in the geometrical language.

Note that this generalizes and provides a geometrical interpretation of the well-known algorithm by Fisher and Pan (see [7]) in the particular case $p = 1$. Indeed, we have

$$\#S_n(R) = \sum_{x_1=-\lfloor R \rfloor}^{\lfloor R \rfloor} \#S_{n-1}\left(\sqrt[p]{R^p - |x_1|^p}\right). \tag{11}$$

When $p = 1$, this formula becomes

$$\#S_n(R) = \#S_{n-1}(R) + 2\left(\#S_{n-1}(R-1) + \cdots + \#S_{n-1}(1) + 1\right). \tag{12}$$

We obtain the equation from which Fisher and Pan deduced their algorithm. Note that their equation is a bit more complicated due to the weights associated to the different axes.

An example of geometrical interpretation is given in Fig. 2.

E. Algorithm

Let M_k denote $(m_k, \ldots, m_n)$, and call $\mathrm{Index}(M_k)$ the function computing the index of M_k, $\mathrm{Below}(M_k)$ the function computing the number of points such that $x_k < m_k$. According to the preceding paragraph, we have

$$\mathrm{Index}(M_1) = \mathrm{Index}(M_2) + \mathrm{Below}(M_1) \tag{13}$$

then by recurrence

$$\mathrm{Index}(M_1) = \mathrm{Index}(M_n) + \sum_{k=1}^{n-1} \mathrm{Below}(M_k). \tag{14}$$

This is the key formula for our algorithm. The algorithm holds for any value of p considered. In the Laplacian or the Gaussian case, it can be further developed thanks to our corollary.

Let us first make a few remarks of practical interest.

- Actually, the $\mathrm{Below}(M_k)$ will be added up with k decreasing from $n-1$ to 1. So that the "dimension" of the *theta* series will increase and that, at each step, we can obtain the new *theta* series by multiplying the last one by θ_1.

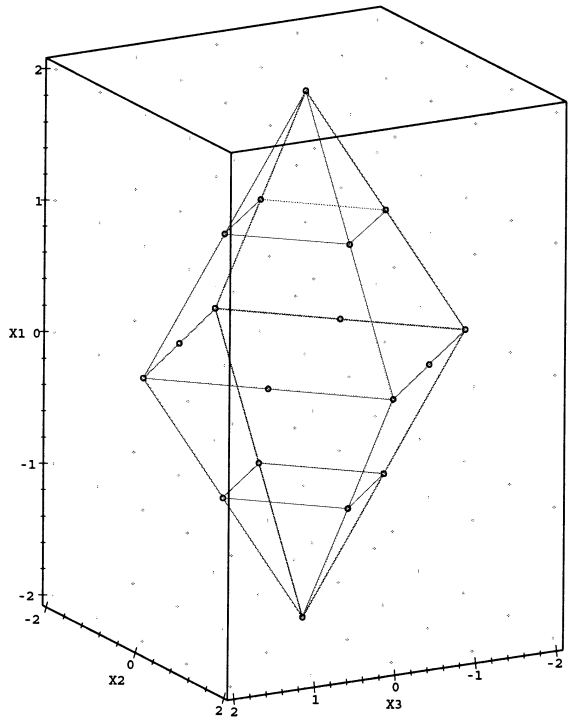


Fig. 2. Geometrical interpretation in the case of $S_3(R)$: L_1-sphere with $R = 2$.

- In dimension 1, spheres are reduced to two points at most. If $m_n < 0$, $\mathrm{Index}(M_n) = 0$ since M_n is the first point in the sphere, otherwise $\mathrm{Index}(M_n) = 1$.
- The computation of $\mathrm{Below}(M_k)$ requires the computation of the coefficients of $\theta_{n-k}(z)$ in a certain range. This range is determined by R which is an upper bound for every radius encountered throughout the algorithm. Thus, an upper bound for the maximum index and the actual maximum index in our algorithm is, respectively, R in the Laplacian case or R^2 in the Gaussian case (which are integers).

Algorithms Index and Below are developed in the Appendix. In next section, we give an example in the case $p = \frac{1}{2}$.

IV. Coding Example: Generalized Gaussian Case

Let $p = 1/2$ and let $M(1,\ 2,\ -3,\ 7)$ the point to be indexed. Note that, as p is a noninteger, it is no more possible to use convolutional products to compute the coefficients of the corresponding *theta* series. However, formulas (5) and (10) remain true. Thus, here we set

$$\theta_1 = 1 + 2z + 2z^{\sqrt{2}} + 2z^{\sqrt{3}} + 2z^2 + 2z^{\sqrt{5}} + 2z^{\sqrt{6}} + 2z^{\sqrt{7}}.$$

Then, the series θ_2 and θ_3 can be easily expanded, for example, with Matlab, using formula (10) and the coefficients can be stored in appropriate tables. The algorithm runs as follows:

• Initial step:
$\mathrm{Index} \leftarrow 0$.
• $k = 4$:
Since $x_4 = 7 > 0$, $\mathrm{Index} \leftarrow \mathrm{Index} + 1$.
• $k = 3$:

TABLE I
Coefficients of Theta Series Associated to $\mathbf{Z}^1$, Used for the Example of Section IV

x	-9	-8	-7	-6	-5	-4	-3
$\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}$	...	...	$\sqrt{3}$	...	...	...	$\sqrt{7}$
$[z^{\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}}]\theta_1(z)$	0	0	2	0	0	0	2

We consider $M_3(-3,\ 7) \in S_2((\sqrt{3}+\sqrt{7})^2)$ and we count points on layers below.

$$\mathrm{Below}(M_3) \leftarrow \sum_{x=-\lfloor(\sqrt{3}+\sqrt{7})\rfloor^2}^{-4} \#S_1((\sqrt{3}+\sqrt{7}-\sqrt{|x|}\,)^2).$$

The relevant parameters are collected in Table I (for sake of simplicity, we have omitted the radii of the shells where there are no integer points).

$\mathrm{Index} \leftarrow \mathrm{Index} + 2$.
• $k = 2$:

We consider $M_2(2,\ -3,\ 7) \in S_3((\sqrt{2}+\sqrt{3}+\sqrt{7}\,)^2)$ and we count points on layers below

$$\mathrm{Below}(M_2) \leftarrow \sum_{x=-\lfloor(\sqrt{2}+\sqrt{3}+\sqrt{7}\,)^2\rfloor}^{1} \#S_2((\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{|x|}\,)^2).$$

The relevant parameters are collected in Table II.

$\mathrm{Index} \leftarrow \mathrm{Index} + 16$.
• $k = 1$:

We consider

$$M_1(1,\ 2,\ -3,\ 7) \in S_4((1+\sqrt{2}+\sqrt{3}+\sqrt{7}\,)^2)$$

and we count points on layers below.

$$\mathrm{Below}(M_1) \leftarrow \sum_{x=-\lfloor(1+\sqrt{2}+\sqrt{3}+\sqrt{7}\,)^2\rfloor}^{0} \#S_3(1+\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{|x|}\,)^2).$$

The relevant parameters are collected in Table III.

$\mathrm{Index} \leftarrow \mathrm{Index} + 96$.

Finally, the point $M(1,\ 2,\ -3,\ 7)$ is indexed by the number 115.

V. Decoding

A. Principle

Decoding is the reverse process of encoding, mainly described in Section III and Algorithm 1 in the Appendix. Decoding must be obviously unique, that is, from any index we have to retrieve the unique corresponding vector. Recall that the coding technique used here generates codewords in two parts: a prefix (for the energy of the vector) and a suffix (for the position of the vector on the layer corresponding to this energy). The radius R can be straightforwardly deduced from the prefix. Furthermore, using our notations, the suffix corresponds to $\mathrm{Index}(M)$. Thus, the decoding process consists in retrieving the coordinates of an n-dimensional vector M from n, R, and $\mathrm{Index}(M)$. The general algorithm for decoding corresponds to Algorithm 4 in the Appendix. It is based on the same principle as the coding, that is, counting points on parallel layers with variable radius and dimension: $\#S_{k-1}(\sqrt[p]{R_k^p - |x|^p}\,)$, where p is also known. The cardinalities of these layers are computed through Algorithms 3 and 5 using appropriate *theta* series. But, in order to decrease the computational complexity they can be also precomputed and stored in tables.

TABLE II
COEFFICIENTS OF THETA SERIES ASSOCIATED TO $\mathbf{Z}^2$, USED FOR THE EXAMPLE OF SECTION IV

x	-33	...	-3	-2	-1	0	1	2
$\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}$	...	...	$\sqrt{2}+\sqrt{7}$	$\sqrt{3}+\sqrt{7}$	...	$\sqrt{2}+\sqrt{3}+\sqrt{7}$	...	$\sqrt{3}+\sqrt{7}$
$[z^{\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}}]\theta_2(z)$	0	0	8	8	0	0	...	8

TABLE III
COEFFICIENTS OF THETA SERIES ASSOCIATED TO $\mathbf{Z}^3$, USED FOR THE EXAMPLE OF SECTION IV

x	-46	...	-7	...	-3	-2	-1	0	1
$1+\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}$	...	...	$1+\sqrt{2}+\sqrt{3}$	...	$1+\sqrt{2}+\sqrt{7}$	$1+\sqrt{3}+\sqrt{7}$	$\sqrt{2}+\sqrt{3}+\sqrt{7}$	...	$\sqrt{2}+\sqrt{3}+\sqrt{7}$
$[z^{1+\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}}]\ \theta_3(z)$	0	0	24	0	24	24	24	0	24

In the following subsection, we give an example of decoding corresponding to the one given in Section IV for coding.

B. Example: Generalized Gaussian Case

As we saw in Section IV, the convolutional products cannot be used in this case straightforwardly. A good alternative is then to compute a table with the required coefficients of the *theta* series as explained in Section IV. Here, we give an example of decoding.

Let $p = 1/2$ and let $\mathrm{Index}(M) = 115$.

$$\|\vec{OM}\|_p = R = (1+\sqrt{2}+\sqrt{3}+\sqrt{7})^2$$

is also known (since it is transmitted as the prefix of the index). We thus have to retrieve the point $M(1,\ 2,\ -3,\ 7)$. The algorithm runs as follows:

- Initial step:

$\mathrm{Index} \leftarrow 115$.
$R_1 \leftarrow R = (1+\sqrt{2}+\sqrt{3}+\sqrt{7})^2$.

- $k = 1$:

Since $\|M\|_p = R_1 = (1+\sqrt{2}+\sqrt{3}+\sqrt{7})^2$, the minimum possible value of x_1 is $x_1 = -\lfloor R_1 \rfloor = -\lfloor (1+\sqrt{2}+\sqrt{3}+\sqrt{7})^2 \rfloor = -46$. We search the first value of x_1 such that

$$\sum_{x=-\lfloor (1+\sqrt{2}+\sqrt{3}+\sqrt{7})^2 \rfloor}^{x_1} \#S_3\left(\left(1+\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) > \mathrm{Index}.$$

Using the relevant parameters collected in Table III, we find for $x_1 = 1$

$$\sum_{x=-\lfloor (1+\sqrt{2}+\sqrt{3}+\sqrt{7})^2 \rfloor}^{x_1} \#S_3\left(\left(1+\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) = 120 > \mathrm{Index}.$$

Now we have $M_1(1,\ x_2,\ x_3,\ x_4)$. The number of points below M_1 is given by

$$\mathrm{Below}(M_1) = \sum_{x=-\lfloor (1+\sqrt{2}+\sqrt{3}+\sqrt{7})^2 \rfloor}^{0} \#S_3\left(\left(1+\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) = 96.$$

$\mathrm{Index} \leftarrow \mathrm{Index} - \mathrm{Below}\,(M_1) = 19$.
$R_2 \leftarrow (\sqrt{R_1} - \sqrt{\lvert x_1\rvert})^2 = (\sqrt{2}+\sqrt{3}+\sqrt{7})^2$.

- $k = 2$:

We consider $M_2(x_2,\ x_3,\ x_4) \in S_3((\sqrt{2}+\sqrt{3}+\sqrt{7})^2)$ and we count points on layers below until we have

$$\sum_{x=-\lfloor (\sqrt{2}+\sqrt{3}+\sqrt{7})^2 \rfloor}^{x_2} \#S_2\left(\left(\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) > \mathrm{Index}.$$

Using the relevant parameters collected in Table II, we find for $x_2 = 2$

$$\sum_{x=-\lfloor (\sqrt{2}+\sqrt{3}+\sqrt{7})^2 \rfloor}^{x_2} \#S_2\left(\left(\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) = 24 > \mathrm{Index}.$$

Now we have $M_2(2,\ x_2,\ x_4)$. The number of points below M_2 is given by

$$\mathrm{Below}(M_2) = \sum_{x=-\lfloor (\sqrt{2}+\sqrt{3}+\sqrt{7})^2 \rfloor}^{1} \#S_2\left(\left(\sqrt{2}+\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) = 16.$$

$\mathrm{Index} \leftarrow \mathrm{Index} - \mathrm{Below}\,(M_2) = 3$.
$R_3 \leftarrow (\sqrt{R_2} - \sqrt{\lvert x_2\rvert})^2 = (\sqrt{3}+\sqrt{7})^2$.

- $k = 3$:

We consider $M_3(x_3,\ x_4) \in S_2((\sqrt{3}+\sqrt{7})^2)$ and we count points on layers below until we have

$$\sum_{x=-\lfloor (\sqrt{3}+\sqrt{7})^2 \rfloor}^{x_3} \#S_1\left(\left(\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) > \mathrm{Index}.$$

Using the relevant parameters collected in Table I, we find for $x_3 = -3$

$$\sum_{x=-\lfloor (\sqrt{3}+\sqrt{7})^2 \rfloor}^{x_3} \#S_1\left(\left(\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) = 4 > \mathrm{Index}.$$

Now we have $M_3(-3,\ x_4)$. The number of points below M_3 is given by

$$\mathrm{Below}(M_3) = \sum_{x=-\lfloor (\sqrt{3}+\sqrt{7})^2 \rfloor}^{-2} \#S_1\left(\left(\sqrt{3}+\sqrt{7}-\sqrt{\lvert x\rvert}\right)^2\right) = 2.$$

$\mathrm{Index} \leftarrow \mathrm{Index} - \mathrm{Below}\,(M_3) = 1$.
$R_4 \leftarrow (\sqrt{R_3} - \sqrt{\lvert x_3\rvert})^2 = (\sqrt{7})^2 = 7$.

- $k = 4$:

Since it is the last coordinate, we have $\lvert x_4\rvert = 7$. Considering the rule of encoding: $\mathrm{Index} = 1$ yields a positive coordinate. Then, we have $x_4 = 7$.

Finally, we retrieve the point $M(1,\ 2,\ -3,\ 7)$ whose index was 115.

VI. COMPLEXITY

In this section, we analyze the complexity of both coding and decoding algorithms in terms of computational cost and memory requirement.

As we explained before we need to distinguish two cases:

- p in the range $0 < p \leq 2$: the coefficients of the *theta* series are precomputed and stored in tables to be used straightforwardly;

- $p = 1, 2$: the coefficients of the *theta* series can be computed on line using convolutional products (there is no storage requirement).

The complexity, i.e., the number of operations per sample is evaluated in terms of additions ($\boldsymbol{A}$), powers ($\boldsymbol{P}$), and convolutional products ($\boldsymbol{C}$) for the last case. Indeed, when the algorithms are implemented on DSPs including the convolutional product in their libraries, it is fair to count it as one operation. Otherwise, one should count at most $R^p + 1$ multiplications and R^p additions per product. Note that we neglected comparisons since they involve complexity in $\frac{1}{n}$ operations per sample.

Furthermore, as the complexity strongly depends on the coordinates of vectors, we propose to upper-bound it by taking the worst case where $x_k = \lfloor R \rfloor$ and all the other coordinates are null. Note that R means here the maximum radius of the codebook.

A. Computational Cost at Coding

1) p in the Range $0 < p \leq 2$: Here we consider Algorithms 1 and 2. The cost of Algorithm 1 is easily determined by

$$(2n-1)\boldsymbol{A} + (n-1) \text{ times Algorithm 2.} \tag{15}$$

The structure of Algorithm 2 is more complicated. First, we have to compute R_k, which costs $(n-k+2)\boldsymbol{P}$ and $(n-k)\boldsymbol{A}$ (k varying from $n-1$ to 1, according to Algorithm 1). This leads to

$$\left[\frac{1}{2}(n+1)(n+2)-3\right]\boldsymbol{P} \quad \text{and} \quad \left[\frac{1}{2}n(n-1)\right]\boldsymbol{A}.$$

Then, as the bounds of the loop depend on the coordinates of the current vector, we take the worst case. Thus, the loop runs at most $2\lfloor R \rfloor$ and performs $2\boldsymbol{A}$ and $1\boldsymbol{P}$ (R_k^p has been already computed).

Finally, the total computational complexity per sample at coding when *theta* series are precomputed can be evaluated by

$$C_p^{\text{coding}} \leq \left[\frac{1}{2}n + \left(4\lfloor R \rfloor + \frac{3}{2}\right) - \frac{1}{n}(1 + 4\lfloor R \rfloor)\right]\boldsymbol{A} + \left[\frac{1}{2}n + \left(2\lfloor R \rfloor + \frac{3}{2}\right) - \frac{2}{n}(1 + \lfloor R \rfloor)\right]\boldsymbol{P} \text{ per sample.} \tag{16}$$

2) p Equals 1 or 2: In this case, the coefficients of the *theta* series are computed on line using Algorithm 3 whose complexity is limited to one convolutional product (except the first time where it is called: it consists only in loading the coefficients of the *theta* series in dimension 1). Algorithm 3 being called $(n-1)$ times by Algorithm 1, the involved complexity is $(n-2)\boldsymbol{C}$.

Then, according to (16), the total computational complexity per sample at coding when the *theta* series are computed on line can be evaluated by

$$C_{p=1,2}^{\text{coding}} \leq \left[\frac{1}{2}n + \left(\lfloor R \rfloor + \frac{3}{2}\right) - \frac{1}{n}(1 + 4\lfloor R \rfloor)\right]\boldsymbol{A} + \left[\frac{1}{2}n + \left(2\lfloor R \rfloor + \frac{3}{2}\right) - \frac{2}{n}(1 + \lfloor R \rfloor)\right]\boldsymbol{P} + \frac{1}{n}(n-2)\boldsymbol{C} \text{ per sample.} \tag{17}$$

Let us recall that in this case no storage is required.

B. Computational Cost at Decoding

1) p in the Range $0 < p \leq 2$: Here we consider Algorithms 4 and 2. In Algorithm 4, one has to retrieve the vector M, coordinate by coordinate. We know the index of M and its radius R^n which have been both transmitted to the decoder. It is important to notice that the complexity of Algorithm 2 is lower when used in the decoding process than in the coding process. Indeed, the computation of R_k^p is done here in Algorithm 4 from the starting value R^n by subtracting the contribution of each new found coordinate x_k. Furthermore, the loops of both Algorithms 4 and 2 are assumed to run at most $2\lfloor R \rfloor + 1$ times. Then, the complexity of Algorithms 4 and 2 can be upper-bounded, respectively, by

$$[(2\lfloor R \rfloor + 5)n - (2\lfloor R \rfloor + 5)]\boldsymbol{A} + (2\lfloor R \rfloor + 1)(n-1) \text{ times algorithm } 2 + 3(n-1)\boldsymbol{P} \tag{18}$$

and

$$(2\lfloor R \rfloor + 1)(2\boldsymbol{A} + 1\boldsymbol{P}). \tag{19}$$

Finally, the total computational complexity per sample at decoding when *theta* series are precomputed can be evaluated by

$$C_p^{\text{decoding}} \leq \left[\left(8\lfloor R \rfloor^2 + 10\lfloor R \rfloor + 7\right) - \frac{1}{n}\left(8\lfloor R \rfloor^2 + 10\lfloor R \rfloor + 7\right)\right]\boldsymbol{A} + 4\left[\left(\lfloor R \rfloor^2 + \lfloor R \rfloor + 1\right) - \frac{1}{n}\left(\lfloor R \rfloor^2 + \lfloor R \rfloor + 1\right)\right]\boldsymbol{P} \text{ per sample.} \tag{20}$$

2) p Equals 1 or 2: To avoid storage requirement, the coefficients of the *theta* series can be computed on line using Algorithm 5 whose complexity is evaluated to $(n-k-1)$ convolutional products (except the last time where it is called: *theta* series in dimension 1). Algorithm 5 is called $(n-1)$ times (for $k = 1$ to $n-1$) by Algorithm 4, which involves $[\frac{1}{2}(n-1)(n-2)]\boldsymbol{C}$.

Then, according to (20), the total computational complexity per sample at decoding when the *theta* series are computed on line can be evaluated by

$$C_{p=1,2}^{decoding} \leq \left[\left(8\lfloor R \rfloor^2 + 10\lfloor R \rfloor + 7\right) - \frac{1}{n}\left(8\lfloor R \rfloor^2 + 10\lfloor R \rfloor + 7\right)\right]\boldsymbol{A} \cdot 4\left[\left(\lfloor R \rfloor^2 + \lfloor R \rfloor + 1\right) - \frac{1}{n}\left(\lfloor R \rfloor^2 + \lfloor R \rfloor + 1\right)\right]\boldsymbol{P} + \frac{1}{2}\left(n - 3 + \frac{2}{n}\right)\boldsymbol{C} \text{ per sample.} \tag{21}$$

C. Storage Requirement at Coding

In this subsection, we detail the memory cost due to the coefficients of the *theta* series when they are not computed on line.

For a lattice codebook with radius R and a norm p, we have to store the $R^p + 1$ first coefficients of each *theta* function. If we count 4-byte words per coefficient and $n-1$ *theta* series, the storage complexity is defined by

$$St_{p=1,2}^{coding} \leq 4(n-1)(R^p + 1) \text{ bytes.} \tag{22}$$

Note that this formula is available for $p = 1$ or 2. In this case, the cells of the array can be addressed straightforwardly.

For other noninteger values of p, one has to design an additional entry to the previous array for the powers of the coefficients. This yields a search in the table that we neglected as the powers can be sorted in increasing order.

The storage requirement in the case of noninteger values of p is thus easily determined, using (22), to be

$$St_p^{\text{coding}} \leq 8(n-1)(\lfloor R \rfloor + 1) \text{ bytes.} \tag{23}$$

D. Storage Requirement at Decoding

The storage complexity at decoding can be straightforwardly deduced from the one at coding. One has just to store an additional coefficient of the *theta* series, according to the condition of the loop *while* of Algorithm 4. Thus, the storage requirements for $p = 1, 2$ and any other noninteger value of p are given, respectively, by

$$St_{p=1,2}^{\text{decoding}} \leq 4(n-1)(R^p + 2) \text{ bytes} \tag{24}$$

$$St_p^{\text{decoding}} \leq 8(n-1)(\lfloor R \rfloor + 2) \text{ bytes.} \tag{25}$$

VII. Conclusion

In this correspondence, we propose an efficient method for enumerating and indexing lattice points lying on L_p-spheres, p in the range $0 < p \leq 2$. This is of most interest for product code lattice vector quantization of generalized Gaussian distribution sources, like wavelet coefficients. This method was developed for $\boldsymbol{Z}^n$ lattices and can be easily extended to other lattices like $\boldsymbol{D_n}$.

On one hand, the main advantage of our approach is that, in some cases ($p = 1$ and $p = 2$), the algorithm can be reduced to the calculation of a few convolutional products. It is thus well adapted to dedicated DSP architectures since convolutional products are easy to implement and require few arithmetic operations. On the other hand, our method also holds for any p in the range $0 < p \leq 2$.

Furthermore, the proposed indexing method can be easily integrated in a whole coding chain and permits an efficient solution for low bit rate compression.

Appendix
Coding and Decoding Algorithms

ALGORITHM 1 (CODING)

$\mathrm{Index} = \mathrm{Index}(M)$
% main function.
% M is supposed to be a global variable, known by every subroutine.
$\mathrm{Index} \leftarrow 0$
If $x_n > 0$, $\mathrm{Index} \leftarrow \mathrm{Index} + 1$
for $k = n - 1$ to 1 step -1
  $th = \mathrm{Theta}(n - k)$
  % th is the vector formed by the $R^p + 1$ first coefficients of $\theta_{k-1}(z)$.
  $\mathrm{Index} \leftarrow \mathrm{Index} + \mathrm{Below}\,(M_k)$
endfor
write Index

ALGORITHM 2

Function $\mathrm{Below} = \mathrm{Below}(M_k)$
$\mathrm{Below} \leftarrow 0$
for x from $-\lfloor R_k \rfloor$ to $x_k - 1$
  % R_k and x_k are implicitly known from M.
  $\mathrm{Below} \leftarrow \mathrm{Below} + th[R_k^p - |x|^p]$
  % recall that $[z^{R_k^p - |x|^p}]\theta_{k-1}(z) = \#S_{k-1}(\sqrt[p]{R_k^p - |x|^p})$.
endfor
return Below

ALGORITHM 3 (THETA FUNCTION FOR CODING: $p = 1$ or 2)

Function $\mathrm{Theta}k = \mathrm{Theta}(k)$
% returns a vector of $R^p + 1$ integers
Static Theta1, LastTheta
% Static means that the declared variables keep the same values from one call to the next.
if $k = 1$ then
  $Theta1 \leftarrow \cdots$ % strongly depends on p; for example, for $p = 1$
    $\mathrm{Theta1} = [1, 2, \ldots, 2]$
  $\mathrm{Theta}k \leftarrow \mathrm{Theta1}$
else
  $\mathrm{Theta}k \leftarrow \mathrm{conv}(\mathrm{LastTheta}, \mathrm{Theta1})$
  % conv denotes the convolutional product.
endif
$\mathrm{LastTheta} \leftarrow \mathrm{Theta}k$
return $\mathrm{Theta}k$

ALGORITHM 4 (DECODING)

$M = M(x_1, x_2, \ldots, x_n)$
% n, p, R, $\mathrm{Index}(M)$ are given.
$I \leftarrow \mathrm{Index}(M)$
$R_1 \leftarrow R$
% loop on the coordinates:
for $k = 1$ to $n - 1$, step 1
  $th = \mathrm{Theta}(n - k)$
  % th is the vector formed by the $R^p + 1$ first coefficients of $\theta_{k-1}(z)$.
  $x_k \leftarrow -\lfloor R_k \rfloor$
  $B(M_k) \leftarrow 0$
  % loop on the radius of the layers: counting points on these layers.
  % stop when the number of points is greater than the current index which means x_k has been found.
  while $B(M_k) \leq I$
    $B \leftarrow B(M_k)$
    $x_k \leftarrow x_k + 1$
    $B(M_k) \leftarrow \mathrm{Below}(M_k)$
  endwhile
  $x_k \leftarrow x_k - 1$
  $I \leftarrow I - B$
  $R_{k+1} \leftarrow (R_k^p - |x_k|^p)^{1/p}$
endfor
if $I = 1$ then $x_n \leftarrow R_n$
else $x_n \leftarrow -R_n$
endif
write M

ALGORITHM 5 (THETA FUNCTION FOR DECODING: $p = 1$ or 2)

Function $\mathrm{Theta}k = \mathrm{Theta}(k)$
% returns a vector of $R^p + 1$ integers.
$\mathrm{Theta1} = \cdots$ % strongly depends on p; for example, for $p = 1$,
    $\mathrm{Theta1} = [1, 2, \ldots 2]$.
if $k = 1$ then $\mathrm{Theta}k \leftarrow \mathrm{Theta1}$
endif
$\mathrm{LastTheta} \leftarrow \mathrm{Theta1}$
for $i = 1$ to $k - 1$ step 1
  $\mathrm{Theta}k \leftarrow \mathrm{conv}(\mathrm{LastTheta}, \mathrm{Theta1})$
  % conv denotes the convolutional product.
  $\mathrm{LastTheta} \leftarrow \mathrm{Theta}k$
endfor
return $\mathrm{Theta}k$

Acknowledgment

The authors want to acknowledge the anonymous reviewers for their advice which improved the quality of the correspondence.

References

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Apr. 1992.
[2] M. Barlaud, P. Solé, T. Gaidon, M. Antonini, and P. Mathieu, "Lattice vector quantization for multiscale image coding," *IEEE Trans. Image Processing*, vol. 3, pp. 367–381, July 1994.
[3] F. Chen, Z. Gao, and J. Villasenor, "Lattice vector quantization of generalized Gaussian sources," *IEEE Trans. Inform. Theory*, vol. 43, pp. 92–103, Jan. 1997.
[4] J. H. Conway and N. J. A. Sloane, "Fast quantizing and decoding algorithms for lattice quantizers and codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 227–232, Mar. 1982.

[5] ——, *Sphere Packing, Lattices and Group*. New York: Springer-Verlag, 1988.
[6] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 568–583, July 1986.
[7] T. R. Fischer and J. Pan, "Enumeration encoding and decoding algorithms for pyramid cubic lattice and trellis codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 2056–2061, Nov. 1995.
[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.
[9] D. G. Jeong and J. D. Gibson, "Uniform and piecewise uniform lattice vector quantization for memoryless Gaussian and Laplacian sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 786–804, May 1993.
[10] R. Laroia and N. Farvardin, "A structured fixed-rate vector quantizer derived from a variable-length scalar quantizer: Part I—Memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 851–867, May 1993.
[11] P. Loyer, "Réseaux arithmétiques et aspect des communications numériques," Ph.D. dissertation, Univ. Nice-Sophia Antipolis, France, 1995.
[12] P. Loyer, J. M. Moureaux, and M. Antonini, "Solving lattice codebook enumeration problem for generalized Gaussian sources," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, June 2000.
[13] J. M. Moureaux, M. Antonini, and M. Barlaud, "Counting lattice points on ellipsoids: Application to image coding," *IEE Electron. Lett.*, vol. 31, no. 15, pp. 1224–1225, July 1995.
[14] J. M. Moureaux, P. Loyer, and M. Antonini, "Efficient indexing method for lattice quantization applications," in *Proc. IEEE ICIP*, Lausane, Switzerland, Sept. 16–19, 1996, pp. 447–450.
[15] ——, "Low complexity indexing method for $\boldsymbol{Z^n}$ and $\boldsymbol{D_n}$ lattice quantizers," *IEEE Trans. Commun.*, vol. 46, pp. 1602–1609, Dec. 1998.
[16] P. Raffy, M. Antonini, and M. Barlaud, "Distortion-rate models for entropy-coded lattice vector quantization," *IEEE Trans. Image Processing*, vol. 9, pp. 2006–2017, Dec. 2000.
[17] P. Solé, "Counting lattice points in pyramids," in *Actes du Congrès Séries Formelles et Combinatoire Algébrique*. Montréal, QC, Canada: LACIM, June 1992, pp. 343–355.
[18] ——, "Generalized theta functions for lattice vector quantization," in *Coding and Quantization*, R. Calderbank, G. D. Forney, Jr., and N. Moayeri, Eds. Providence, RI: Amer. Math. Soc., Oct. 1992, vol. 14, DIMACS Series, pp. 27–32.

Annexe E

Distortion-Rate Models for Entropy-Coded Lattice Vector Quantization

IEEE Transactions on Image Processing vol.9, no.12, pp. 2006-2017, décembre 2000

J'ai co-écrit cet article avec Philippe Raffy et Michel Barlaud.

Résumé Dans cet article, nous avons proposé un algorithme de quantification vectorielle algébrique (QVA) combiné à un codeur entropique. Nous avons concentré nos travaux sur la modélisation de l'EQM ainsi que sur celle du code préfixe utilisé par le codeur. Dans un premier temps, nous avons généralisé la formule de Jeong et Gibson à des réseaux Λ quelconques et dans un deuxième temps nous avons proposé un modèle de débit pour la QVA à entropie contrainte efficace même à bas débits. Les résultats expérimentaux prouvent la précision de nos modèles.

Distortion-Rate Models for Entropy-Coded Lattice Vector Quantization

Philippe Raffy, Marc Antonini, and Michel Barlaud, *Member, IEEE*

***Abstract*—The increasing demand for real-time applications requires the use of variable-rate quantizers having good performance in the low bit rate domain. In order to minimize the complexity of quantization, as well as maintaining a reasonably high PSNR ratio, we propose to use an *entropy-coded lattice vector quantizer* (ECLVQ). These quantizers have proven to outperform the well-known EZW algorithm's performance in terms of rate-distortion tradeoff.**

In this paper, we focus our attention on the modeling of the mean squared error (mse) distortion and the *prefix code* rate for ECLVQ. First, we generalize the distortion model of Jeong and Gibson on fixed-rate cubic quantizers to lattices under a high rate assumption. Second, we derive new rate models for ECLVQ, efficient at low bit rates without any high rate assumptions. Simulation results prove the precision of our models.

***Index Terms*—Distortion-rate theory, entropy-coded LVQ, high rate quantization theory, lattice VQ, low rate distortion and rate models, prefix code rate, subband image coding.**

I. Introduction

SHANNON theory implies that the performance of a VQ[1] can come arbitrarily close to the theoretical optimal performance if the vector dimension is sufficiently high. Unfortunately, the computational complexity of an unconstrained code increases exponentially with dimension. In addition, the storage requirements can be very large. One solution to overcome this problem of dimensionality is to use some form of constrained VQ such as lattice VQ (LVQ). Lattice quantization can be viewed as a vector generalization of uniform scalar quantization. Like VQ, lattice VQ is able to take into account spatial dependencies between adjacent pixels as well as to take advantage of the n-dimensional space filling gain [28]. Whatever the source distribution is, lattice vector quantizers will always outperform uniform scalar quantizers. A lattice Λ in R^n is composed by all integer combinations of a set of linearly independent vectors $\mathbf{a_i}$ (the lattice's basis) such that

$$\Lambda = \{\mathbf{y} | \mathbf{y} = u_1\mathbf{a_1} + u_2\mathbf{a_2} + \cdots + u_n\mathbf{a_n}\}$$

where the u_i are integers. The partition of the space is hence regular and depends only on the chosen basis vectors $\mathbf{a_i} \in R^m$ $(m \geq n)$. Note that each set of basis vectors define a different lattice. Using lattice points as codewords avoids the task of designing the codebook. Fast encoding and decoding algorithms making use of simple rounding operations have been proposed by Conway and Sloane [9]. Consequently, the encoding and decoding speed does not depend on the number of codewords within the codebook. Thus, lattice vector quantizers offer the possibility of a substantial reduction in computational and storage complexity over unstructured full-search VQ designed by the GLA algorithm [27]. This is a reason for their great popularity [3], [11], [15], [35]. The interest of many researchers has also been stimulated by the attractive asymptotic (i.e., high bit rate) distortion-rate performance of lattices. Indeed, Gersho conjectured that in the asymptotic case and for variable-rate, a lattice is approximately optimal [14]. Note that this near optimality is also realized for any source with a uniform density. Although a constrained lattice VQ code may not be mathematically optimal in the low bit rate domain, the reduction in computational complexity may allow higher dimensions to be practical. This may lead to a better overall performance for the lattice VQ code at a given rate. However, little is known about lattice VQ in the low bit rate range using an entropy code. In [35], it is shown that good performance can be achieved by combining lattice VQ and a variable-rate coder [24], [35], [40]. This provides strong motivation for the use of entropy coded LVQ in the wavelet domain [3], [40]. Before introducing our contribution, we first discuss the choice of the lattice (which is not necessarily the same for high bit rates and low bit rates).

Until now, two methods have been explored to study lattices, both using the high-rate assumption. The first method is geometrical and is due primarily to the work of Conway and Sloane [9]. Under space filling considerations, they have reported the best known lattices for low dimensions ($n \leq 8$): the root lattices A_n ($n \geq 1$), D_n ($n \geq 2$), E_n ($n = 6, 7, 8$), the Barnes–Wall lattice Λ_{16}, and the Leech lattice Λ_{24} were found to be optimal in that sense. The second method is a generalization of the first one. High-resolution theory which applies immediately to lattice VQ, is here explicitly used and turns out to be a powerful tool for deriving analytical formulas as well as a convenient mean for analyzing and understanding LVQ [10], [22], [28], [41]. For low bit rates, much less is known about variable-rate LVQ. The granular distortion[2] is not only a function of the shape

Manuscript received January 18, 1999; revised June 20, 2000. This paper is based in part upon work supported by the Ministère des Affaires Etrangères under LAVOISIER's grant (1997). This work was presented in part at the International Conference on Image Processing, Washington, DC, October 1995. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

P. Raffy is with Identive Corporation, Palo Alto, CA 94305 USA (e-mail: philippe@identive.com).

M. Antonini and M. Barlaud are with the University of Nice-Sophia Antipolis, BP-121, F-06903 Sophia Antipolis Cedex, France (e-mail: am@i3s.unice.fr).

Publisher Item Identifier S 1057-7149(00)10074-0.

[1]VQ will be used both for vector quantization and vector quantizer.

[2]When treating nonasymptotic distortion (or rate models), we consider as negligible the total contribution to the average distortion from the overload region in the low-bit rate range since we can take arbitrary large codebooks. This assumption is justified since variable length coding is used.

of the lattice cell but also a function of the source distribution [36]. As a consequence, the hierarchy of lattices originally established in the high bit rate domain assuming piecewise constant density functions is no longer valid in the low-bit rate range since now the source distribution must be taken into account. In the case of generalized Gaussian sources with a shape parameter less than one, the superiority of the cubic $\mathbb{Z}$ lattice over the E_8 and Leech lattices has been recently established [38]. This result demonstrates the great interest in using a combined wavelet transform and a cubic LVQ scheme.

This paper deals with the modeling of the mean squared error (*mse*) distortion and the rate for LVQ. We propose a new distortion model valid for any lattices under a high rate assumption. We also derive two new rate models respectively valid for $\mathbb{Z}$ and any lattices, efficient at low bit rates without any high rate assumptions. Previous work on that topic provides upper bounds, lower bounds, asymptotic results for the distortion and the entropy, and sometimes they only apply to $\mathbb{Z}$ lattices. Most of the results are given in terms of the distortion-rate function for comparison to Shannon theory bounds. If the reader is interested, a thorough review of quantization can be found in [20]. These works give an appreciable insight into quantizers performance, but generally remain inadequate in the case of low bit rates. Furthermore, knowing the distortion-rate function of one quantizer does not straightforwardly give a way of optimizing its parameters (for example, the quantization step for a uniform scalar quantizer).

An important issue of lattice VQ is the tuning of its parameter, or *scaling factor*, γ which has the effect of increasing or decreasing the size of the basic quantization cell [22], [37]. Unlike previous approaches using high quantizer bit rate formulas, our paper gives an n-dimensional formulation of the nonasymptotic models of the distortion given originally by [36] in the scalar case, and also derives nonasymptotic models of the rate $R(\gamma)$ for several root lattices ($\mathbb{Z}^n$, D_n, E_n). The formulas can be easily extended to mixtures or even used for companders. The proposed estimation formulas allow one to avoid learning or measuring processes so that it can lead to reduction of tedious experimental processing, especially when using iterative processes or bit allocation algorithms. The distortion formulas come from work related to quantization noise spectra [8], [33], [36], [39], [44], [45]. Note that this "exact analysis" approach was also applied to feedback quantization systems such as Delta and Delta-Sigma modulators [6], [18], [21], [43]. The rate is estimated by using a prefix code [12], which is well-suited to the regular structure of a lattice and also allows particularly efficient indexing methods [13], [31], [40]. In this paper, we propose two models for the rate: one is dedicated to $\mathbb{Z}$ lattices only and turns out to be very precise whatever the source distribution is, while the second can apply to other root lattices under a mild geometrical approximation and remains valid only for Gaussian source distributions. To our knowledge, no significant work has been reported previously on the low bit rate modeling of prefix code rates, except the recent work of [24] which focused on the estimation of the entropy for a specific Laplacian pdf (probability density function). All the proposed models have proven to yield very accurate results compared to experimental values, usually below an average estimation error of 10%. Part of this work was presented in [2].

The remainder of this paper is organized as follows. In Section II, we give an insight into previous works related to the modeling of $D(\gamma)$. Then, we extend to arbitrary lattices the high-resolution formulation of the distortion given by Jeong and Gibson [22] for cubic lattices. In Section IV, nonasymptotic formulas of the LVQ prefix-code rate are derived as functions of the scaling factor γ and the accuracy of these formulas is discussed. Finally, in Section V, we present simulation results using a LVQ-wavelet coder scheme which we compare with those obtained with the formulas previously derived. We then demonstrate their accuracy when applied to subband coding.

II. Distortion: Previous Work

We assume that the source pdfs are symmetric with zero mean. For convenience's sake, the expressions of the rate and the distortion are given per vector.

A. High-Resolution Theory

High-resolution quantization theory is based on the following assumption: because there are so many output points, the probability density of the input is approximately constant across any particular input bin. Therefore, given the centroid $\mathbf{y_i}$ of the Voronoi cell V_i, we can estimate the probability of source vector $\mathbf{X}$ being quantized by $\mathbf{y_i}$ by the mean value theorem [14]

$$\Pr(\mathbf{X} \in V_i) = \int_{\mathbf{X} \in V_i} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} \cong f(\mathbf{y_i}) vol(V_i) \qquad \mathbf{y_i} \in V_i.$$

High-resolution quantization theory provides tractable equations for the performance of quantizers. Most of the distortion-rate theory literature has focused on the estimation of the distortion for fixed-rate quantizers. Nevertheless, the high bit rate approximation can be successfully used for the estimation of entropy as well [16], [17], [26].

Bennett [4] first applied the high-resolution approximation to develop a system performance formula for scalar quantizers. Later, Zador [41] proposed a vector quantizer version of the problem, and derived an asymptotic distortion bound for vector quantizers, also known as the Zador–Gersho formula [14]

$$\hat{D}_{p,\,L} \cong \hat{G}_{p,\,n} \frac{1}{L^{p/n}} \left[\int f_{\mathbf{X}}(\mathbf{x})^{(n/n+p)}\, d\mathbf{x} \right]^{(n+p/n)}$$

where

- L codebook size;
- n vector size;
- $\hat{G}_{p,n}$ coefficient of quantization of the n-dimensional quantizer corresponding to a L_p norm.

However, this result assumes that there is no overload distortion. When considering LVQ, truncation of the infinite lattice is required in order to define a finite length codebook or, similarly, to limit the maximum bit rate of the quantizer.

For quantization purposes, the source vectors must be scaled inside the codebook. Since the lattice is symmetric, centered at the origin and constituted by concentric shells of radius r_m ($m = 1, \cdots, +\infty$), this can be performed using the radial scaling γ [11]. Consequently, for high bit rates, designing the LVQ requires choosing the scaling factor $\hat{\gamma}$ which results in the best trade-off between the granular noise and the overload noise.

In [22], Jeong and Gibson proposed a distortion formula for fixed-rate lattice VQ based on the $\mathbb{Z}$ lattice. This formula takes into account both the granular and the overload noises

$$D_{p,\gamma,m} = \frac{\gamma^2}{12} \int_{\|\mathbf{x}\|_p \le \gamma r_m} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} + D_{overload}. \quad (1)$$

They also derived an expression for $\hat{\gamma}$ as a function of the source distribution. In the next section, we propose an extension of Gibson's results to other root lattices, particularly to the D_n and E_n lattices.

B. Nonasymptotic Granular Distortion

Contrary to the high bit rate case, the search for the optimal $\hat{\gamma}$ is not of interest anymore. Indeed, the contribution of the overload noise is negligible since we assume an infinite lattice. This hypothesis is justified because variable length coding is assumed. Consequently, the $D(\gamma)$ curve has no minimum anymore. Instead, we address how distortion can be accurately modeled without any asymptotic assumptions.

Based on the work of Clavier [8] and Widrow [39], we give below an analytical expression of the granular distortion for $\mathbb{Z}^n$ lattices, written as a function of the joint characteristic function $\Phi_{\mathbf{X_1}\cdots\mathbf{X_n}}$. Related developments can be found in [5], [23], and [42]. If the reader is interested, the proof of the proposed vector version of the distortion formula can be obtained from the work of Sripad and Snyder [36], who derived conditions under which the quantization error of a scalar quantizer is white. Here, we give the expression of the distortion in the form of the following proposition.

Proposition 1: Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a random vector of size n with joint probability density function $f_{X_1\cdots X_n}(x_1, \cdots, x_n)$ and marginal characteristic function $\Phi_{X_i}(u_i)$. Then, the per vector granular distortion resulting from the cubic LVQ of $\mathbf{X}$ with *mse* distorsion measure and scaling factor γ is given by

$$D_g(\gamma) = \frac{\gamma^2}{12} \sum_{p=1}^{n} \left[1 + \frac{6}{\pi^2} \sum_{k \neq 0} \frac{(-1)^k}{k^2} \Phi_{X_p}\left(\frac{2\pi k}{\gamma}\right) \right].$$

The proposed distortion model requires no assumption on the source distribution but only the knowledge of the marginal characteristic function $\Phi_{X_p}(u_p) = \Phi_{X_1\cdots X_p\cdots X_n}(0, \cdots, u_p, \cdots, 0)$. It is noteworthy that Proposition 1, similarly to the work of Sripad and Snyder, works for nonstationary sources. This is the main reason of its interest. Fig. 1 shows the validity and the efficiency of the proposed vector version of the distortion. A source with elliptical statistics is tested and values of the n-dimensional model turn out to be very close to the experimental ones. The source in Fig. 1 is a sequence of i.i.d. 2-D Gaussian vectors (x_i, y_i) with mean $(0, 0)$ and covariance matrices K, where $K_{11} = \sigma_x^2$, $K_{22} = \sigma_y^2$, and $K_{12} = K_{21} = 0$. It is noteworthy that, in the case of elliptical statistics, this extension of Sripad and Snyder's work is superior to any other reported works [22], [39].

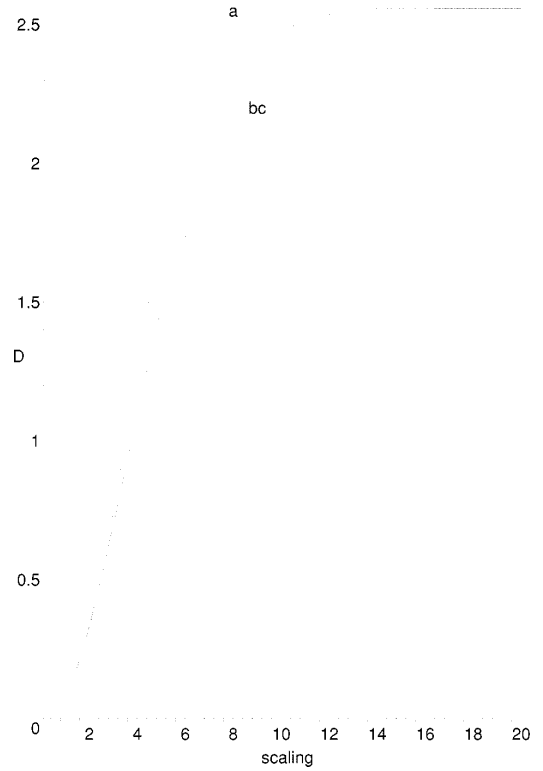


Fig. 1. Synthetic elliptical statistics ($\mu = 0, \sigma_x = 1$ and $\sigma_y = 2$): comparison of Widrow's distortion model [39] and the proposed model (Proposition 1). (a) Widrow's model, (b) proposed model, and (c) Experimental results (points).

III. Asymptotic Distortion: Extension to Arbitrary Lattices

It is well known that dense lattices such as D_4, E_8 or Λ_{16} provide lower granular distortion than the simple cubic lattice $\mathbb{Z}^n$. This remark stimulates our interest and we propose in this section a generalization of Jeong and Gibson asymptotic distortion model [22] [cf. formula (1)] for any lattices Λ and an *mse* distortion measure.

A. Background: Second Moment of a Polytope

First, recall the second moment of a polytope $P(\mathbf{y})$ defined by the following relationship:

$$G_n = \frac{1}{n} \frac{\int_{P(\mathbf{y})} \|\mathbf{x} - \mathbf{y}\|^2\, d\mathbf{x}}{vol(P(\mathbf{y}))^{1+(2/n)}}$$

with $vol(P(\mathbf{y})) = \int_P d\mathbf{x}$ and where $\mathbf{y}$ is a n-dimensional vector corresponding to the centroid of the elementary polytope $P(\mathbf{y})$. Note that G_n is a dimensionless quantity. Assuming that the source is scaled by a scaling factor γ, the second moment becomes

$$\begin{aligned} G_n &= \frac{1}{n} \frac{\int_{P(\mathbf{y})} \left\| \frac{\mathbf{x}}{\gamma} - \mathbf{y} \right\|^2 d\frac{\mathbf{x}}{\gamma^n}}{vol(P(\mathbf{y}))^{1+(2/n)}} \\ &= \frac{1}{n} \frac{\int_{P(\gamma\mathbf{y})} \|\mathbf{x} - \gamma\mathbf{y}\|^2\, d\mathbf{x}}{\gamma^{2+n} vol(P(\mathbf{y}))^{1+(2/n)}} \end{aligned} \quad (2)$$

where $P(\gamma\mathbf{y})$ is the polytope dilated by a scaling factor γ.

B. Distortion of a Scaled Polytope

The following demonstration is valid for any lattice Λ. If we consider a scaled Voronoi cell $V_i = P(\gamma \mathbf{y_i})$ with output vector $\gamma \mathbf{y_i}$, we can express the *mse* per dimension inside the cell by the relation

$$D_i = \frac{1}{n} \int_{P(\gamma \mathbf{y_i})} ||\mathbf{x} - \gamma \mathbf{y_i}||^2 f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}.$$

For high rate approximations, i.e., for large number of Voronoi cells of small volume (large codebook size) and a smooth pdf, we can make the approximation [14]

$$f_{\mathbf{X}}(\mathbf{x}) \approx f_{\mathbf{Y}}(\gamma \mathbf{y_i}), \qquad \text{for } \mathbf{x} \in P(\gamma \mathbf{y_i})$$

and simplify the expression of D_i as

$$D_i = \frac{1}{n} f_{\mathbf{Y}}(\gamma \mathbf{y_i}) \int_{P(\gamma \mathbf{y_i})} ||\mathbf{x} - \gamma \mathbf{y_i}||^2\, d\mathbf{x}. \tag{3}$$

Then, according to formula (2), formula (3) becomes[3]

$$D_i = \Pr\{\mathbf{X} \in P(\gamma \mathbf{y_i})\} G_n \gamma^2 vol(P(\mathbf{y_i}))^{(2/n)}.$$

C. Overall Distortion of a Scaled Lattice

Since a lattice corresponds to a set of regularly spaced vectors, then $P(\mathbf{y_i}) = P(\mathbf{y_j})$ and $vol(P(\mathbf{y_i})) = vol(P) = (\det \Lambda)^{1/2} \ \forall \ i, j$. The *mse* of a lattice consists of the *mse* occurring inside the codebook and the *mse* due to the overload distortion in the truncated region. Then, for a truncated lattice of size L, the total asymptotic distortion is given by

$$D_{\gamma, L} = \sum_{i=1}^{L} G_n \gamma^2 vol(P)^{2/n} \Pr\{\mathbf{X} \in P(V_i)\} + D_{overload}$$

which can be approximated by

$$D_{\gamma, m} = G_n \gamma^2 vol(P)^{2/n} \int_{\|\mathbf{x}\|_2 \le \gamma r_m} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} + D_{overload}. \tag{4}$$

When the number of codevectors, or equivalently m, is large, Moo and Neuhoff [29] showed that the granular distortion asymptoticaly dominates the overload distortion. Using their result, we have, according to formula (4)

$$\lim_{m \to +\infty} D_{\gamma, m} = G_n \gamma^2 vol(P)^{2/n} \tag{5}$$

and, thus, we have rederived Lemma 4 given by [29].

Our model [formula (4)] is a good approximation of the distortion for small values of γ, i.e., for high bit rates (see Fig. 2). However, for high values of γ (i.e., for low bit rates), it does not fit the real distortion, as does Jeong and Gibson's model.

IV. Non-Asymptotic Rate Models

To our knowledge, the rate estimation of variable-length lattice VQ in the low-bit rate range has not been investigated, except for [2] and, more recently, the work of Kim *et al.* [24]. The latter gives an approximation of the entropy estimation for a Laplacian distribution source. Here, we derive nonasymptotic models for the *prefix code* rate that remain valid for a class of monomodal sources (generalized Gaussian). Before presenting our models, let us introduce the prefix code we use.

[3]For a lattice $\mathbb{Z}^n$, $G_n = 1/12$ and $vol(P(\mathbf{y_i})) = 1$. Thus, $D_i = \Pr\{\mathbf{X} \in P(\gamma \mathbf{y_i})\}(\gamma^2/12)$ which corresponds to the result given by Jeong and Gibson in the case of the cubic lattice [22].

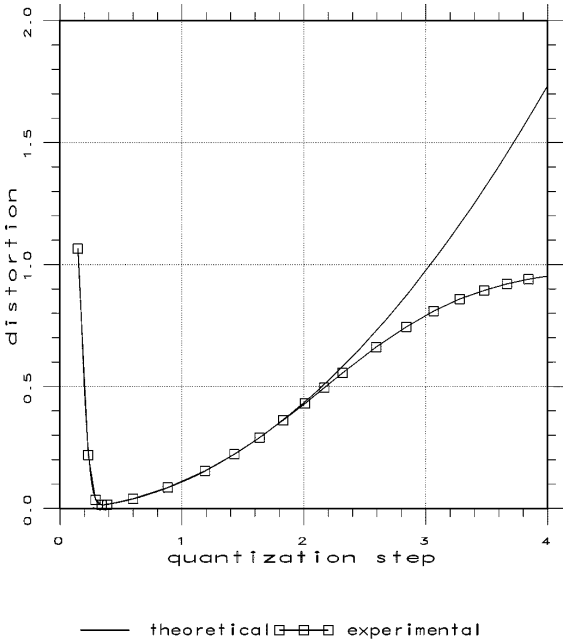


Fig. 2. Validity of the proposed asymptotic model [cf. (4)] for the lattice D_4 and a synthetic i.i.d. Gaussian distribution ($\mu = 0$, $\sigma = 1$).

A. Prefix Code

When defining prefix codes for lattices, the length of a codeword depends on the structure and density of the lattice, as well as on the shape considered to truncate it according to the source statistics [3], [11], [30]. In fact, each codeword is constructed by considering the probability of occurrence of the energy levels and the population of each energy level, i.e., the number of lattice vectors belonging to the same (hyper-)surface (with respect to a L_p norm). Here, we use an entropy code to describe the energy levels (a small number of levels implies an easy construction of this code). We propose the following starting formulation for the entropy of the quantized vectors $\mathbf{Y} = Q(\mathbf{X})$ (in bits per vector):

$$H(\mathbf{Y}) = -\sum_i \sum_j p(\mathbf{y_j}|r_i)\, \log_2\, p(\mathbf{y_j}|r_i) \tag{6}$$

where $p(\mathbf{y_j}|r_i) = \Pr\{\mathbf{Y} = \mathbf{y_j}|R = r_i\}$ is the probability of the source vector $\mathbf{X}$ being encoded by the vector $\mathbf{y_j}$ on the surface of radius r_i. Note that the radius $r(\mathbf{y_j})$ corresponding to quantized vector $\mathbf{y_j}$ is a deterministic function of $\mathbf{y_j}$, with $p(\mathbf{y_j}|r_i) = 0$ when $r_i \neq r(\mathbf{y_j})$. Let $\mathbf{Y}$ be the position random vector on a given surface of radius r_i with probability $\Pr\{\mathbf{Y} = \mathbf{y_j}\}$ and R the radius random variable with probability $p(r_i)$ [$p(r_i) = \Pr\{R = r_i\}$ is the probability of occurrence of the energy level r_i^2 after quantization]. The conditional entropy is [19]

$$H(\mathbf{Y}|R) = -\sum_i \sum_j p(r_i) p(\mathbf{y_j}|r_i)\, \log_2\, p(\mathbf{y_j}|r_i).$$

It can be also decomposed into two entropy terms

$$H(\mathbf{Y}|R) = H(\mathbf{Y}, R) - H(R)$$

where $H(R) = -\sum_i p(r_i) \log_2 p(r_i)$, and $H(\mathbf{Y}, R) = H(\mathbf{Y})$ since R is a deterministic function of $\mathbf{Y}$. Hence, the expression of entropy $H(\mathbf{Y})$ given by (6) becomes

$$H(\mathbf{Y}) = -\sum_i p(r_i) \cdot \left\{ \log_2 p(r_i) + \sum_j p(\mathbf{y_j}|r_i) \log_2 p(\mathbf{y_j}|r_i) \right\}. \quad (7)$$

The first term is the entropy of surfaces concerned with quantization (energy of vectors) and the second term is the entropy of the vectors located on a given surface of radius r_i. For large dimension, it is well known that the pdf of a Laplacian or a Gaussian source is constant on a hyper-sphere (Gaussian distribution) or hyper-pyramid (Laplacian distribution) of radius r_i. The source vectors are then uniformly distributed on that surface. A step further consists in assuming that the probability that a given source vector is encoded by a lattice point among the lattice points on the surface is equal to $(1/N(r_i))$, where $N(r_i)$ denotes the number of vectors lying on the same surface of radius r_i. $N(.)$ corresponds to the enumeration coefficients N_m given by the *theta* series of the lattice in the case of a Gaussian statistics [9], the *Nu* series in the case of a Laplacian statistics [3] or the *modified theta* series in the case of elliptical statistics [30]. Note that this hypothesis implies that $H(Y|R)$ is maximum. Therefore, if we assume that the lattice truncation energy m^2 is large enough to avoid overload distortion, then we are able to give an expression for the granular rate. According to formula (7), we have

$$R_g = -\sum_{i=0}^{m-1} p(r_i) \log_2 p(r_i) + \sum_{i=1}^{m-1} p(r_i) \log_2 N(r_i).$$

Note that we use a variable-length code[4] for the position of the vectors on the shells of constant energy (or constant pdf). Furthermore, we allow codewords to have noninteger lengths. R_g can also be conveniently expressed as a function of the scaling factor γ

$$R_g(\gamma) = -\sum_{i=0}^{m-1} p(\gamma r_i)[\log_2 p(\gamma r_i) - \log_2 N(r_i)]. \quad (8)$$

It follows from (8) that the estimation of the rate depends only on a good evaluation of the radius probability $p(\gamma r_i)$. This evaluation is the purpose of the next two subsections.

B. Signal Approach for the $\mathbb{Z}$ Lattice and Arbitrary Distribution Sources

For a given source with joint pdf $f_{X_1 \cdots X_n}(x_1 \cdots x_n)$, we demonstrate in this section that a discrete approximation of the radius probability is given by Proposition 2. The following takes into account the geometry of $\mathbb{Z}^n$ lattices and thus, is only valid for this kind of lattice. Furthermore, the proposed model authorizes both low and high rate applications.

Proposition 2: Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a random vector of size n with joint probability density function $f_{X_1 \cdots X_n}(x_1, \cdots, x_n)$. Then, the probability of a surface of constant energy r^p or equivalently the probability of radius r, inside the codebook granular region, resulting from the cubic LVQ of $\mathbf{X}$ with scaling factor γ and a smooth pdf is given by

$$\Pr\{\|\mathbf{Y}\|_p = \gamma r\} = \sum_{m_1=-\infty}^{+\infty} \cdots \sum_{m_n=-\infty}^{+\infty} \delta\left(\sum_{k=1}^{n} |m_k|^p - r^p\right) \times \int_{-\gamma/2}^{\gamma/2} \cdots \int_{-\gamma/2}^{\gamma/2} \times f_{X_1 \cdots X_n}(x_1 - m_1\gamma, \cdots, x_n - m_n\gamma)\, dx_1 \cdots dx_n$$

with δ the Kronecker, i.e., $\delta(u) = 1$ if $u = 0$ and 0 otherwise.

Proof: According to [25], it is possible to represent the expression of the distribution of a discrete quantized variable by a *symbolic probability density*, i.e., by a superposition of impulse functions concentrating the probabilities $\Pr\{\mathbf{X} \in P(\gamma \mathbf{y_i})\}$ at the Voronoi centroids $\mathbf{y_i}$. Extending this concept to vector quantization, it is possible to write

$$\begin{aligned}\Pr\{\mathbf{Y} = \mathbf{y}\} = &\sum_{m_1=-\infty}^{+\infty} \cdots \sum_{m_n=-\infty}^{+\infty} \\ &\times \delta(y_1 - m_1\gamma) \cdots \delta(y_n - m_n\gamma) \\ &\times \int_{\gamma(m_1-1/2)}^{\gamma(m_1+1/2)} \cdots \int_{\gamma(m_n-1/2)}^{\gamma(m_n+1/2)} \\ &\times f_{X_1 \cdots X_n}(x_1, \cdots, x_n)\, dx_1 \cdots dx_n \end{aligned} \quad (9)$$

where $\mathbf{y} = (y_1, \cdots, y_n)$ represents the quantized source vector $\mathbf{X}$. Equation (9) can be rewritten as

$$\begin{aligned}\Pr\{\mathbf{Y} = \mathbf{y}\} = &\sum_{m_1=-\infty}^{+\infty} \cdots \sum_{m_n=-\infty}^{+\infty} \\ &\times \delta(y_1 - m_1\gamma) \cdots \delta(y_n - m_n\gamma) \\ &\times \int_{-\gamma/2}^{\gamma/2} \cdots \int_{-\gamma/2}^{\gamma/2} f_{X_1 \cdots X_n} \\ &\times (x_1 - m_1\gamma, \cdots, x_n - m_n\gamma)\, dx_1 \cdots dx_n \end{aligned} \quad (10)$$

which corresponds to the probability of occurrence of a quantization vector $\mathbf{Y}$. Note that $\delta(y_1 - m_1\gamma) \times \cdots \times \delta(y_n - m_n\gamma) = 1$, if $y_i = m_i\gamma \ \forall\ i$.

Furthermore, the probability $\Pr\{\|\mathbf{Y}\|_p = \gamma r\}$ corresponds to the set $S = \{\mathbf{y}/\|\mathbf{y}\|_p^p = \gamma^p r^p\}$, i.e., the set of vectors $\mathbf{y}$ such that $\sum_i |y_i|^p = \sum_i |m_i\gamma|^p = \gamma^p r^p$. Then

$$\Pr\{\|\mathbf{Y}\|_p = \gamma r\} = \sum_{\mathbf{y} \in S} \Pr\{\mathbf{Y} = \mathbf{y}\}. \quad (11)$$

Let us introduce the following Kronecker function

$$\delta\left(\sum_{k=1}^{n} |m_k|^p - r^p\right) = \begin{cases} 1, & \text{if } r^p = \sum_{k=1}^{n} |m_k|^p \\ 0, & \text{elsewhere.} \end{cases}$$

[4]Despite the fact that variable length coding is not achievable in all cases, the proposed rate formula can be considered as accurate. Practically, $\lceil \log_2 N(r_i) \rceil$, e.g., a binary code is used for the densest surfaces while $\log_2 N(r_i)$, e.g., a variable length code is used otherwise.

Hence, (11) becomes

$$\Pr\{\|\mathbf{Y}\|_p = \gamma r\} = \sum_{\mathbf{y}} \delta(\|\mathbf{y}\|_p = \gamma r) \Pr\{\mathbf{Y} = \mathbf{y}\}$$
$$= \sum_{m_1} \cdots \sum_{m_n} \delta\left(\sum_{k=1}^{n} |m_k|^p - r^p\right) \cdot \Pr\{\mathbf{Y} = (\gamma m_1, \cdots, \gamma m_n)\}.$$

Therefore, the substitution of $\Pr\{\mathbf{Y}\}$ given by formula (10) into the previous equation implies Proposition 2.

Introduction of Proposition 2 in formula (8) permits us to estimate the prefix code bit rate, given the joint distribution of the source. However, it is valid only for cubic lattices and its extension to other lattices is difficult. In the next section, we propose another approach based on geometrical considerations which gives a close approximation of the rate for any lattices and Gaussian distribution sources.

C. Geometrical Approach for Arbitrary Lattices and Gaussian Distribution Sources

In this approach, we take into account the geometrical structure of the lattice. We need the knowledge of an analytical expression for the *radius density* $f_n(r)$ of a given source pdf $f_{X_1 \cdots X_n}(x_1, \cdots, x_n)$. In the case of a i.i.d. generalized Gaussian distribution

$$f_{X_1 \cdots X_n}(X_1, \cdots, X_n) = A^n \exp\left(-\sum_{i=1}^{n} |bx_i|^\alpha\right)$$
$$\text{with } A = \frac{b\alpha}{2\Gamma\left(\frac{1}{\alpha}\right)} \text{ and } b = \frac{1}{\sigma}\sqrt{\frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}} \quad (12)$$

where σ is the standard deviation and α the shape parameter. Given $r = (X_1^\alpha + \cdots + X_n^\alpha)^{1/\alpha}$, it was shown in [7] that

$$f_n(r) = K_n r^{n-1} e^{-(br)^\alpha} \quad \text{with} \quad K_n = \frac{\alpha b^n}{\Gamma\left(\frac{n}{\alpha}\right)}. \quad (13)$$

Note that for $\alpha = 2$ and $\alpha = 1$, we respectively recognize the Gaussian and Laplacian cases. In order to increase the accuracy of the rate model, the random vector space is separated into two regions ($r_i = 0$ and $r_i > 0$) according to the geometric shape of the source pdf. As a matter of fact, the case at the origin is very important since it corresponds to very low bit rates ($\gamma \to \infty$). We give the expression of the prefix code rate in the form of the following proposition.

Proposition 3: Let $f_n(r)$ be the radius density function of the n-dimensional random vector $\mathbf{X}$ with joint i.i.d. Gaussian probability density function $f_{X_1 \cdots X_n}(x_1, \cdots, x_n)$. Then, an accurate estimate of the radius probability of an arbitrary lattice Λ is given by

$$\begin{cases} \hat{p}_\Lambda(0) = \left[\int_{-\gamma r_{eq}}^{\gamma r_{eq}} f_X(x)\,dx\right]^n \\ \hat{p}_\Lambda(\gamma r_i) \cong C^{st} N(r_i) Pol_n(\theta(r_i)) \int_{\gamma(r_i - r_{eq})}^{\gamma(r_i + r_{eq})} f_n(r)\,dr \\ \qquad \forall\, r_i > 0 \end{cases}$$

where

- r_{eq} function of the packing radius ρ_Λ of the cubic lattice;
- $N(r_i)$ number of vectors lying on the hyper-sphere with radius r_i;
- Pol_n polynomial function which characterizes the geometry of the lattice;
- C^{st} a constant.

Proof: We first establish the estimate of the radius probability for a cubic lattice $\mathbb{Z}^n$ ($r_{eq} = \rho$) before generalizing to other root lattices.

1) At $r = 0$

For the Voronoi cell at the origin, it is possible to determine the true probability $p_{\mathbb{Z}^n}(0)$ since we are able to integrate over the correct volume. Thus, given that $r_{eq} = \rho$ the solution given in Proposition 3 is straightforward.

2) At $r > 0$

To estimate the probability $p(\gamma r_i)$, we propose the following geometrical approximation in order to find a tractable solution for any lattices. Instead of integrating over the cubic Voronoi, we rather integrate over a domain H_n (see Fig. 3) such that every source vector in the domain H_n are quantized by $\mathbf{y_j}$, centroid of the cubic Voronoi. This approximation is convenient because we transform our original integration problem over a nonseparable domain into a separable integration over a spherical domain. Thus, we can write

$$\hat{p}(\gamma r_i) \cong N(r_i) p\{\gamma(r_i - \rho) < r < \gamma(r_i + \rho)\}$$
$$= N(r_i) \int \cdots \int_{H_n} f_X(x_1, \cdots, x_n)\,dx_1 \cdots dx_n. \quad (14)$$

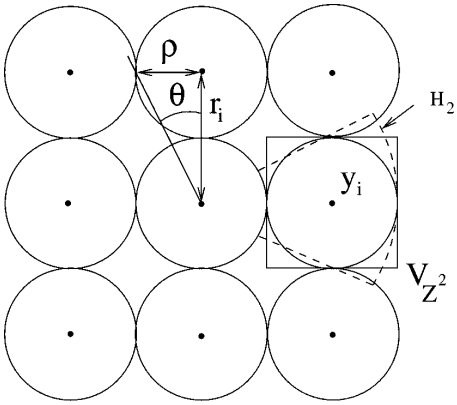


Fig. 3. $\mathbb{Z}^2$ lattice with packing radius ρ. The approximation consists in integrating over the domain H_2 (dashed lines) instead of the squared Voronoi $V_{\mathbb{Z}^2}$.

Translate the problem into spherical coordinates

$$\begin{cases} x_1 = r\cos\theta_{n-1}\cos\theta_{n-2}\cdots\cos\theta_2\cos\theta_1 \\ x_2 = r\cos\theta_{n-1}\cos\theta_{n-2}\cdots\cos\theta_2\sin\theta_1 \\ \vdots \\ x_{n-1} = r\cos\theta_{n-1}\sin\theta_{n-2} \\ x_n = r\sin\theta_{n-1}. \end{cases}$$

Hence, we have n variables: $\theta_1, \cdots, \theta_{n-1}, r$. Angle θ depends on the packing radius ρ of the $\mathbb{Z}^n$ lattice and is defined by $\theta(r_i) = \arctan(\rho/r_i)$ (see Fig. 3). The Jacobian of this system is

$$\begin{cases} |J| = r^{n-1}J_1 \\ \text{with } J_1 = (\cos\theta_{n-1})^{n-2}(\cos\theta_{n-2})^{n-3}\cdots(\cos\theta_3)^2 \cdot(\cos\theta_2) \\ \text{and } \forall\, n \in N \quad -\dfrac{\pi}{2} \le \theta_n \le \dfrac{\pi}{2} \end{cases}.$$

Thus, we can rewrite (14) as follows:

$$\hat{p}(\gamma r_i) \cong N(r_i)\frac{A}{K_n}\int\cdots\int_{H_n} f_n(r)J_1\,dr\,d\theta_1\cdots d\theta_{n-1}$$

with constants A and K_n of the Gaussian distribution given by (12) and (13) for $\alpha = 2$. The integration is therefore easy since we can separate the multiple integral into simple integrals. This yields

$$\begin{aligned}\hat{p}(\gamma r_i) \cong & N(r_i)\frac{A}{K_n}\left(\int_{-\theta(r_i)}^{\theta(r_i)} d\theta_1\right)\cdots \\ & \cdot\left(\int_{-\theta(r_i)}^{\theta(r_i)}(\cos\theta_{n-1})^{n-2}\,d\theta_{n-1}\right) \\ & \cdot\int_{-\gamma(r_i-\rho)}^{\gamma(r_i+\rho)} f_n(r)\,dr. \end{aligned} \tag{15}$$

The product of simple integrals depending on angle θ gives a function of θ that we call the *polynomial characteristic function*. In order to determine its expression, we need to evaluate $I_n = \int_0^a(\cos\theta)^n\,d\theta$. One easily shows that this suite is defined by

$$\begin{cases} nI_n = B_{n-1} + (n-1)I_{n-2} \\ I_0 = a,\ I_1 = \sin a,\ B_{n-1} = (\cos a)^{n-1}\sin a \end{cases}. \tag{16}$$

From now on, we only consider even n-dimensional lattices ($n = 2p$) so that (16) simplifies to

$$\begin{aligned} I_{2p} = & \frac{1\times 3\times 5\cdots(2p-1)}{2\times 4\times 6\cdots 2p}I_0 + \frac{B_{2p-1}}{2p} \\ & + \sum_{k=3}^{2p-1}\frac{(2p-1)\cdots[2p-(k-2)]}{2p\cdots[2p-(k-1)]}B_{2p-k}. \end{aligned}$$

Substituting this result into expression (15) yields the final expression of the radius probability. In what follows, we give formulas for small dimensions ($n \le 8$). Note that they can be easily extended to large ones. Expressions are given with parameter $\alpha = 2$ of the Gaussian distribution

$$\begin{cases} \hat{p}_{\mathbb{Z}^1}(r_i) = \frac{1}{2}N(r_i)Pol_{\mathbb{Z}^1}\{\theta(r_i)\}\displaystyle\int_{\gamma(r_i-\rho)}^{\gamma(r_i+\rho)} f_1(r)\,dr \\ \hat{p}_{\mathbb{Z}^2}(r_i) = \dfrac{1}{2\pi}N(r_i)Pol_{\mathbb{Z}^2}\{\theta(r_i)\}\displaystyle\int_{\gamma(r_i-\rho)}^{\gamma(r_i+\rho)} f_2(r)\,dr \\ \hat{p}_{\mathbb{Z}^4}(r_i) = \dfrac{1}{2\pi^2}N(r_i)Pol_{\mathbb{Z}^4}\{\theta(r_i)\}\displaystyle\int_{\gamma(r_i-\rho)}^{\gamma(r_i+\rho)} f_4(r)\,dr \\ \hat{p}_{\mathbb{Z}^8}(r_i) = \dfrac{3}{\pi^4}N(r_i)Pol_{\mathbb{Z}^8}\{\theta(r_i)\}\displaystyle\int_{\gamma(r_i-\rho)}^{\gamma(r_i+\rho)} f_8(r)\,dr. \end{cases}$$

TABLE I
ACCURACY OF THE GEOMETRICAL $R(\gamma)$ MODEL FOR A SYNTHETIC i.i.d. GAUSSIAN SOURCE ($\mu = 0$, $\sigma = 1$)

γ	$R^{th}_{Z^8}$	$R^{exp}_{Z^8}$	$R^{th}_{D_4}$	$R^{exp}_{D_4}$	$R^{th}_{E_8}$	$R^{exp}_{E_8}$
0.5	3.048	3.041	2.705	2.700	2.667	2.669
1.0	2.099	2.091	1.918	1.862	2.147	2.084
1.5	1.574	1.576	1.425	1.356	1.670	1.556
2.0	1.217	1.230	1.057	1.023	1.364	1.219
2.5	0.928	0.949	0.732	0.733	1.084	0.963
3.0	0.679	0.687	0.445	0.454	0.780	0.739
3.5	0.467	0.471	0.231	0.244	0.499	0.505
4.0	0.301	0.297	0.100	0.101	0.272	0.271
4.5	0.183	0.185	0.037	0.041	0.122	0.126
5.0	0.106	0.104	0.011	0.016	0.045	0.050

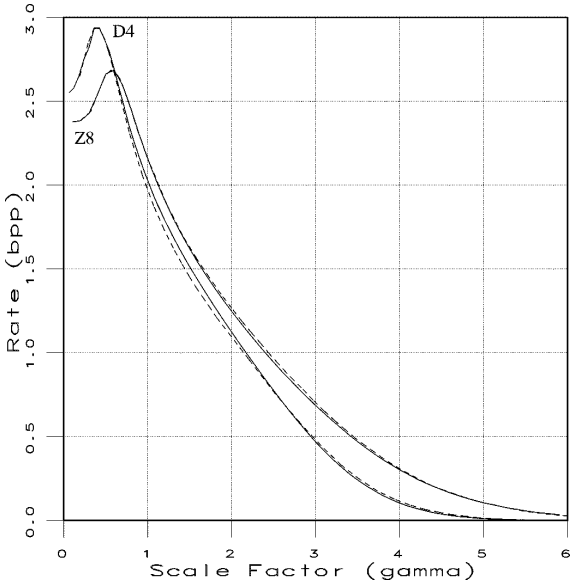


Fig. 4. Validity of the geometrical proposed rate model for two different lattices ($\mathbb{Z}^8$ and D_4) in the case of a synthetic i.i.d. Gaussian distribution. Dashed line—analytical model. Solid line—experimental results.

Fig. 5. Image "Nimes," 488 × 488 pixels coded on 8 bpp.

Their corresponding characteristic polynomial functions are

$$\begin{cases} Pol_{\mathbb{Z}^1}(\theta) = 1 \\ Pol_{\mathbb{Z}^2}(\theta) = 2\theta \\ Pol_{\mathbb{Z}^4}(\theta) = 4\theta \sin\theta(\theta + \cos\theta \sin\theta) \\ Pol_{\mathbb{Z}^8}(\theta) \\ \quad = 4\theta \sin\theta(\theta + \cos\theta \sin\theta)\left(\frac{2}{3}(\cos\theta)^2 \sin\theta + \frac{4}{3}\sin\theta\right) \\ \qquad \times \left(\frac{1}{2}(\cos\theta)^3 \sin\theta + \frac{3}{4}\cos\theta \sin\theta + \frac{3}{4}\theta\right) \\ \qquad \times \left(\frac{2}{5}(\cos\theta)^4 \sin\theta + \frac{8}{15}(\cos\theta)^2 \sin\theta + \frac{16}{15}\sin\theta\right) \\ \qquad \times \left(\frac{1}{3}(\cos\theta)^5 \sin\theta + \frac{5}{8}\theta + \frac{5}{8}\cos\theta \sin\theta \right. \\ \qquad\quad \left. + \frac{5}{12}(\cos\theta)^3 \sin\theta\right). \end{cases}$$

The complete demonstration of Proposition 3 requires a normalization operation of the radius probability such that

$$\hat{p}(0) + K_s \sum_{i=1}^{+\infty} \hat{p}(r_i) = 1$$

where K_s is a constant introduced in the final expression of $\hat{p}(r_i)$ in order to meet the normalization equation. Consequently, C^{st} of Proposition 3 can be written as follows:

$$C^{st} = K_s 2^{n-1} \frac{\Gamma\left(\frac{n}{2}\right)}{\left[2\Gamma\left(\frac{1}{2}\right)\right]^n}.$$

Fig. 6. Image "Nimes" coded on 1.08 bpp (compression ratio of about 7.40 : 1) using wavelet LVQ, with $\mathbb{Z}^8$ lattice and bit allocation based on exact distortion-rate models. The wavelet coefficients' pdf is modeled as a Laplacian. Peak SNR = 37.09 dB.

In the general case of a lattice Λ, the proposed estimated probability model given previously remains valid. However, we need to find an *equivalent radius* r_{eq} such that the density of the $\mathbb{Z}^n$ lattice with radius r_{eq} corresponds to that of Λ. It follows that

$$r_{eq} = \frac{\rho_\Lambda}{\det M^{1/n}} \tag{17}$$

where ρ_Λ is the packing radius of a lattice Λ, M its generator matrix and $\det M$ the volume of the fundamental region of Λ. Then, for any lattices ρ is replaced by r_{eq} and θ updated accordingly. Hence, we obtain the expression of $p(\gamma r_i)$ given in Proposition 3.

V. Experimental Results

A. Validity of the Asymptotic Distortion Model

We present results for the lattice D_4 and a synthetic i.i.d. Gaussian source ($\mu = 0$ and $\sigma = 1$). Fig. 2 shows the validity of the asymptotic distortion model given by (4). We can see that a good approximation is done for small values of γ, i.e., high bit rates. Obviously, for large values of γ, the model is no longer valid due to the high-resolution hypothesis.

B. Validity of Rate Models

1) Exact $\mathbb{Z}^n$ Model: Experiments show that the exact rate model is very accurate over the full range of γ. In the case of a synthetic source with i.i.d. Laplacian distribution and a $\mathbb{Z}^8$ lattice, the average estimation error is below 0.32%.

2) Approximate Model for Arbitrary Lattices: In Table I and Fig. 4, we present the average rate for various γ's and different lattices, in the case of a synthetic source with i.i.d. Gaussian distribution. Assuming that there is no overload, the rate estimation is very accurate over a large range of γ. For a $\mathbb{Z}^n$ lattice, the rate model is very accurate since the average estimation error is below 3%. In the general case of a lattice Λ (see Proposition 3), the *extra-geometrical* approximation given by (17) is needed to correct the fact that the geometry of the Voronoi cell differs from that of the hyper-cube. This approximation gives quite satisfying results since the average estimation error for both D_4 and E_8 lattices is below 10%, which is acceptable for most cases (Table I). Since the prefix code rate estimation has not been derived previously, it is not possible to compare our estimation with other works.

C. Subband Image Coding

In this section, we investigate the performance of the proposed distortion and rate models over real life images. The considered images are a simulated satellite image called Nimes[5] (see Fig. 5) and the well-known Lena image (from the Rensselaer Polytechnic Institute site). For image compression we used the biorthogonal 9-7 tap filters [1] and perform optimal bit allocation using the algorithm proposed in [32]. In the following experiments we used the cubic lattice $\mathbb{Z}^n$, and provide distortion-rate theoretical estimates using the nonasymptotic distortion model of Proposition 1 and the prefix code rate given by (8) with exact probability given by Proposition 2.

Rate-distortion curves are plotted on Figs. 7 and 8 with bit allocation performed on both theoretical estimations and experimental ones. The lattice we used is a $\mathbb{Z}^8$ lattice and the wavelet coefficients' pdf model considered by the theoretical approximations is Laplacian. The Laplacian models are adjusted for each subband by making the variance of the model equal to the measured subband variance. The experimental rate is estimated using a Huffman code for the prefix, i.e., for the energy part of the prefix code. We can see that the theoretical and experimental rate-distortion curves are very close, providing identical reconstructed images. However, the Laplacian model fits better the wavelet coefficients of Nimes than those of Lena, which explain the gap between the two plots of Fig. 8.

An example of coded/decoded image "Nimes" is presented on Fig. 6 for a compression ratio of 7.40 : 1 (1.08 bpp). In that case, bit allocation was performed using the exact theoretical distortion and rate models given proposition 1 and 2 and (8), with Laplacian wavelet coefficients models.[6] For that image, the peak SNR is 37.09 dB that is about 0.5 dB below that of the

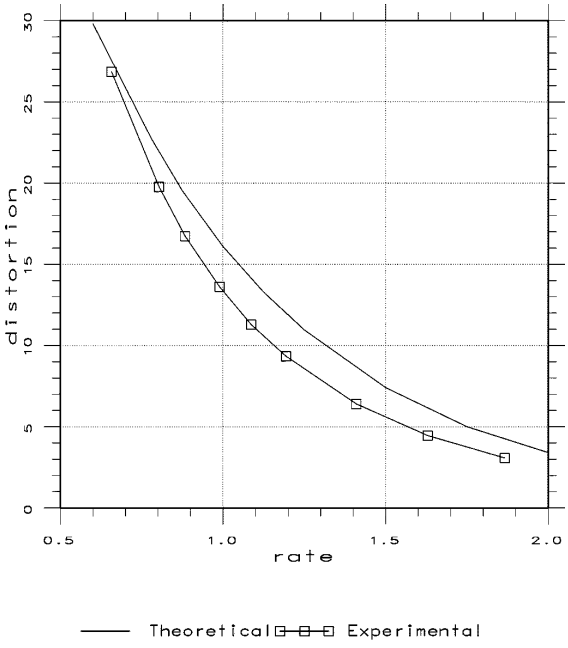


Fig. 7. Exact theoretical and experimental rate-distortion curves for Nimes using wavelet LVQ ($\mathbb{Z}^8$ lattice) in the high bit rate range. The different subband wavelet coefficients' pdfs are modeled as Laplacian. The vectors are represented by a prefix code given by (8). For the experimental results, the prefix is coded using a Huffman code.

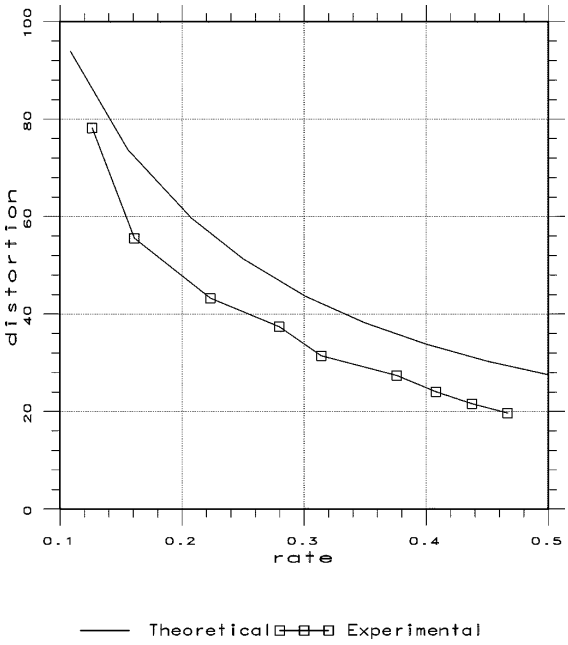


Fig. 8. Exact theoretical and experimental rate-distortion curves for Lena using wavelet LVQ ($\mathbb{Z}^8$ lattice) in the low bit rate range. The different subband wavelet coefficients' pdfs are modeled as Laplacian. The vectors are represented by a prefix code given by (8). For the experimental results, the prefix is coded using a Huffman code.

[5]This image is provided by the Centre National d'Etudes Spatiales (CNES) of Toulouse—France.

[6]Note that although the bit allocation algorithm was performed using theoretical distortion and rate models, final bit rate given on Fig. 6 is real estimate (Huffman coding performed on the prefix code).

coded/decoded image performing bit allocation on the real data. This difference is essentially due to the pdf model we used. Furthermore, comparison with the well-known SPIHT algorithm [34] shows that our results are about 0.7 dB lower in the case of theoretical approximation. This gap becomes about 0.2 dB when performing bit allocation on real data.

It is noteworthy that the performance of our distortion-rate models for ECLVQ quantizers relies on the goodness of fit of the assumed source pdf to the real data. For a few images, the development of models based on a generalized Gaussian pdf may be of interest but still remains an open issue.

VI. Conclusion

We have developed analytical models of both the distortion and the prefix code rate for ECLVQ quantizers. The distortion has been studied in two situations: i) within the high-resolution framework and ii) when no asymptotic assumption is explicitly made. In the first case, we extended the asymptotic models of Jeong and Gibson on $\mathbb{Z}$ lattices [22], to arbitrary lattices. We also rederived a formulation of the asymptotic granular distortion [(5)], recently found by Moo and Neuhoff [29]. In the second case, we gave a vector version of Sripad and Snyder's work on scalar quantization [36] (Proposition 1). We have then derived new formulas of the prefix code rate, valid in the low bit rate range as well as in the high bit rate range. To our knowledge, the prefix code rate model has not been investigated before. We have proposed two models. One is an exact estimation of the prefix code rate for $\mathbb{Z}^n$ lattices and arbitrary pdf (Proposition 2), the other is an accurate approximation of the prefix code rate for any lattices and Gaussian pdf (Proposition 3). Experiments prove the precision of our models. Thanks to the estimation formulas, the coding algorithm does not require any learning or measuring processes. Hence, computational complexity of the subband coder design based on ECLVQ is considerably reduced.

Acknowledgment

The authors are indebted to Prof. R. M. Gray and Dr. C. Pépin for helpful comments and discussion. They also wish to thank the reviewers for their many helpful suggestions to improve the paper.

References

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Apr. 1992.

[2] M. Antonini, P. Raffy, and M. Barlaud, "Toward entropy constrained lattice vector quantization," in *Proc. Int. Conf. Image Processing*, Washington, DC, Oct. 1995, pp. 121–124.

[3] M. Barlaud, P. Solé, T. Gaidon, M. Antonini, and P. Mathieu, "Pyramidal lattice vector quantization for multiscale image coding," *IEEE Trans. Image Processing*, vol. 3, pp. 367–381, July 1994.

[4] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, 1948.

[5] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with rth power distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 239–247, Mar. 1982.

[6] J. C. Candy and O. J. Benjamin, "The structure of quantization noise from sigma-delta modulation," *IEEE Trans. Commun.*, vol. COM-29, pp. 1316–1323, Sept. 1981.

[7] F. Chen, Z. Gao, and J. Villasenor, "Lattice vector quantization of generalized Gaussian sources," *IEEE Trans. Inform. Theory*, vol. 43, pp. 92–103, Jan. 1997.

[8] A. G. Clavier, P. F. Panter, and D. D. Grieg, "PCM distortion analysis," *Elect. Eng.*, pp. 1110–1122, Nov. 1947.

[9] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Berlin, Germany: Springer-Verlag, 1988.

[10] M. V. Eyuboglu and G. D. Forney Jr., "Lattice and trellis quantization with lattice and trellis-bounded codebooks—High-rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 46–59, Jan. 1993.

[11] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, vol. 32, pp. 568–583, July 1986.

[12] ——, "Geometric source coding and vector quantization," *IEEE Trans. Inform. Theory*, vol. 35, pp. 137–145, Jan. 1989.

[13] T. R. Fischer and J. Pan, "Enumeration encoding and decoding algorithms for pyramid cubic lattice and treillis codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 2056–2061, Nov. 1995.

[14] A. Gersho, "Asymptoticaly optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[15] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.

[16] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sept. 1968.

[17] T. J. Goblick and J. L. Holsinger, "Analog source digitalization: A comparison of theory and practice," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 323–326, Apr. 1967.

[18] R. M. Gray, "Quantization noise spectra," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1220–1244, Nov. 1990.

[19] ——, *Source Coding Theory*. Norwell, MA: Kluwer, 1990.

[20] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, Oct. 1998.

[21] J. E. Iwersen, "Calculated quantizing noise of single-integration delta-modulation coders," *Bell Syst. Tech. J.*, pp. 2359–2389, Sept. 1969.

[22] D. G. Jeong and J. D. Gibson, "Uniform and piecewise uniform lattice vector quantization for memoryless Gaussian and Laplacian sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 786–804, May 1993.

[23] J. C. Kieffer, "Stochastic stability for feedback quantization schemes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 248–254, Mar. 1982.

[24] W. H. Kim, Y. H. Hu, and T. Nguyen, "Joint optimization of lattice vector quantizer and entropy coder in subband coding," *Proc. SPIE*, vol. 3078, 1997.

[25] G. A. Korn, "Hybrid-computer techniques for measuring statistics from quantized data," *Simulation*, vol. 7, pp. 229–239, Apr. 1965.

[26] V. Koshélev, "Estimation of mean error for a discrete successive-approximation scheme," *Probl. Pered. Informat.*, vol. 17, pp. 20–33, July–Sept. 1981.

[27] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.

[28] T. D. Lookabaugh and R. M. Gray, "High resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1020–1033, 1989.

[29] P. W. Moo and D. L. Neuhoff, "An asymptotic analysis of fixed-rate lattice vector quantization," in *Proc. Int. Symp. Information Theory Applications*, Victoria, BC, Canada, Sept. 1996, pp. 409–412.

[30] J. M. Moureaux, M. Antonini, and M. Barlaud, "Counting lattice points on ellipsoids: Application to image coding," *Electron. Lett.*, vol. 31, pp. 1224–1225, July 1995.

[31] J. M. Moureaux, P. Loyer, and M. Antonini, "Low-complexity indexing method for $\mathbb{Z}^n$ and D_n lattice quantizers," *IEEE Trans. Commun.*, vol. 46, pp. 1602–1609, Dec. 1998.

[32] P. Raffy, M. Antonini, and M. Barlaud, "Optimal subband bit allocation procedure for very low bit rate image coding," *Electron. Lett.*, vol. 34, pp. 647–648, Apr. 1998.

[33] S. O. Rice, "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Ed. New York: Dover, 1954, pp. 133–294.

[34] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996.

[35] T. Senoo and B. Girod, "Vector quantization for entropy coding of image subbands," *IEEE Trans. Image Processing*, vol. 1, pp. 526–532, Oct. 1992.

[36] A. B. Sripad and D. L. Snyder, "A necessary and sufficient condition for quantization errors to be uniform and white," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, pp. 442–448, Oct. 1977.

[37] P. Swaszek, "A vector quantizer for the laplacian source," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1355–1365, Sept. 1991.
[38] Z. Gao, F. Chen, B. Belzer, and J. Villasenor, "A comparison of the Z, E_8 and Leech lattices for image subband quantization," *IEEE Data Compression Conf.*, pp. 312–321, Mar. 1995.
[39] B. Widrow, "Statistical analysis of amplitude quantized sampled data systems," *Trans. Amer. Inst. Elec. Eng.—Pt. II, Applicat. Ind.*, vol. 79, pp. 555–568, Jan. 1960.
[40] Z. M. Yusof and T. Fischer, "An entropy-coded lattice vector quantizer for transform and subband image coding," *IEEE Trans. Image Processing*, vol. 5, pp. 289–298, Feb. 1996.
[41] P. L. Zador, "Asymptotic quantization error of continuous signals and their quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, 1982.
[42] ——, "Topics in the asymptotic quantization of continuous random variables," Bell Labs. Tech. Memo., 1966.
[43] J. E. Iwersen, "Calculated quantizing noise of single-integration delta-modulation coders," *Bell Syst. Tech. J.*, vol. 44, Oct. 1998.
[44] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282–332, 1944.
[45] ——, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 24, pp. 46–156, 1945.

Philippe Raffy received the B.S. degree in electronic engineering from the University of Paris XI (Orsay), Paris, France, in 1992, and the M.S. (D.E.A. program) and Ph.D. degrees in electrical engineering from the University of Nice-Sophia Antipolis, Nice, France, in 1994 and 1997, respectively.

In February 1998, he joined the Information Systems Laboratory, Stanford University, Stanford, CA, as a Postdoctoral Member. Since October 1999, he has been with Identive Corporation, Palo Alto, CA, as a Senior Research Scientist in projects related to interactive television. His research interests include video sequence analysis, data compression, and transmission over noisy channels.

Marc Antonini received the Ph.D. degree in electrical engineering from the University of Nice-Sophia Antipolis, Nice, France, in 1991.

He was a Postdoctoral Fellow with the Centre National d'Etudes Spatiales, Toulouse, France, in 1991 and 1992. Since 1993, he has been with CNRS, I3S Laboratory, and the University of Nice-Sophia Antipolis. He is a regular reviewer for *Signal Processing*, *Information Theory*, and *IEE Electronics Letters*. He participates in several national research and development projects with French industries, and several international academic collaborations. His research interests include multidimensional image processing, wavelet analysis, vector quantization, information theory, still image and video coding, inverse problem for decoding, and multispectral image coding.

Dr. Antonini is a reviewer for the IEEE TRANSACTIONS ON IMAGE PROCESSING.

Michel Barlaud (M'88) received the "Agregation" from ENS Cachan and These d'Etat from the University of Paris XII, Paris, France.

He is currently a Professor of image processing with the University of Nice-Sophia Antipolis, Nice, France, and at the Head of the Image Processing Group, I3S Laboratory. He is a regular reviewer for several journals and a member of the technical committees of scientific conferences. He is the author of a large number of publications in the area of image processing, and the editor of the book *Wavelets and Image Communication* (Amsterdam, The Netherlands: Elsevier, 1994). His current research interests are still image and video Coding using wavelets and vector quantization, segmentation of video and 3-D $+t$ medical images using PDEs. He leads several national research and development projects with French industries, and participates in several international academic collaborations.

Dr. Barlaud served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.

Annexe F

Fractal Image Compression Based on Delaunay Triangulation and Vector Quantization

IEEE Transactions on Image Processing vol.5, no.2, pp. 521-528, février 1996

J'ai co-écrit cet article avec Franck Davoine, Jean-Marc Chassery et Michel Barlaud.

Résumé Dans cet article, nous avons proposé un nouveau schéma de compression par fractals basé sur une triangulation de Delaunay adaptative. Dans le but de réduire la complexité de codage des coefficients de la transformation, nous avons introduit une version modifiée de l'algorithme de Lloyd. Cette version modifiée s'applique directement sur les histogrammes des pixels contenus par les triangles issus de la triangulation et permet aussi de réduire la complexité de recherche liée à la tranformation fractale par IFS ("Iterated Function System"). L'algorithme que nous avons proposé présente de bonnes performances en terme de Rapport Signal-à-Bruit pour des débits supérieurs à 0,25 bits par pixels.

Fractal Image Compression Based on Delaunay Triangulation and Vector Quantization

Franck Davoine, Marc Antonini, Jean-Marc Chassery, and Michel Barlaud, *Member, IEEE*

Abstract—**This paper presents a new scheme for fractal image compression based on adaptive Delaunay triangulation. Such a partition is computed on an initial set of points obtained with a split and merge algorithm in a grey level dependent way. The triangulation is thus fully flexible and returns a limited number of blocks allowing good compression ratios. Moreover, a second original approach is the integration of a classification step based on a modified version of the Lloyd algorithm (vector quantization) in order to reduce the encoding complexity. The vector quantization algorithm is implemented on pixel histograms directly generated from the triangulation. The aim is to reduce the number of comparisons between the two sets of blocks involved in fractal image compression by keeping only the best representative triangles in the domain blocks set. Quality coding results are achieved at rates between 0.25–0.5 b/pixel depending on the nature of the original image and on the number of triangles retained.**

I. Introduction

IMAGES are increasingly present in data exchange processes. Compression techniques have made it possible, for example, for people from different countries to work together simultaneously, interacting on a common image through a communication medium.

Data compression reduces redundancy in data in order to store or transmit only a minimal number of bits per sample from which a good approximation of the original data can be reconstructed in accordance with human visual perception [19].

Many compression methods have been developed. One usually distinguishes between different coding schemes: transform coding; multiresolution coding; vector quantization; predictive methods; and other more recent schemes such as fractal image coding. Generally, bit rates around 0.8–0.4 b/pixel are reported for still monochrome images without a great loss of visual quality. Higher ratios can be obtained with hybrid coders incorporating different techniques with respect to local image properties.

With linear transform coding, the objective is to decorrelate the image signal using projections on a specified basis of functions. For example, the higher energy of a 2-D discrete cosine transformed (DCT) signal [8] is concentrated in the first area of the transformed domain, representing the lower frequency cosine functions. The compression is then achieved by quantizing and coding the transform coefficients.

With multiresolution coding, the image signal is decomposed into a finite set of subimages, which are usually encoded separately using different codebooks at different rates [1]. These codes are then recombined to restore the whole image. This class of methods regroups subband coding where each subimage lies in a specific frequency band, the Laplacian pyramids scheme, in which the different subimages correspond to distinct scale levels and the wavelet transform.

In vector quantization (VQ), sequences (or vectors) of pixels are encoded rather than each pixel separately. According to Shannon's rate-distortion theory, VQ should perform better than scalar quantization because it takes the samples' correlations into account. VQ can be applied on different types of samples including DCT or wavelet coefficients [1], [2].

Fractal image compression, based on concepts of iterated function systems (IFS) [3], can be seen as a kind of vector quantization and multiresolution coding. The method consists of partitioning blocks (vectors) and approximates each vector by a transformed codebook block derived from the image itself [17]. Each transform, described by a linear term and a translation term, maps a block onto another block with a different resolution and composes the coded information.

This paper focuses on a fractal image coding technique based on the Delaunay triangular partition, in association with a classification (vector quantization) of the triangles in order to reduce the complexity of the coding phase. The partitioning has the advantage of being fully flexible. It is computed on a set of points placed on the image support in a grey level-dependent way, with the help of a split and merge approach. The main advantage of such an algorithm is to reduce the block effects in the reconstructed image, at least on the diagonal image edges.

II. Fractal Image Coding

A. State of the Art

A number of papers on fractal image compression have been proposed by researchers since the original idea of Barnsley and Sloan in 1988 [3], implemented in a fully automated algorithm by Jacquin in 1989 [17]. Barnsley showed that it is possible to model a self-similar image as an attractor of an IFS, with the help of the collage theorem. Such an image is encoded with the very limited number of coefficients of the contractive mappings defining the IFS. The nonautomatic encoding algorithm he proposed (assistance of a "human

Manuscript received June 28, 1994; revised July 30, 1995.

F. Davoine and J.-M. Chassery are with TIMC-IMAG, URA D 1618 CNRS, Equipe Infodis, Institut Albert Bonniot, Faculté de médecine, 38706 La Tronche, France.

M. Antonini and M. Barlaud are with GDR PRC-ISIS, CNRS, MESR, Université de Nice-Sophia Antipolis, Valbonne, France.

Publisher Item Identifier S 1057-7149(96).

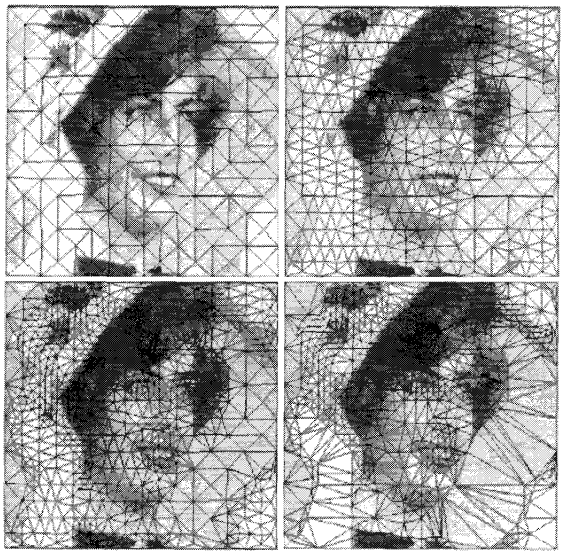


Fig. 1. Delaunay triangulation of the image Woman issued from a split-and-merge approach: (a) Initialization on a lattice of points and computation of its Delaunay triangulation (256 triangles); (b) first split step—each nonhomogeneous triangle is split by insertion of a point on its barycenter; the new Delaunay triangulation is composed of 676 triangles; (c) second split step—the new Delaunay triangulation is composed of 1288 triangles; (d) merge step applied after the second split step—by grouping similar adjacent triangles, the total number of Delaunay triangles is restricted to 1000.

operator" is required) consists of segmenting the image into self-similar objects. Each object, defining a part of the image, is covered by a set of contractive affine transformations of itself, with the property that the resulting union approximates the object. Each part of the image is coded by the coefficients of its associated set of transformations. The decoded image is formed by the association of the attractors of each IFS.

The encoding algorithm of Jacquin is fully automated because it does not require an "intelligent" segmentation of the image into distinct objects. The method is based on the observation that a real-world image is "affine-redundant." It exploits the partial self-transformability of the image.

The first step of his algorithm is to partition the image support into nonoverlapping square range blocks and larger square domain blocks. Given a range block, a domain block is searched such that it provides the best affine mapping to the range block in the root mean squared error (MSE) sense. The encoder attempts to find, for a given image, the set of transformations under which the distortion between the original image and the union of the transformed domain blocks (mapped onto the range blocks) is minimal. In his thesis, Jacquin proposed improvements by the use of a classification scheme and a split of range blocks to adapt the mappings to the local properties of the image.

Since the publication of this original work, a number of research areas have been investigated. They are summarized [11], [18], and [25] as follows:

- construction of the range and domain block partitions (different block shapes, different mappings)

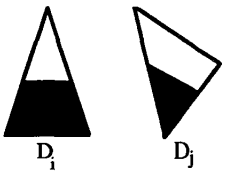


Fig. 2. Histogram-based classification.

- quantization of affine transformation parameters
- improvement of the encoding step— complexity reduction, introduction of fixed basis blocks, orthogonalization of the blocks
- decoding step—standard iteration, noniterative decoding, hierarchical decoding
- theoretical studies on the convergence to the attractor image and on the collage theorem.

B. Principle

Fractal image coding exploits the piecewise self-similarity of the image. The basic idea is to construct a contractive operator W in the metric space X of digital images, for which the image to be encoded is the unique fixed point x_0 .

For this, the class of affine operators is preferred because of their particularity to be "nearly linear," which makes them easy to analyze. The fixed point x_0 of such an operator $W: X \rightarrow X$ verifies

$$\begin{aligned} Wx_0 &= Ax_0 + b \\ &= x_0, \quad \text{where } A \text{ is a linear operator} \\ &\qquad \text{and } b \text{ belongs to } X. \end{aligned} \tag{1}$$

Real-world images are usually not self-similar—except contrived examples—and it is impossible to find an operator that maps a whole image x onto another image x_0 such as x equals x_0.

However, real-world images present piecewise self-similarity [17]: parts of the image resemble other parts. Starting from this observation, one can define an operator as the sum of piecewise mappings on X. A solution is to decompose the image into N nonoverlapping *range* blocks R and into M different *domain* blocks D and define the operator W as

$$Wx = W\left(\bigcup_{i=1}^{N} R_i\right) = \bigcup_{i=1}^{N} \omega_i(D_i). \tag{2}$$

We use the notation $D_i = x|_{Di}$ to denote the image x restricted to the domain part D_i, and ω_i to denote the transformation mapping the domain block D_i onto the range block R_i. The blocks R_i and D_i can have different sizes and shapes. The number M of blocks D_i is less than, greater to, or equal to N and therefore the blocks D_i can be overlapping or picked only from parts of the image.

An additional constraint that must be imposed on the operator W is that it is eventually contractive. W is *contractive* if there exists a constant $s < 1$ such that $d[W(x), W(y)] \leq s.d(x, y), \quad \forall x, y \in X$, where d is a given distance measure

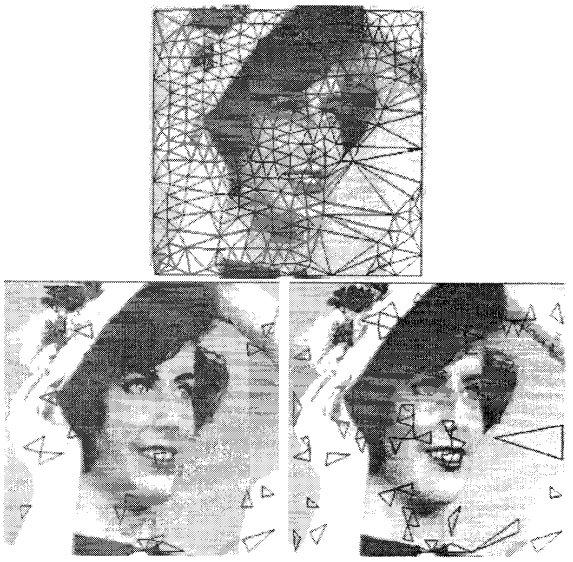


Fig. 3. Classification results. Top: the entire partition D composed of 576 triangles. Bottom: a codebook composed of 32 triangles D_i and a codebook composed of 64 triangles D_i.

and s is called the contractivity of W. W is said to be *eventually contractive* (at the Kth iterate) if there exists a constant K such that the Kth iterate of W (noted W^{oK}) is contractive.

If this is the case, the operator W has a unique fixed point x_0 obtained with $x_0 = \lim_{k\to\infty} W^{ok}x$, for an arbitrary x. The collage theorem proves that if the distance between an image x and its transformation by the eventually contractive operator W is small, then the distance between the image x and the fixed point x_0 of W will also be small. It provides a boundary given by:

$$d(x, x_0) \leq \frac{1}{1-s_k} \cdot \frac{1-s_1^K}{1-s_1} \cdot d(x, Wx) \tag{3}$$

where s_1 is the Lipschitz constant of W, K is the iterate number under which the operator W becomes contractive, and s_k is the contractivity of W^{oK}. Those results were initially published in [16], [22].

C. Partitionings

Block-based compression methods can, in theory, achieve high compression ratios by coding large blocks [26]. In practice, block sizes are limited by algorithm complexity. The aim of an adaptive partitioning is thus to group large blocks onto highly compressible regions of the image (homogeneous grey scales) and smaller size blocks onto detailed parts (random textures, edges).

Fisher [11] introduced adaptive partitions in order to increase the decoded image quality of fractal methods. As in Bedford *et al.* [4], he used a quadtree partitioning providing nonoverlapping square range blocks of arbitrary sizes from $32 * 32$ to$2 * 2$ pixels.

This type of partitioning has the property of being adaptive while staying rigid; the segmentation starts from a square and runs through a tree by recursive splitting. Each square not satisfying the property of uniformity (grey level mean, variance) is split into four equal-sized subsquares. The number and location of the squares is not precisely controlled due to the use of fixed-shape blocks.

A generalization of this scheme is the semirigid H–V (horizontal-vertical) partitioning, which recursively splits the initial image support into rectangles. The resulting H–V structure isolates long rectangles covering exactly or nearly horizontal or vertical edges [12].

These two segmentation procedures provide an effective combination between adaptivity to the image content, the blocks manipulation facilities (use of a tree structure), and the number of bits required to code the partition.

Triangular partitions have been proposed by Fisher [11] and used in [9] and [23] for fractal image compression. The advantage of such a partitioning is to break away from the rigid 90° angle rotation of the quadtree and H–V partitionings, and thus to be much more adaptable to the natural edges of the image. Fisher described an algorithm that recursively subdivides a triangle into four others in such a way that they share self-similar properties. We propose a more flexible scheme based on the well-known Delaunay triangulation, which allows control of the compression ratio as well as the quality of the attractor image.

III. Implementation

A. Tools

The affine operator ω_i we use is defined by

$$\omega_i \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_i & b_i & 0 \\ c_i & d_i & 0 \\ 0 & 0 & s_i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \\ o_i \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix} + b \tag{4}$$

where A is the linear operator of ω_i and b is a translation vector.

A spatial transformation defined by the coefficients a_i, b_i, c_i, d_i, e_i, and f_i maps the coordinates (x, y) of the block D_i pixels to the pixels in the block R_i, and a grey scale transformation (in the z direction) defines how the pixel grey values are transformed. This last one is written as

$$v_i(z) = s_i.z + o_i. \tag{5}$$

Fractal image coding requires the definition of a similarity measure between the image block R_i and the transformed image block D_i. For this, the simple L $_2$-metric, also called root mean square (RMS) distortion measure, is preferred because it is easily computed and fits well our visual perception for relatively small block sizes. It is defined as the square root of the sum over the range block R_i of the squared differences of

B pixels' values r_i and d_i , i.e.

$$d(R_i, D_i) = \left[\sum_{i=1,B} (d_i - r_i)^2 \right]^{1/2}. \quad (6)$$

Our implementation is one of the many variants of fractal image-coding schemes (optimized or not) proposed in the literature [10], [14], [17], [24]. The main difference is that we use triangular arrays of pixels to reduce the visible artifacts in the decoded images. The encoding and decoding algorithms will be described in Sections IV-B and IV-C.

B. Delaunay Triangulation

Fractal image coding, more than other methods, removes redundant information by capturing "local affine redundancy," i.e., the local self-similarity present in images. The objective of the partition is thus to provide a maximum number of similar blocks at different resolutions (range and domain blocks). For this reason, the partition must fit the visual image content well. Delaunay triangulation provides a powerful tool for rapidly computing a partition suitable for fractal coding.

Delaunay triangulation of a vertex set S is defined as the unique triangulation with "empty circles," i.e., no vertex lies inside the circumscribing circle of any Delaunay triangle, as follows:

$$DEL(S) = \{(p_i, p_j, p_k) \in S^3,\ B(p_i, p_j, p_k) \cap S \backslash (p_i, p_j, p_k) = \emptyset\} \quad (7)$$

where $B(p_i, p_j, p_k)$ is the circle circumscribed by the three points p_i, p_j, p_k forming a Delaunay triangle [27]. This triangulation is used in many applications because of the property that it is the only one, among all possible triangulations of the same vertex set, that maximizes the minimum interior angle of all the triangles. For fractal image coding, interiors of different triangles are compared. Thus, this local angle maximization is attractive because it avoids numerical problems arising from the presence of thin triangles.

Our method of computing Delaunay triangulation is an incremental approach [7] working by local modification of the diagram by insertion of a new vertex. The image can be seen as a "driver" for the partition. It guides the evolution and location of triangles. The method is based on a split-and-merge approach initialized on a small number of regular triangles computed on a regular grid of vertices [5].

The following details the two steps of split and merge:

Split and Merge algorithm

Initialization:

Construct a lattice S (triangle vertices) on the image support (see Fig. 1)

Split:

Repeat until convergence:

a: calculate the Delaunay triangulation of the set of vertices

b: for each triangle: *if* the triangle is not homogeneous *then* insert a vertex on its barycenter

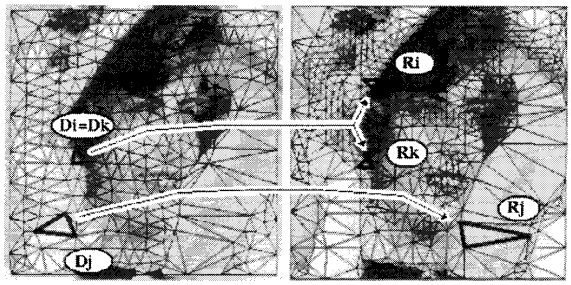


Fig. 4. Encoding step. Each triangle in the partition R is associated to its more similar triangle in the partition D. Each triangle R_i has an antecedent. A triangle D_i can be mapped on more than one triangle R_i, as is shown in Fig. 3. A triangle D_i can be ignored during the encoding step, as is done with the use of vector quantization of the triangles D_i.

Merge:

extract the useless vertices

A vertex $p_i \in S$ is said to be *useless* if all the triangles for which p_i is a vertex are similar with respect to their grey level variance and mean. A triangle is said to be *homogeneous* if its associated parameter of interest (variance) is less than a prefixed threshold. The split process *converges* when all the triangles are homogeneous, or when their size is greater than a given threshold. Examples of such adaptive triangulations are given in Fig. 1(c).

All triangles are recorded in a graph structure. A triangle is coded by its address (position) in the graph. Two parameters are used in this algorithm of split and merge: the grey level variance and the mean value. They are related to the concepts of homogeneity and similarity between adjacent triangles. As we will see further on, two partitions R and D will be extracted from the split-and-merge algorithm. They will differ by the choice of these parameters.

We will now present a classification algorithm performed on the partition D followed by a description of the encoding scheme.

IV. Applications

A. Classification

VQ-based classification and archetype classification methods have been investigated with square blocks for fractal image compression, respectively, in [20] and [6].

To speed up the matching process of the encoding algorithm (see Section IV-B), we propose to reduce the number of triangles D_i resulting from the Delaunay triangulation. This reduction can be done using a *classification* algorithm that permits the construction of a codebook of triangles D_i. The goal of such an algorithm is to design a codebook containing the $\overline{M} = 2^r$ "best representative" triangles D_i. Thus, during the coding process the search for the best domain block D_i corresponding to a given range block R_i is restricted to the domain blocks contained in the previously defined codebook.

The classification algorithm we use is based on the observation that the classical Lloyd algorithm (1957) for scalar

Fig. 5. Classification results. Five triangles are kept in the partition D as shown in the upper left illustration. At the end of the encoding step, each triangle R_i (thin triangles in the five other illustrations) is associated to one of the five triangles (thick triangles) kept in the partition D.

quantization can be generalized for vectors or image blocks [13]. This generalization, well known in the literature as the LBG algorithm, has been developed by Linde, Buzo, and Gray for vector quantization applications in 1980 [21]. This algorithm is based on the Voronoi partitioning of a training set into clusters, and permits the construction of a codebook by taking the centroid of each cluster. For a fixed value of $\overline{M}$, this codebook is optimal for the training set and the chosen classification criterion ρ.

The LBG algorithm is described as follows:

LBG algorithm

Step 0: initialization

Given a training set τ
Compute the centroid $y_1^{(0)}$ of the training set
Construct the codebook $\mathrm{Y}^{(0)} = y_1^{(0)}$

Step 1:

Split the codebook $\mathrm{Y}^{(r)} = \{y_i^{(r)}; \quad i = 1, \cdots, 2^r\}$ by
$\begin{cases} y_{2i-1}^{(r+1)} = y_i^{(r)} - \varepsilon \\ y_{2i}^{(r+1)} = y_i^{(r)} + \varepsilon \end{cases}$

Step 2:

Perform the **Generalized Lloyd algorithm** to generate the improved codebook
$\mathrm{Y}^{(r+1)} = \{y_i^{(r+1)}; \quad i = 1, \cdots, 2^{r+1}\}$
If 2^{r+1} is the expected codebook size then
  Stop
Else
  $r + 1 \rightarrow r$
  goto Step 1
End if

Generalized Lloyd algorithm

Step 1:

Begin with an initial codebook $\mathrm{Y}_1 = \mathrm{Y}^{(r+1)}$
Set $m = 1$
Choose κ

Step 2

Given the codebook Y_m perform the **Lloyd iteration** to generate the improved codebook Y_{m+1}

Step 3

Compute the average distortion $\overline{\rho}$ for Y_{m+1}
If $\overline{\rho} \leq \kappa$ then
  $\mathrm{Y}^{(r+1)} = \mathrm{Y}_{m+1}$
  Stop
Else
  $m + 1 \rightarrow m$
  goto Step 2
End if

Lloyd iteration

(a) Given a codebook $\mathrm{Y}_m = \{y_i\}$, partition the training set into cluster sets C_i such that:
  $C_i = \{x \in \tau / \rho(x, y_i) \leq \rho(x, y_j) \quad \text{for all} \quad j \neq i\}$

(b) For each cluster C_i compute the centroid $\mathrm{cent}\,(C_i)$
The **improved codebook** is given by:
  $\mathrm{Y}_{m+1} = \{y_i = \mathrm{cent}\,(C_i)\}$

Histogram-Based Classification: Two triangles with any orientation and shape, each composed of two distinct regions (a bright and a dark average intensity part), can be mapped onto the same range triangle if the regions are equally separated in the two triangles, as is shown in Fig. 2.

Nevertheless, two such triangles must be considered identical by the classification process.

Furthermore, a triangle with n_d pixels ordered as a n_d-dimensional column vector is greatly influenced by the spatial shape of the triangle. For these two reasons, it is not a good way to compare pixel value vectors.

In order to avoid such problems, we work with the normalized and centered grey level values histogram of each triangle instead of with the pixel values themselves. In fact, the two triangles in the previous example have the same histogram. Furthermore, it is easy to characterize each triangle of the Delaunay triangulation by its corresponding histogram. Thus, the training set for the LBG algorithm is composed of histograms computed for each domain block D_i, and the

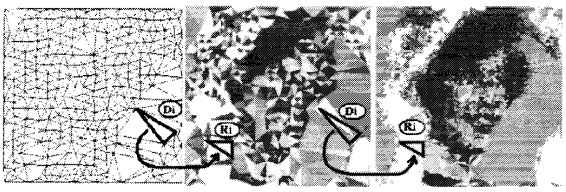


Fig. 6. First two iterations of the decoding step. Left: reconstructed triangulation D on an initial white image; Middle: first iteration of the operator W, applied on the reconstructed triangulation R (blocks D_i in the partition D are mapped onto the blocks R_i of the partition D are mapped onto the blocks R_i of the partition R). Right: second iteration (eight to ten iterations are sufficient for convergence).

classification is performed on these histograms rather than on pixel values.

Different similarity criteria can be used for the classification process. By denoting the normalized histogram of a domain triangle by $f_D(d)$, the usual criterion is the MSE defined by

$$\rho(D_i, D_j) = \sum_k [f_{D_i}(d_k) - f_{D_j}(d_k)]^2$$

and the average similarity can be computed by

$$\overline{\rho} = \sum_{j=1}^{\overline{M}} \min_{Dj \in \mathrm{Y}_m} \rho(D_i, D_j).$$

Note that for fractal coding applications, a modified version of the Lloyd algorithm was developed. Here, the best representative element of a cluster is not its centroid but the training vector nearest to the centroid according to the criterion ρ [see step (b) of the modified Lloyd iteration].

Modified Lloyd iteration

(a) Given a codebook $\mathrm{Y}_m = \{y_i\}$, partition the training set into cluster sets C_i such that:
$C_i = \{x \in \tau / \rho(x, y_i) \leq \rho(x, y_j) \quad \text{for all} \quad j \neq i\}$

(b) For each cluster C_i compute the centroid $\mathrm{cent}\,(C_i)$
The **improved codebook** is given by
$\mathrm{Y}_{m+1} = \{y_i = \arg \min_{x \in C_i} \rho[x, \mathrm{cent}\,(C_i)]\}$

To illustrate this new algorithm, a classification result is shown in Fig. 3. The aim of the classification process is to keep the most significant grey level histograms. As observed in Fig. 3, the two codebooks are composed of a limited number of blocks regularly distributed on the image support, and provide representative sets of blocks D_i depending on image content.

B. Encoding

The encoding algorithm starts with an adaptive Delaunay triangulation R providing the nonoverlapping blocks R_i. Another Delaunay triangulation D provides the blocks D_i from which it is possible to keep the $\overline{M}$ "best representative triangles" (see Section IV-A). Fig. 4 illustrates the matching process of the encoding algorithm, which is described below.

For each triangle R_i of the partition R, find the best triangle in the set of the $\overline{M}$ triangles D_j ($j = 1, \cdots, \overline{M}$), minimizing the RMS error $d[R_i, w_j(D_j)]$ and considering

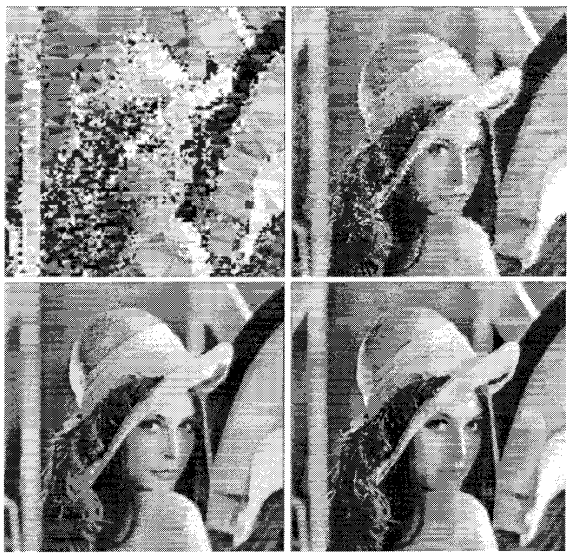


Fig. 7. Decoding of the Lena image (512×512) starting with a black image. Compression ratio = 15.2 : 1, PSNR = 31.95 dB. Top: First and second iteration. Bottom: tenth iteration (reconstructed image) and original image.

the best isometry $\mathrm{k_j}$ ($1 \leq \mathrm{k_j} \leq 6$) for mapping the triangle D_j with n_d pixels onto the triangle R_i with n_r pixels. The index i is assigned to such a triangle in the partition D.

Then associate the following with R_i:

- the block D_i address in the graph of the partition D
- the grey tone scaling coefficient s_i
- the grey tone offset coefficient o_i
- the isometry index $\mathrm{k_i}$

The encoding algorithm we propose is based on triangular blocks of undetermined sizes and shapes. If $n_d > n_r$, the comparison between blocks is done by pure subsampling of D_i for simplicity reasons. Each pixel in R_i is compared (in the RMS error sense) to the pixel in the correct position in the block D_i under the spatial affine transformation.

It has been observed that authorization of n_d to be less than n_r can improve visual quality of reconstructed images as well as their peak signal-to-noise ratio (PSNR). In practice, the partition D is composed, on average, of larger blocks than the blocks in the partition R. As a consequence, we allow the mapping of a block D_i onto a larger similar range block. In this way, the PSNR of the reconstructed image can increase by 0.4 dB for natural images without any strict spatial contraction constraints. The only constraint we impose on the method is to have R_i not equal to D_i.

Fig. 5 illustrates the different associations at the end of the coding step considering only five triangles in the partition D.

An important problem in fractal encoding concerns the property of the operator W defined in (2) to be contractive or eventually contractive. This property is essential for the iterative decoding process to converge to an attractor lying close to the original image.

A classical solution ensuring convergence in the sense of the $\mathrm{L_2}$ metric is to impose $|s_i| < 1, \quad \forall 1 \leq i \leq N$ where N

Fig. 8. Decoding results on Woman image 256 × 256, 8 b/pixel. From top to bottom, left to right: Reconstructed images encoded with codebook sizes, respectively, of (a) 32; (b) 64; (c) 128; and (d) 256. (compression ratios are, respectively, 28.5, 26.3, 22.8, and 18.3); (e) reconstructed image without vector quantization of the triangles D_i (compression ratio is equal to 18.3); (f) original image.

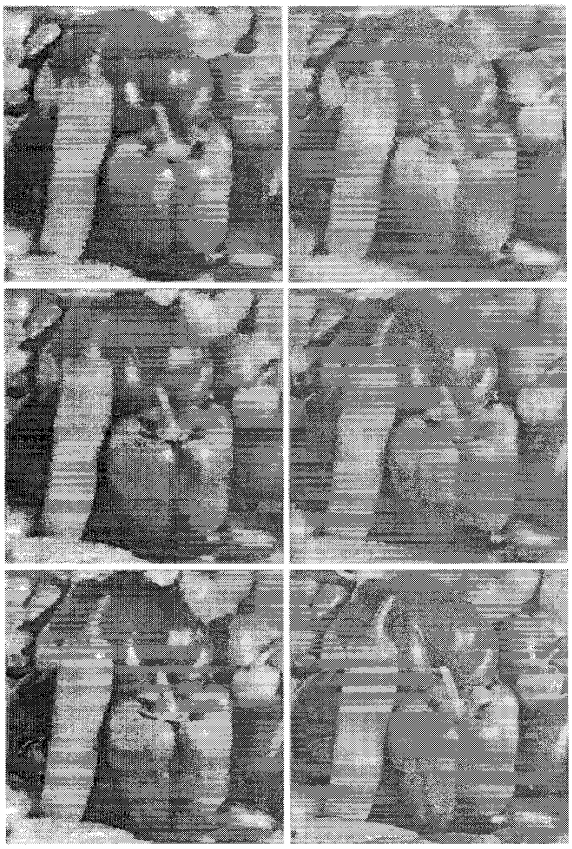


Fig. 9. Decoding results on the image Pepper 256 × 256, 8 b/pixel. From top to bottom, left to right: (a), (b), (c), (d) Reconstructed images encoded with codebook sizes, respectively, of (a) 32; (b) 64; (c) 128; and 256 (compression ratios are, respectively, 24.49, 22.8, 20.15, and 16.3); (e) reconstructed image without vector quantization of the triangles D_i. The compression ratio is equal to 15.6; (f) original image.

is the number of range blocks and s_i is the scale coefficient of the affine transform w_i.

However, this severe constraint is sufficient but not necessary, and therefore can impair reconstructed image quality. Jacobs *et al.* [16] showed by experimental studies that $s_{i\,\max} < 1.5$ improves the PSNR-versus-compression results. They did their analysis on the classical Jacquin's scheme based on regular square blocks. Different theoretical studies have then been made on the contractivity of W for different classes of simple mappings based on regular square blocks [15], [22] and quadtree partitions [24]. The aim of their works is to better control contractivity of the operator W in order to leave more freedom in the choice of scale coefficients so that decoder complexity can be reduced and reconstructed image quality increased. The control considers different norms or the spectral radius of the linear part of W.

In our case, based on irregularly shape triangles, experimental results on a large number of images confirm that the severe constraint $|s_i| < 1$ ensures the contractivity of the operator W. Thus, we require that all grey level transformations be contractive under this constraint, but we do not require that all spatial transformations be contractive.

C. Decoding

Starting with two known lattices of points in the image support, the two partitions R and D are reconstructed (see Section IV-D).

The mapping W is then iterated on an arbitrary image until the resulting image converges. During an iteration, a mapping is done on each triangle R_i after transformation of the pixel values contained in the corresponding triangle D_i. Fig. 6 shows the first two iterations of W, initialized on a white image. The two partitions are fixed during the iteration process and details inside each triangle R_i appear and get smaller as one goes along the iterations.

The numerical evaluation of the coder's performance is achieved by computing the PSNR between the original image $x(m, n)$ and the decoded image $\hat{x}(m, n)$. It is given by

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{E\{[x(m, n) - \hat{x}(m, n).]^2\}} \right)$$

Fig. 7 shows the iterative decoding of the Lena image computed on a Delaunay partition with a compression ratio equal to 15.2 : 1 and a 31.95 dB PSNR.

D. Compression Aspects

In the case of fractal image compression, an operator W with its associated partitions uniquely encodes one image. The compression ratio is thus related to the storage required for the operator.

The information necessary to describe an operator consists, for each ω_i (4) of

- the address of the domain triangle D_i in the graph of the partition D
- the s_i and o_i coefficients (scale and offset)
- the isometry index

At the decoding phase, the two partitions R and D are reconstructed, starting from two known regular triangulations. Their computations only require the knowledge of the split-and-merge processes memorized during the encoding phase. The reconstruction algorithm simply makes a series of binary decisions.

During the split steps, the division of a triangle is coded with one bit equal to "1," a nondivision with a "0," and finally the extraction of a vertex during the merge step is coded with a "1," a nonextraction with a "0." This chain of bits, specifying an adaptive Delaunay triangulation, represents a small number of overhead bits. For example, the split-and-merge process in Fig. 1 is coded with $X = 1564$ bits without entropy coding.

The scalar quantization of coefficients s_i and o_i is done during the RMS error minimizations of the encoding phase with, respectively, n_s and n_o bits. We found that $n_s = 6$ and $n_o = 6$ is the best compromise choice in order to sufficiently compress classical images with a fairly good PSNR.

The isometry indices are coded with $n_i = 3$ bits because of the six possible symmetries for mapping one triangle onto another.

The compression ratio formula (for a $2^n * 2^n$ 8–b/pixel image) can be written differently depending on whether we use the entire triangulation D during the encoding process (a) or not (b), as follows:

$$(a): \quad T_c = \frac{2^n * 2^n * 8}{N * (n_a + n_i + n_o + n_s) + X}$$

where X is the number of overhead bits to code the partitions D and R, $n_a = \lceil \log_2 (M) \rceil$ is the number of bits to code the address of a triangle D_i, and N is the number of triangles R_i in the partition R.

$$(b): \quad T_c = \frac{2^n * 2^n * 8}{N * (n_i + n_o + n_s) + \overline{M}(3 * 2 * n) + X^\bullet}$$

where $\overline{M}$ is the number of "best representative" triangles of the partition R ($\overline{M}$ is the size of the codebook, see Section IV-A), and $X^\bullet$ is the number of overhead bits to code the partitions D only. Classification of the triangles D_i is thus useful in terms of compression if

$$\overline{M}(2 * 3 * n) + X^\bullet < N(\lceil \log_2 M \rceil) + X.$$

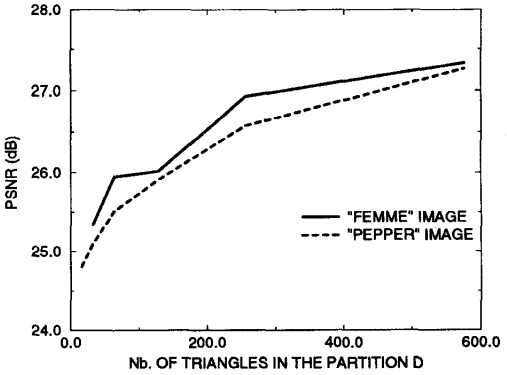


Fig. 10. PSNR of the decoded image versus the number of triangles D_i.

V. Experimental Results

Coding simulations have been performed on the two images Woman and Pepper at $256 * 256$ and 8-b/pixel (see Figs. 8 and 9). The different codebook sizes that have been used are 32, 64, 128, 256, and finally the entire partition D composed of 576 triangles. The number 576 is important because of the fact that triangles in partition D must be larger than triangles in partition R. Decoded images are close to the original image. Artifacts are visible on contrasted contours and finely textured areas due to the restricted number of triangles R_i. Better visual results are possible with a finer triangulation, to the detriment of the compression ratio. Classification of the triangles D_i in order to reduce encoding complexity does not damage decoded image quality too perceptibly. Fig. 10 shows PSNR of reconstructed images versus different codebook sizes (16, 32, 64, 128, 256, and finally the entire partition D composed of 576 triangles).

VI. Conclusions and Perspectives

A new partitioning scheme based on Delaunay triangles has been investigated for fractal image compression. Such a partition has been shown to be attractive because it provides a reduced number of blocks compared with square-based partitions and, thus, minimizes the number of mappings. The triangulation we propose, computed on a set of points distributed on the image support, is fully flexible and efficiently coded. Moreover, a classification scheme based on vector quantization has been used in order to reduce the number of blocks in the domain partition. This results in reduced encoding complexity while preserving good decoding quality at rates between 0.25 and 0.5 depending on the nature of the image.

In a future work, we propose to improve our classification scheme based on the comparison of the grey level histograms (MSE). The aim will be to find a grey level shape criterion, invariant under spatial affine transformations of the blocks.

Current work also involves the improvement of the partitioning scheme, e.g., how to enable it to provide a minimum number of blocks in order to decrease the bit rate while maintaining PSNR-compression performances. We propose to

merge neighboring stretched triangles into quadrilaterals. The triangulation is better suited than quadtrees or rectangles for the modeling of the piecewise self-similarity, at least for the diagonal image edges. A way to benefit from this advantage is to position initial points regularly onto the edges before the construction of the partition so that the triangle edges fit the image edges. We are currently investigating this method.

REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing,* vol. 1, no. 2, pp. 205–220, 1992.

[2] M. Barlaud, P. Solé, T. Gaidon, M. Antonini, and P. Mathieu, "Pyramidal lattice vector quantization for multiscale image coding," *IEEE Trans. Image Processing,* vol. 3, no. 4, pp. 367–381, 1994.

[3] M. Barnsley and L. Hurd, *Fractal Image Compression.* Wellesley, MA: AK Peters, 1993.

[4] T. Bedford, F. M. Dekking, M. Breeuwer, M. S. Keane, and D. Van Schooneveld, "Fractal coding of monochrome images," *Signal Processing: Image Commun.,* vol. 6, pp. 405–419, 1994.

[5] E. Bertin, F. Parazza, and J.-M. Chassery, "Segmentation and measurement based on 3D Voronoi diagram: pplication to confocal microscopy," *Computerized Med. Imaging and Graphics,* vol. 17, no. 3, pp. 175–182, 1993.

[6] R. D. Boss and E. W. Jacobs, "Archetype classification in an iterated transformation image compression algorithm," in *Fractal Image Compression—Theory and Application,* Y. Fisher, Ed. New York: Springer-Verlag, 1995.

[7] A. Bowyer, "Computing Dirichlet tessellations," *Computer J.,* vol. 24, no. 2, pp. 162–166, 1981.

[8] R. J. Clarke, *Transform Coding of Images.* Orlando, FL: Academic, 1985.

[9] F. Davoine and J.-M. Chassery, "Adaptive Delaunay triangulation for attractor image coding," in *12th Int. Conf. Patt. Recog.,* Jerusalem, Israel, Oct. 9–13, 1994, pp. 801–803.

[10] F. Dudbridge, "Least-squares block coding by fractal functions," in *Fractal Image Compression—Theory and Application,* Y. Fisher, Ed. New York: Springer-Verlag, 1995.

[11] Y. Fisher, Ed., *Fractal Image Compression—Theory and Application.* New York: Springer-Verlag, 1995.

[12] Y. Fisher and S. Menlove, "Fractal encoding with HV partitions," in *Fractal Image Compression—Theory and Application,* Y. Fisher, Ed. New York: Springer-Verlag, 1995.

[13] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Boston: Kluwer, 1992.

[14] M. Gharavi-Alkhansari and T. S. Huang, "A fractal-based image block-coding algorithm," in *Proc. ICASSP,* vol. 5, 1993, pp. 345–348.

[15] B. Hürtgen and T. Hain, "On the convergence of fractal transforms," in *Proc. ICASSP,* vol. 5, 1994, pp. 561–564.

[16] E. W. Jacobs, Y. Fisher, and R. D. Boss, "Image compression: A study of the iterated transform method," *Signal Processing,* vol. 29, pp. 251–263, 1992.

[17] A. E. Jacquin, "A fractal theory of iterated Markov operators with applications to digital image coding," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA, Aug. 1989.

[18] ——, "Fractal image coding: A review," in *Proc. IEEE,* vol. 81, no. 10, pp. 1451–1465. 1993.

[19] J. C. Kieffer, "A survey of the theory of source coding with a fidelity criterion," *IEEE Trans. Inform. Theory,* vol. 39, no. 5, pp. 1473–1490, 1993.

[20] S. Lepsøy and G. E. Øien, "Fast attractor image encoding by adaptive codebook clustering," in *Fractal Image Compression—Theory and Application,* Y. Fisher, Ed. New York: Springer-Verlag, 1995.

[21] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.,* vol. COM-28, no. 1, pp. 84–95, 1980.

[22] L. Lundheim, "A discrete framework for fractal signal modeling," in *Fractal Image Compression—Theory and Application,* Y. Fisher, Ed. New York: Springer-Verlag, 1995.

[23] M. Novak, "Attractor coding of images," in *Proc. Picture Coding Symp.,* Lausanne, Switzerland, 1993.

[24] G. E. Øien and S. Lepsøy, "Fractal-based image coding with fast decoder convergence," *Signal Processing,* vol. 40, pp. 105–117, 1994.

[25] D. Saupe and R. Hamzaoui, "A review of the fractal image compression literature," *Comput. Graphics,* vol. 28, no. 4, pp. 268–276, 1994.

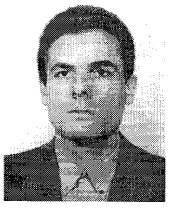
[26] J. Vaisey and A. Gersho, "Image compression with variable block size segmentation," *IEEE Trans. Signal Processing,* vol. 40, no. 8, pp. 2040–2060, 1992.

[27] D. F. Watson, "Computing the n-dimensional Delaunay tessellation with application to Voronoi polytopes," *Computer J.,* vol. 24, no. 2, pp. 167–172, 1981.

Franck Davoine was born in France on April 16, 1967. He received the DEA degree in signal, image, and speech processing, in 1992, from the National Polytechnic Institute of Grenoble, France.

Since 1992, he has been working toward the Ph.D. degree at the TIMC-IMAG Laboratory in Grenoble.

His research interests include image and signal processing, still image, and image sequence coding.



Marc Antonini was born in Nice, France, on August 29, 1965. He received the Ph.D. degree in electrical engineering from the University of Nice–Sophia Antipolis, France, in 1991.

He is currently working with CNRS .

His research interests include multidimensional image processing, wavelet analysis, lattice vector quantization, information theory, and image coding.

Jean-Marc Chassery was born in France on April 1949. He received the Doctorat d'état degree in 1984.

He is currently Director of Research at CNRS and manages the image group INFODIS at TIMC-IMAG Laboratory in Grenoble, France.

His primary research interests are digital geometry and image processing including multiscale and adaptive spatial partitioning. Such models have been applied in image reconstruction, segmentation, and compression.

Michel Barlaud (M'88), for photograph and biography, please see p. 199.

Annexe G

Optimal Decoder for Block-Transform Based Video Coders

IEEE Transactions on Multimedia vol.5, no.2, juin 2003

J'ai co-écrit cet article avec Joël Jung et Michel Barlaud.

Résumé Dans cet article nous introduisons un algorithme de décodage de vidéos comprimées au moyen de codeurs tels que M-JPEG, H26x ou MPEG. L'algorithme proposé prend en compte à la fois les artefacts liés à la compression mais aussi ceux introduits par la transmission, l'acquisition ou le stockage. La nouveauté de notre méthode réside dans le traitement simultané de tout ces problèmes en utilisant une approche variationnelle. Le décodage réalise une segmentation spatio-temporelle, et un traitement différent est effectué sur les objets en mouvement par rapport au fond fixe. L'information contenue par le train binaire reçu est aussi prise en compte (pas de quantification et vecteurs mouvement). De nombreux résultats expérimentaux montrent l'efficacité de la méthode proposée par rapport au post-traitement développé pour MPEG-4.

Optimal Decoder for Block-Transform Based Video Coders

Joël Jung, Marc Antonini, and Michel Barlaud, *Fellow, IEEE*

***Abstract*—In this paper, we introduce a new decoding algorithm for DCT-based video encoders, such as Motion JPEG (M-JPEG), H26x, or MPEG. This algorithm considers not only the compression artifacts but also the ones due to transmission, acquisition or storage of the video. The novelty of our approach is to jointly tackle these two problems, using a variational approach. The resulting decoder is object-based, allowing independent and adaptive processing of objects and backgrounds, and considers available information provided by the bitstream, such as quantization steps, and motion vectors. Several experiments demonstrate the efficiency of the proposed method. Objective and subjective quality assessment methods are used to evaluate the improvement upon standard algorithms, such as the deblocking and deringing filters included in MPEG-4 postprocessing.**

***Index Terms*—Blocking effect, cell loss, dropout, M-JPEG, MPEG, object-based decoding, optimization, variational approach.**

I. Introduction

A. Need for Efficient Video Decoders

WITH THE recent advent of digital technologies, and the ever-increasing need for speed and storage, compression is more and more widespread. Compression algorithms are integrated in most recent devices: JPEG for digital cameras and Motion-JPEG (M-JPEG) for digital camcorders. MPEG-2 is used for digital TV and DVDs, H263 in videophones. MPEG-4 is used in "best-effort" applications such as video streaming on the Internet, and soon will be integrated in mobiles and PDAs, sharing the wireless field with the emerging H.264 standard [17]. All these algorithms are block-transform based; they consequently produce visually annoying compression artifacts, such as the well-known blocking effect. Moreover, most of these applications suffer from transmission over noisy channels leading also to some artifacts. Last but not least, storage and playback of the video can also introduce some artifacts. For instance, tape damage or head clogging can produce block loss, mosaic effects on small areas of pixels, blotches, banding, etc. In the rest of the paper, the term "dropout" will stand for any defect introduced by the transmission chain, except compression artifacts.

Manuscript received April 3, 2000; revised April 17, 2002. Theoretical parts of this work (Sections II and III) were performed during J. Jung's Ph.D. thesis at I3S laboratory, and experimental results (Section IV) at Philips Research France. The associate editor coordinating the review of this paper and approving it for publication was D. Faouzi Kossentini.

J. Jung is with Philips Research France, 92156 Suresnes, France (e-mail: joel.jung@philips.com).

M. Antonini and M. Barlaud are with the I3S Laboratory, CNRS UMR-6070, University of Nice-Sophia Antipolis, 06903 Sophia Antipolis, France (e-mail: am@i3s.unice.fr; barlaud@i3s.unice.fr).

Digital Object Identifier 10.1109/TMM.2003.811616

In all the applications above, image quality is a key issue. For "best effort" applications, it is up to the manufacturer to include algorithm that will improve the visual quality as much as possible. Better video quality can be obtained if and only if, each kind of artifact is jointly processed. So, when using the standard compression algorithm at low bitrates, it may be necessary to improve the quality at the decoder side. Similarly, equivalent visual quality can be achieved with lower bitrates using such a post-processing algorithm.

Very few approaches have been proposed to tackle both problems simultaneously. Consequently, we propose in this paper a new decoding method, adapted to DCT-based compression algorithms, that will deal simultaneously with compression and transmission artifacts, and dropouts.

B. Review of Existing Techniques

The proposed decoding algorithm deals with blocking effects and dropouts, using an object-based approach. We start with a survey of existing restoration techniques.

1) Removal of Blocking Artifacts: Very different approaches have been proposed to reduce blocking effects, particularly in still images. Post-processing methods were initiated in [34] by Gersho, who applied a nonlinear space-variant filtering to block coding methods. More recently, methods based on global and local filtering [29], wavelet thresholding [15], and criterion minimization [45] have also been suggested. Advanced methods, adopting a global approach for decoding, with the minimization of a criterion measuring blocking effects [31], or wavelet thresholding and projection [21] avoid the classic problems (such as edge smoothing) of video post-processing methods. Nevertheless, video sequence processing requires more than an independent processing of consecutive images. The temporal characteristics of blocking effects must be considered in the decoding scheme. Consequently, these methods are not suitable for video sequences. Actually, there are less approaches dedicated to DCT-based video sequences that we might expect. In [46], Galatsanos applies a decoding method for MPEG, based on the theory of projections onto convex sets. In [42], and [6], DCT domain algorithms are applied: coefficients are adjusted to remove discontinuities at block corners. In [8], local slopes are evaluated and modified to achieve global smoothness, a noniterative post-processing technique is applied in [32], and a Wiener-filter based restoration method is proposed in [44]. Most of the methods developed for blocking effects reduction in sequences have the main drawback of being post-processing approaches: they do not consider the information provided by the bitstream, such as quantization values, motion vectors, type of macroblock (I,P) etc. Due to this lack of information, they generally do not solve the problem of

luminance variation inside the DCT-blocks, and local filtering applied to block corners causes smoothing that tends to reduce the details in the sequence. More recently, algorithms like MPEG-4 informative deblocking [33] applies a detection of the artifacts in the spatial domain, while the correction is frequential and uses 1×4 DCTs. Filtering operations are performed along the 8×8 block edges. Both luminance and chrominance data are filtered. Two different filtering modes are used depending on the strength of the artifact: the "DC offset mode," or the "Default mode." In the "Default mode," a signal adaptive smoothing scheme is applied by differentiating image details at the block discontinuities using the frequency information of neighbor pixel arrays. The filtering scheme in "default mode" is executed by replacing the immediate boundary pixel values. In the "DC offset mode," a stronger filtering is applied, due to the DC offset, for all the block boundaries first along the horizontal edges followed by the vertical edges. If a pixel value is changed by the previous filtering operation, the updated pixel value is used for the next filtering. More information on this deblocking filtering are available in [16, Annex F].

2) Removal of Transmission Artifacts and Dropouts: Transmission artifacts are usually removed using error control and concealment techniques. These techniques are numerous. We recommend to read the review [41] written by Wang and Zhu for more information. Nevertheless, let us detail some methods of interest: spatial interpolation is very simple and low cost, but often results in blurring [13]. Motion compensated temporal prediction gives generally good results if motion vectors are available [10]. Other methods are using POCS [37], or bayesian approaches [36]. Kokaram suggests a detection method in [26] and a spatial interpolation method in [27] for missing data. In [14], a local analysis of spatio–temporal anisotropic gray-level continuity for film blotch removal is proposed, and in [30] a method for blotch and scratch detection in image sequences is developed. Unfortunately, these dropouts detection and interpolation methods are all post-processing approaches, and are often dedicated to a single kind of artifact.

3) Object-Based Approaches for Restoration: Object-based approaches require efficient motion segmentation. This problem has recently been widely investigated, and very different kinds of approaches have been proposed, for instance, let us mention methods for object segmentation [18], [19] and for object segmentation with tracking [11], [20]. Many methods exploit the temporal and spatial information in the video sequence to differentiate foreground from background: in [40], an automatic spatio–temporal and object based segmentation algorithm is proposed and in [9] a real time object detection one. Noise removal and simultaneous displacement estimation were proposed by Katsaggelos in [3]. The same problem is developed in [28]: Kornprobst deals with the problem of segmentation and restoration in a coupled way using an optimization approach.

Unfortunately, none of these methods considers both blocking effects removal and segmentation. Consequently, on the one hand, results of segmentation are strongly affected by these blocking artifacts, while on the other hand, methods applied for blocking effects reduction cannot benefit from an efficient motion segmentation, to apply separated and adapted processing for moving objects and for the background of the sequence.

C. Main Contribution and Paper Organization

Coding artifact removal is a tricky problem: one might believe that applying a process such as the one used in MPEG-4 VM8 is sufficient. In fact, subjective experiments show that viewers are sensitive to improvements, according to regions of interests and masking properties. The key-issue, in proposing an object-based decoder, is to be able to deal with these properties, and so, smooth/clean more or less backgrounds and objects, according to their own properties. This is a major contribution of our approach.

Numerous standardized decoders already exist on the market. Our motivation for providing an "advanced decoder," where processing is included inside the decoder is twofold.

- First, we take benefit from the information provided by the encoder, that is available after the decoder for standard post-processing methods. The information are the quantizations steps, the motion vectors, the dequantized values in the DCT domain, and the type of the macroblocks (I or P).
- Second, the joint error-concealment/compression artifact reduction approach is a key-point enabling better results than when applying two independent tasks.

Some recent studies address the interaction between compression artifacts reduction and error concealment methods. But, to our knowledge, the problem of jointly reducing blocking effects and dropouts by a global decoding approach has not yet been considered. Our object-oriented method is based on five steps (see Fig. 1).

1) The background of the scene is estimated and blocking effects are removed, while an accurate representation of the moving parts of each image is computed.
2) Each object is spatially isolated from the others and tracked, in order to be processed separately. A database of spatial and temporal characteristics of the objects is built.
3) Objects corresponding to dropouts on the background are removed according to spatio–temporal assumptions.
4) Quantization noise and blocking effects are removed on each object independently.
5) Processed objects and backgrounds are gathered, to build the final sequence.

The paper is organized as follows. In Sections II and III, the decoding method adapted to blocking effects and dropouts removal is proposed. Section IV is devoted to experimental results. We show that our method achieves an enhanced decoding: it increases significantly the visual quality of the sequence both objectively and subjectively.

II. Reduction of Compression Artifacts

A. Notation

In the rest of the paper, we suppose that the k^{th} image $\widetilde{f_k}$ of the sequence corresponds to a projection of the moving objects c_k onto the background f_k with $c_k \in [0;1]$ ($c_k(x,y) = 1$ if the pixel (x, y) belongs to the background and 0 otherwise [28]).

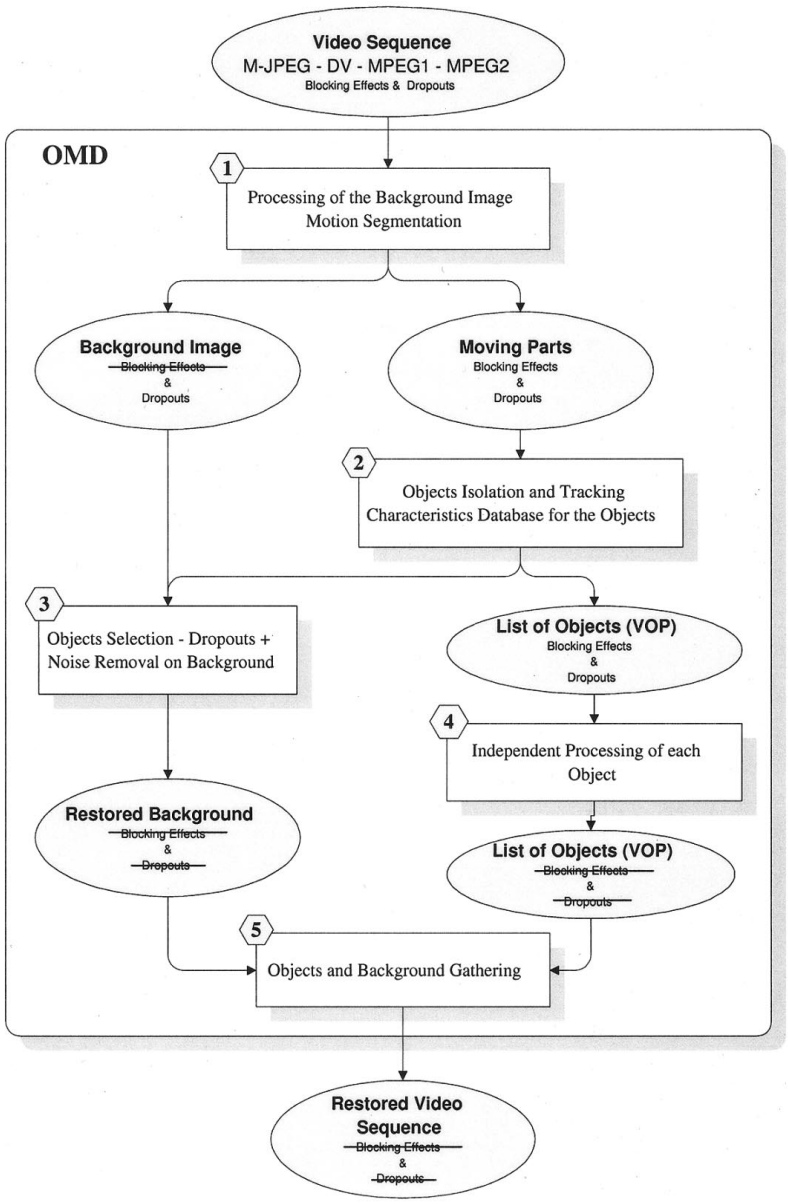


Fig. 1. Overview of the decoding method.

Let us define p_k^*, the k^{th} image of the coded/decoded M-JPEG sequence in the DCT transform domain, by

$$p_k^* = Q^{-1}\left(Q\left(D\tilde{f}_k\right)\right),\ k = 1, 2, \ldots, N \tag{1}$$

where N is the number of images contained by the sequence and D is the DCT operator. The quantizer Q is defined as an additive noise model [35] such that

$$Q\left(D\tilde{f}_k\right) = D\tilde{f}_k + \epsilon,\ k = 1, 2, \ldots, N \tag{2}$$

with ϵ the quantization noise.

B. Spatio–Temporal Segmentation

1) Introduction: This section describes how an improved representation of the background and a map of the moving parts of the sequence are obtained. It corresponds to part 1 of Fig. 1.

The basic idea of several spatio–temporal segmentation methods is to perform a temporal average of the video in order to separate the background and the moving objects.

TABLE I
BEHAVIOR OF CRITERION (3): VALUE OF THE ESTIMATED BACKGROUND PIXEL f_k ACCORDING TO THE PIXEL IN THE OBSERVED IMAGE

Kind of pixel in the observed image $D^{-1}p_k^*$	**Value for f_k**
background ($c_k = 1$)	$f_k \leftarrow D^{-1}p_k^*$
object ($c_k = 0$)	$f_k \leftarrow m_k$
else	f_k is a function of $D^{-1}p_k^*$ and m_k

The principal advantage of this kind of processing is its low computational cost but the drawbacks are numerous, especially

- it is necessary to have an important number of images (generally two or three seconds of video) in order to perform the processing;
- the background needs to be static;
- a scene-cut is disastrous;
- these methods are not robust when images are corrupted by blocking artifacts introduced by compression.

The goal of our approach is clearly to suppress the drawbacks encountered in the classical post-processing methods and to allow

1) obtaining different background f_k for each image k of the sequence, resulting in a more realistic reconstructed sequence;
2) "on the fly" processing, i.e., progressive processing. Then, it is not necessary to know the entire video sequence to perform the spatio–temporal segmentation process for image k;
3) robust segmentation when the camera moves or when there is a scene-cut;
4) restoration efficiency even at low bitrate.

2) Proposed Criterion: The problem consists in finding f_k the estimated background for the frame $k (k = 1, \ldots, N)$ [22], [23]. The main idea of our method is based on the observation that moving objects are characterized by strong temporal discontinuities. Then, to separate the moving objects from the background we perform a temporal smoothing on each pixel that presents strong temporal derivative. This problem is expressed as an inverse problem by introducing the following functional given formula (3):

$$J_1(f_k, c_k) = \underbrace{\int_{\Omega} c_k^2 \left(f_k - D^{-1}p_k^*\right)^2 d\Omega}_{term\ A} + \underbrace{\alpha_p \int_{\Omega} (c_k - 1)^2 \|\nabla_T f_k\|^2 d\Omega}_{term\ B} \quad (3)$$

where D^{-1} is the inverse DCT operator, Ω is the support of the image in the spatial domain and ∇_T stands for the temporal derivative between current image f_k and a temporal average image noted m_k. It is defined by:

$$\nabla_T f_k = f_k - m_k. \quad (4)$$

Note that α_p is a coefficient which permits to control the convergence of the criterion toward $D^{-1}p_k^*$ or m_k. As described in Table I, criterion (3) allows us to

1) extract the background f_k by using the information contained in the observed image p_k^* (term A);
2) replace inside f_k the objects c_k by an adaptive temporal average noted m_k (term B).

For each frame, the estimated background images are given by

$$f_k^* = \arg \min_{f_k, c_k} \left(J_1(f_k, c_k)\right).$$

The optimal solution of this minimization problem is obtained when

$$\frac{\partial J_1(f_k, c_k)}{\partial f_k} = 0 \text{ and } \frac{\partial J_1(f_k, c_k)}{\partial c_k} = 0$$

equivalent to

$$\begin{cases} \left(c_k^2 + \alpha_p (c_k - 1)^2\right) f_k = c_k^2 D^{-1} p_k^* + \alpha_p (c_k - 1)^2 m_k & (a) \\ \text{and } c_k = \frac{\alpha_p (f_k - m_k)^2}{\alpha_p (f_k - m_k)^2 + \left(f_k - D^{-1}p_k^*\right)^2}. & (b) \end{cases} \quad (5)$$

The solution of system (5) is then given by the following algorithm.

$c_k = 1$
Repeat
  Step 1: Solve equation (a) of system (5)
  in f_k with c_k fixed
  Step 2: Solve equation (b) of system (5)
  in c_k with f_k fixed
Until Convergence of f_k

During the first pass, the background is estimated with c_k initialized to 1. Solving (5a) provides the new estimate f_k and the c_k sequence with values in the interval [0;1]. Then, in the next iteration the temporal average m_k is weighted by these values of c_k, not to take into account objects for the computing of the average that produces the estimated background.

The image m_k can be defined as a weighted average of the previous images $f_i (i = 1, \ldots, k-1)$, as given in

$$m_k = \frac{1}{\sum_{i=1}^{k-1}} \sum_{i=1}^{k-1} c_i f_i, \quad for \sum_{i=1}^{k-1} c_i \neq 0. \quad (6)$$

The computation of this weighted average, which takes into account the object position, permits to accelerate the convergence of the minimization problem toward the solution f_k. As we can see in the example presented in Fig. 2, this iterative process is powerful: in the case of a sequence with a static background, it totally removes the moving objects from the background.

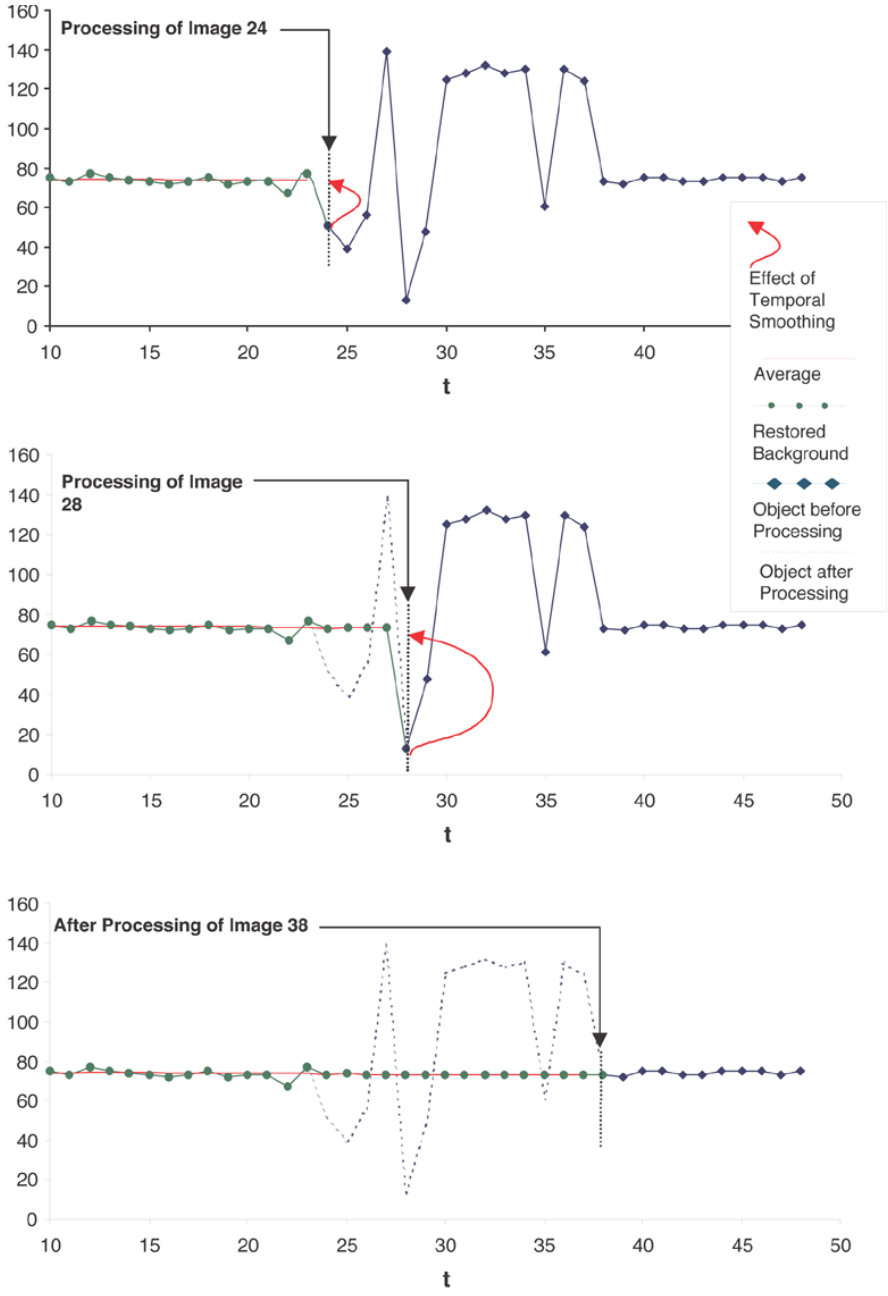


Fig. 2. Temporal evolution of a pixel intensity for a typical video sequence of 50 frames. We can see the effect of the temporal smoothing when the pixel belongs to a moving object.

C. Processing of the Background

1) Noise and Ringing Removal: The criterion proposed in (3) performs the spatio–temporal segmentation but does not remove the artifacts due to DCT coding. Thus, we must introduce in this criterion regularization constraints containing a priori assumptions on the solution to obtain.

The first constraint on the solution we introduce is

$$J_2(f_k) = \lambda_1^2 \int_{\Omega} \varphi\left(\|\nabla f_k\|\right) d\Omega \tag{7}$$

where λ_1 is a Lagrangian parameter and φ a potential function described in [7]. This function has special properties such that

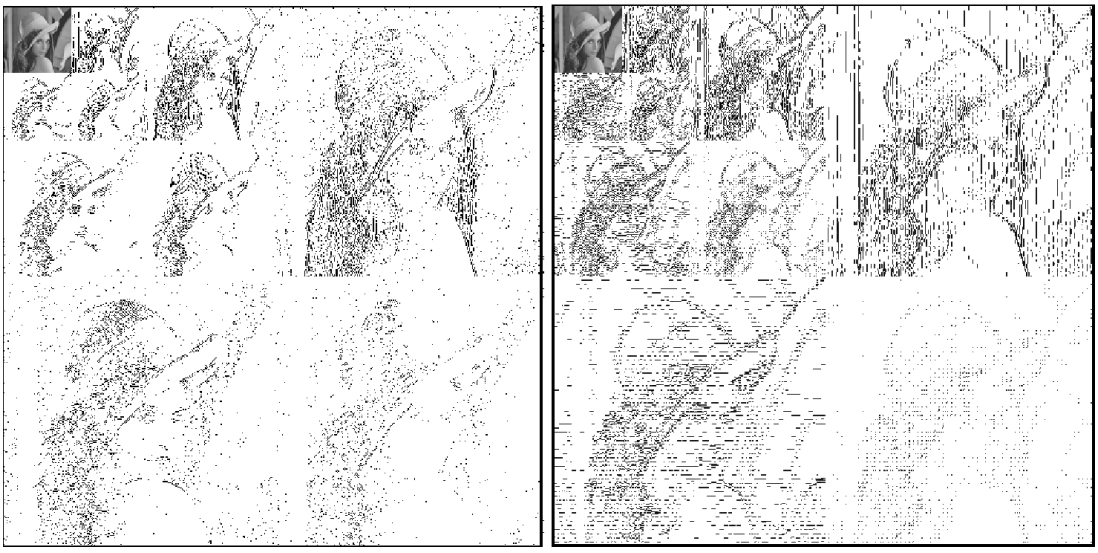


Fig. 3. Wavelet decomposition on 3 levels of the image "Lena" (512 × 512 pixels - 8 bpp). On the left hand side, decomposition of the original image. On the right hand side, decomposition of Lena coded/decoded by JPEG (bitrate of 0.15 bpp).

large gradient associated to edges can be preserved and, at the same time, homogeneous areas are smoothed isotropically. The choice of the potential function as an edge preserving regularization function is addressed in [39]. Furthermore, the potential function needs to satisfy the following properties [7]:

- $0 < \lim_{u \to 0} (\varphi'(u))/(2u) < \infty$: isotropic smoothing in homogeneous areas;
- $\lim_{u \to \infty} (\varphi'(u))/(2u) = 0$: preservation of edges;
- $(\varphi'(u))/(2u)$ is strictly decreasing in order to avoid instabilities;

where $\varphi'(u)$ stands for the derivative of $\varphi(u)$.

2) Removal of Blocking Effects: In our approach, we tackle the problem of DCT blocking effects by reducing temporal variation of luminance inside the blocks with (3). Moreover, we can also work on block edges. In fact, image coding using DCT generates blocking effects which have typical characteristics in the wavelet (or space-frequency) domain [24] (see Fig. 3): wavelet coefficients [1] of block edges appear in horizontal and vertical high frequencies of wavelet coefficient sub-images [21]. Thus, significant wavelet energy results directly from DCT-block edges. The basic idea is to reduce the amplitude of the corresponding wavelet coefficients, while preserving the others. This choice is built on two ideas. The coefficients to remove

- are well located spatially in the wavelet domain (their position depends on the size of the DCT blocks);
- consist of isolated horizontal and vertical lines in high frequencies (because of the structure of the blocks).

It is possible to reduce these DCT artifacts by introducing in the criterion the following second constraint

$$J_3(f_k) = \eta_1^2 \int_{\Omega_R} \psi\left(\frac{|Rf_k|}{\delta}\right) d\Omega_R \tag{8}$$

which performs a soft thresholding of blocking artifacts in the wavelet domain. The support of the image in the wavelet domain is noted Ω_R and R represents the wavelet operator, δ is a parameter adjusting the soft-thresholding, and ψ is a potential function. Indeed, introducing constraint (8) in the minimization process implies that the term $\psi(|Rf_k|/\delta)$ and thus $|Rf_k|$ must be small, depending on the value of δ which control the soft thresholding. The value of η_1 depends on the choice of the wavelet coefficients to remove. η_1 is set to 1 if the coefficient is to be removed, otherwise η_1 is set to 0. With these assumptions, the MSE of coefficients with appropriate 3×3 patterns is evaluated [21], to set the value of η_1. This allows us to remove blocking effects but moreover, it avoids the appearance of "wrong objects" in the moving parts sequence c_k: DCT causes temporal variation of intensity of blocks that could be interpreted as small moving objects.

The functions φ and ψ we used in our approach correspond to the Green's function [12] given by

$$\varphi(u) = 2\log\cosh(u) \text{ with } \frac{\varphi'(u)}{2u} = \begin{cases} 1, & \text{if } u = 0 \\ \frac{\tanh(u)}{u}, & \text{if } u \neq 0 \end{cases} \tag{9}$$

3) Incorporating Quantization Constraint: During the decoding, the introduction of constraint $J_2(f_k)$ performs smoothing and tends to put values out of the quantization interval, increasing quantization noise. This global approach for the decoding, instead of a post-processing one, reduces these alterations: the knowledge of the quantization matrix used by the coder limits the possible values of each pixel of the reconstructed image to the quantization interval [38]. For a transformed coefficient, the possible values for Df_k belong to the interval $[p_k^* - q_k/2, p_k^* + q_k/2]$, with q_k the quantization

value for this coefficient, determined by the quantization matrix. Consequently, we introduce the penalty (10)

$$J_4(f_k) = \frac{1}{4}\mu_1^2 \int_{\Omega_D} \left(|\alpha(f_k)| - \alpha(f_k)\right)^2 d\Omega_D + \frac{1}{4}\mu_1^2 \int_{\Omega_D} \left(|\beta(f_k)| + \beta(f_k)\right)^2 d\Omega_D \quad (10)$$

where Ω_D is the support of the image in the DCT transform domain, μ_1 is a weighting factor for the penalty and

$$\begin{cases} \alpha(x) = Dx - p_k^* + \frac{q_k}{2} \\ \beta(x) = Dx - p_k^* - \frac{q_k}{2}. \end{cases} \quad (11)$$

In (10), if $\alpha(x) \geq 0$, i.e., $Df_k \geq p_k^* + (q_k/2)$, we find f_k such that $D^{-1}(Df_k - p_k^* - (q_k/2))$ tends to 0, i.e., f_k tends to $D^{-1}(p_k^* + (q_k/2))$ or equivalently Df_k tends to $p_k^* + (q_k/2)$. Thus, the value of Df_k is automatically projected onto the higher bound of the quantization interval. Similarly, if $\beta(x) < 0$, i.e., $Df_k < p_k^* - (q_k/2)$ the value of Df_k is projected onto the lower bound of the quantization interval, $p_k^* - (q_k/2)$. This term reduces quantization errors, but moreover, allows multiple successive compression/decompression of the sequence without accumulation of quantization errors. This is a fundamental advantage when transmission media have lower rate capacity than the compressed video, and bitrate transcoding is needed to meet the channel constraints.

4) New Criterion: The complete criterion can be rewritten as

$$J(f_k, c_k) = J_1(f_k, c_k) + J_2(f_k) + J_3(f_k) + J_4(f_k). \quad (12)$$

It performs a spatio–temporal segmentation and remove simultaneously compression artifacts (blocking effects and quantization noise). As we have seen in Section II-B2, the optimal solution of this minimization problem is obtained when $(\partial J(f_k, c_k))/(\partial f_k) = 0$, equivalent to

$$c_k^2\left(f_k - D^{-1}p_k^*\right) + \alpha_p(c_k - 1)^2(f_k - m_k) - \lambda_1^2 div\left(\frac{\varphi'\left(|\nabla f_k|\right)}{|\nabla f_k|} \cdot \nabla f_k\right) \quad (13a)$$

$$+ \eta_1^2 R^{-1} \frac{\psi'\left(\frac{|Rf_k|}{\delta}\right)}{\frac{|Rf_k|}{\delta}} Rf_k + \mu_1^2 D^{-1}\kappa(f_k) = 0 \quad (13b)$$

with

$$\kappa(x) = \begin{cases} Dx - p_k^* - \frac{q_k}{2}, & \text{if } Dx < p_k^* - \frac{q_k}{2} \\ Dx - p_k^* + \frac{q_k}{2}, & \text{if } Dx \geq p_k^* + \frac{q_k}{2} \\ 0, & \text{if } Dx \in \left[p_k^* - \frac{q_k}{2};\ p_k^* + \frac{q_k}{2}\right] \end{cases} \quad (14)$$

while the optimal map of moving objects c_k is still given by (5b).

To solve the nonlinear equation (13), we use the half-quadratic regularization method developed in [7] which permits to ensure convexity of the criterion in f_k. The minimization consists in a sequence of linear systems, solved by an iterative method like Gauss-Seidel or conjugate gradient. In order to find the optimal solution, alternate minimizations in f_k and c_k are performed as defined in Section II-B2 until convergence in f_k. This iteration process, associated with the simultaneous regularization, removes all residual trails of objects from the background, even for slow moving objects.

D. Processing of the Moving Objects

This section corresponds to part 2 of Fig. 1. Remember that the minimization of functional (12) gives

- the estimation of a background image for each frame, on which blocking effects were removed;
- a representation of the moving parts of the sequence, without being negatively influenced by blocking artifacts that appear in the original sequence.

1) Preprocessing: The aim of this section is to extract from this representation a list of objects and their spatial and temporal characteristics, in order to be able to process them separately in the next section. Thus, elementary methods were chosen as tools for spatial segmentation and tracking and are described briefly in Sections II-D1a and II-D1b. More elaborate ones could be applied, such as [4], [11], [20].

a) Spatial segmentation: The c_k images are processed. First, each moving object is spatially isolated from the others in each image. A thresholding controls the amount of objects in the image. Then, binary mathematical morphology operations are performed (combinations of dilations and erosions with a 3×3 structuring element), to connect different components in complete objects. Finally, each object is labeled and its spatial properties are evaluated: height, width, barycentre, size. Small objects of size lower than a given threshold (typically 50 pixels) corresponding to noise are immediately removed.

b) Tracking: For each object in image i, motion estimation is performed using a simple block matching algorithm. The motion vector $(u_{i,j}^l, v_{i,j}^l)$ corresponding to the motion of object l from image i to image j is obtained. Notice that for MPEG, motion vectors are directly available from the bitstream. If the motion vector field of the object is coherent, the tracking of the object is successful. The knowledge of the motion vectors gathered with the spatial ones results in the accurate knowledge of each object of the sequence.

2) Objects Processing: This section corresponds to part 4 of Fig. 1. At this point, we isolated and labeled each object, and we know its motion during the sequence. Thus, each object can be processed independently. The processing detailed in this section is applied to each object: computation time is highly reduced on parallel computers.

c) Blocking effects and quantization noise removal: Let O_k^l represent the object l in the image k, with $l \in [1; L]$, $k \in [1; N]$, N is the number of images, and L the number of objects detected in these N images. For each object l, and for each image k, the new representation of the object is obtained with the minimization of the criterion (15)

$$L\left(O_k^l\right) = L_1\left(O_k^l\right) + L_2\left(O_k^l\right) + L_3\left(O_k^l\right) + L_4\left(O_k^l\right). \quad (15)$$

The data driven term $L_1(.)$ performs a temporal average with motion compensation. It is given by

$$L_1\left(O_k^l\right) = \sum_{i=-n}^{n} \int_{\Omega_{O^l}} \left(O_k^l - D^{-1}p_{k+1}^* \times \left(x + u_{k,k+i}^l, y + v_{k,k+1}^l\right)\right)^2 d\Omega_{O^l} \quad (16)$$

where Ω_{O^l} is the support of the object l in the spatial domain. The value of n depends on the object characteristics. If the shape of the object changes rapidly, n has to be small. The motion vector $(u^l_{i,j}, v^l_{i,j})$ results from the motion estimation detailed in Section II-D1b.

As explained in Section II-C2, the term

$$L_2\left(O^l_k\right) = \lambda_2^2 \int\limits_{\Omega_{O^l}} \varphi\left(\left\|\nabla O^l_k\right\|\right) d\Omega_{O^l} \tag{17}$$

performs the spatial regularization on the object: smoothing while preserving sharp edges. The term

$$L_3\left(O^l_k\right) = \eta_2^2 \int\limits_{\Omega_{O^l_R}} \psi\left(\frac{\left|RO^l_k\right|}{\delta}\right) d\Omega_{O^l_R} \tag{18}$$

performs the soft thresholding in wavelets domain to remove blocking effects. The support of object l in the wavelet transform domain is noted $\Omega_{O^l_R}$. Finally

$$\begin{aligned} L_4\left(O^l_k\right) = &\frac{1}{4}\mu_2^2 \int\limits_{\Omega_{O^l_D}} \left(\left|\alpha\left(O^l_k\right)\right| - \alpha\left(O^l_k\right)\right)^2 d\Omega_{O^l_D} \\ &+\frac{1}{4}\mu_2^2 \int\limits_{\Omega_{O^l_D}} \left(\left|\beta\left(O^l_k\right)\right| + \beta\left(O^l_k\right)\right)^2 d\Omega_{O^l_D} \end{aligned} \tag{19}$$

performs the projection onto the quantization interval to reduce quantization noise on the object. The support of object l in the DCT transform domain is noted $\Omega_{O^l_D}$. The values of $\alpha(O^l_k)$ and $\beta(O^l_k)$ are given in (11).

As well as for the background, blocking effects are reduced spatially on blocks edges, by constraints (17) and (18), and temporally by the term (16). Here, the functions φ and ψ also correspond to the Green's function.

d) Criterion minimization in O_k: The estimated object is given by

$$O^{l^*}_k = \arg\min_{O^l_k}\left(L\left(O^l_k\right)\right) \tag{20}$$

The optimal solution of this minimization problem is obtained with $(\partial L(O^l_k))/(\partial O^l_k) = 0$, equivalent to

$$\begin{aligned} &\sum_{i=-n}^{n}\left(O^l_k - D^{-1}p^*_{k+i}\right) - \lambda_2^2 div\left(\frac{\varphi_3'\left(\left|\nabla O^l_k\right|\right)}{\left|\nabla O^l_k\right|}\nabla O^l_k\right) \\ &+\eta_2^2 R^{-1}\frac{\psi_2'\left(\frac{\left|RO^l_k\right|}{\delta}\right)}{\frac{\left|RO^l_k\right|}{\delta}} RO^l_k + \mu_2^2 D^{-1}\kappa\left(O^l_k\right) = 0 \end{aligned} \tag{21}$$

where $\kappa(x)$ is the function given in (14). The method to find the solution of (21) is the same as the one given in Section II-C4.

E. Handling of MPEG I, P, and B Frames

All the previous explanations were given for either M-JPEG frames, or I-frames of a MPEG codec. For P- and B-frames, slight modification are applied.

TABLE II
DESCRIPTION OF THE TEST SET USED FOR COMPRESSION ARTIFACTS

BBC News	News report, and formula 1. Moderate motion, high amount of details, scene cut.
Formula 1	Formula 1 scene, extremely fast motion, with lot of spatial content and colors.
FlatTV	Philips commercial, moderate motion, high amount of details, scene cuts.
Weather	Weather report, shows the announcer chroma-keyed onto a map. Contains slow motion, some details and colors, and a static background.
Hall	COST-211 European project test-sequence, static background.
Surfing	Philips commercial, fast motion, many textures, scene cut.
Van	OSIAM RNRT project test-sequence, static background.

TABLE III
SCORES OBTAINED ON THE MPEG-4 SEQUENCES WITH AND WITHOUT POST-PROCESSING, AND OMD, BY TWO OBJECTIVE METRICS: GBIM AND OLqM (THE LOWER THE VALUE, THE BETTER THE RESULT)

	GBIM			OLqM		
	MPEG	MPEG+PP	OMD	MPEG	MPEG+PP	OMD
BBC News	1.48	1.13	0.82	348	48	10
Formula 1	1.45	1.0	0.92	539	57	9
FlatTV	1.29	1.02	0.84	176	22	3
Weather	1.65	1.35	1.21	426	18	2
Hall	1.28	0.99	0.93	425	33	5
Surfing	2.72	1.30	0.98	870	56	3
Van	1.55	1.27	1.16	329	100	33

TABLE IV
SUBJECTIVE SCORES OBTAINED ON THE MPEG-4 SEQUENCES WITH AND WITHOUT POST-PROCESSING, AND OMD, USING A PANEL OF 12 VIEWERS, NAIVES AND EXPERTS (THE HIGHER THE VALUE, THE BETTER THE RESULT)

	MPEG	MPEG+PP	OMD
BBC News	50	58	57
Formula 1	50	61	60
FlatTV	65	72	71
Weather	68	74	76
Hall	61	64	71
Surfing	35	45	46
Van	38	42	47

Constraint (10) is removed, since quantization values for P- and B-macroblocks are for the residual images. The operator D^{-1} in term A of (3) is removed, and consequently p^*_k represent the MPEG decoded frames (in the spatial domain). This part requires further studies, for instance to apply specific criterion directly on the residual P- or B-frame, or to take into account macroblocks from P- and B-frames that have been I-encoded, because of non coherence of the motion vectors. Moreover, the notion of GOP, and the detection of the I-frames are indicators of possible scene cut, that could help for resetting the average m_k.

F. Sequence Reconstruction

This section corresponds to part 5 of Fig. 1. The final sequence $\widetilde{p}^*_k$ is reconstructed by projecting the objects $O^{l^*}_k$ on the estimated backgrounds f^*_k:

$$\widetilde{p}^*_k = c_k^2 f^*_k + (c_k - 1)^2 O^*_k \tag{22}$$

with $O^*_k = \bigcup_l O^{l^*}_k$ as the representation of the gathered objects, for the frame k. For a given pixel (x, y), if it belongs to a moving object, $c_k(x, y) = 0$ and the pixel from O^*_k is used. Otherwise, the pixel from f^*_k is used.

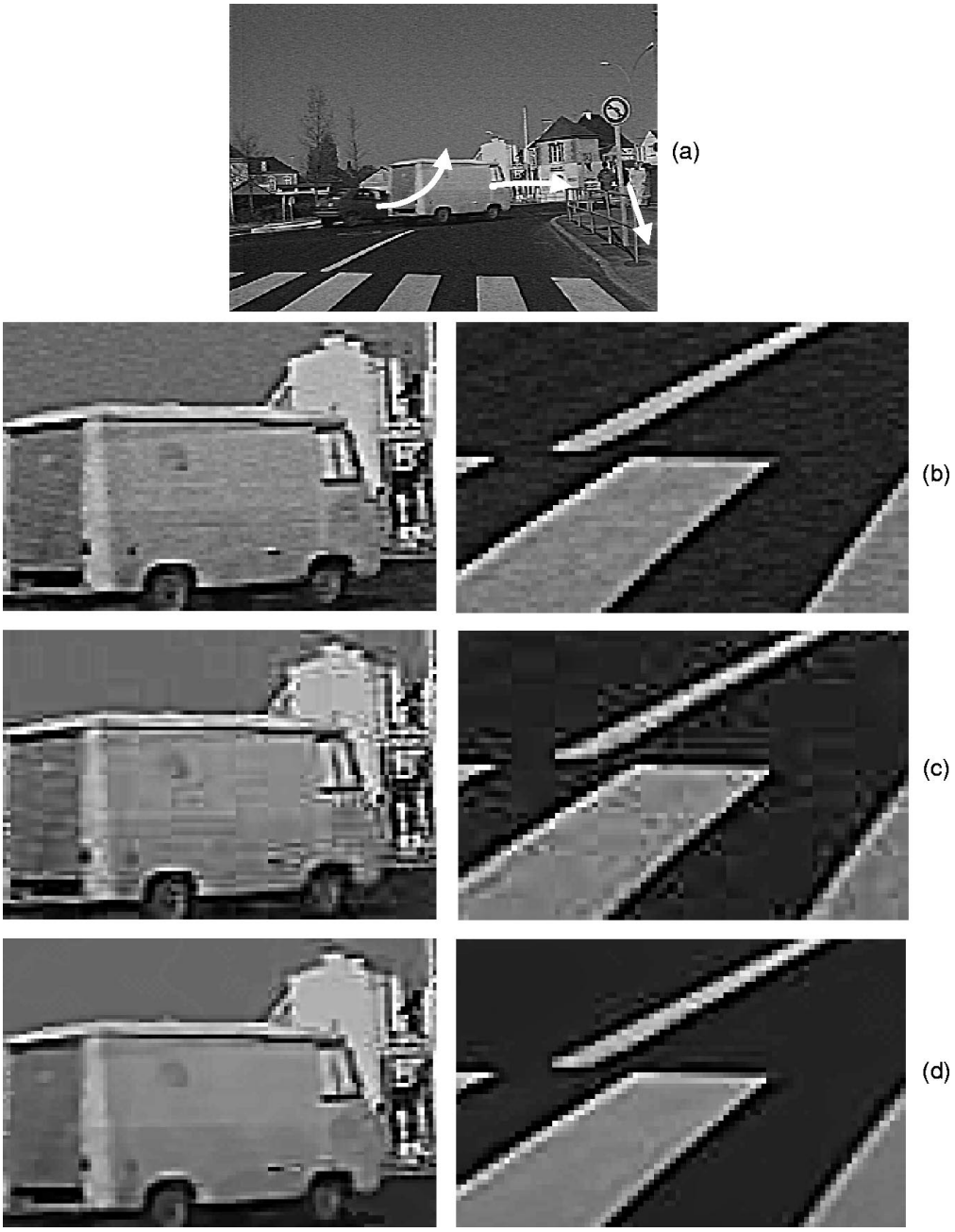


Fig. 4. The 100th image of the sequence "Van" (RNRT/OSIAM project): (a) overview of the scene (b) original uncompressed image (c) image decoded using standard M-JPEG decoding (d) image decoding using proposed decoding algorithm.

III. Removal of Transmission, Acquisition and Storage Artifacts

A. Dropouts: Causes and Consequences

The removal of the dropouts is fully integrated in the decoding method described in Section II. Dropouts in video sequences can result from the following.

- Transmission: current networks are not adapted to the transmission of video sequences. Many efforts consisted in understanding the packet delay and loss behavior [2], to design network algorithms such as routing and flow control. Noisy channels involve losses of packets that

can have very different effects on the visual aspect of the decoded sequence, depending on the role of the corrupted or lost area. In M-JPEG and MPEG I-frames, losses often appear as blocks or horizontal lines. If MPEG B/P-frames are corrupted, losses appear as shiftings in the sequence, objects can be duplicated, or their motion modified.

- Acquisition: a real-time DV compression performed by the new digital camcorders, or acquisition via video capture cards can produce defects on the sequence. In the case of camcorders, dropouts encountered are due to small bugs in the real time codec, that appear most of the time as small rectangular areas with uniform color: they do not appear to replicate any other set of pixels in the image.
- Storage: the principal means of storage used are magnetic tapes. Magnetic tapes are fragile, sensitive to repetitive playbacks, and liable to the unavoidable chemical breakdown of molecules and particles. Head clogging or dust on the tape produces a "banding effect": bands of image freeze and finally dissolve, producing a mosaic of small rectangular dropouts in the next images.

All these causes result in the same damage to the video: loss of blocks, shifting, banding, color alteration, are the most frequent.

B. Dropouts Detection and Removing

The removal of the dropouts requires very few additional computations thanks to the foreground/background separation and the object-based approach of the method. It corresponds to parts 3 and 4 of Fig. 1. The dropouts are removed from the background by (3). But of course, each dropout was detected as a moving object and appears in the c_k sequence, representing the moving objects. The basic idea is to remove these objects according to spatial and temporal assumptions. For instance, the main characteristic of a dropout is its short-life appearance in the sequence. Furthermore, it is also possible to benefit from the transmission channel characteristics if the transport layer protocol is known [25]. Such objects are removed from the c_k sequence, and thus are not considered during the sequence reconstruction. They are replaced by the estimated background: resulting interpolation of missing data is temporal, because background was computed by a temporal average.

On moving objects, dropouts removal is performed by (16). The temporal average and the regularization on each object reduce the variation of intensity between the dropout and its proximity. Moreover, for $i \in [-n; n]$, if one pixel value for $D^{-1}p_{k+i}$ is very different from the $2n$ others, it is assumed to belong to a dropout, and is not considered in the temporal average, not to affect the final value. Resulting interpolation of moving objects is spatial and temporal.

IV. Experimental Results

This section presents the results of the evaluation of the method. In the first part, the improvement in term of compression artifact reduction is assessed using both objective and subjective tests. In the second part, results on transmission, storage and playback artifacts removal are presented. Finally, the complexity of the algorithm is evaluated.

TABLE V
TUNING OF PARAMETERS ACCORDING TO THE VALUE OF q_{fact}

	1 to 10	11 to 20	20 to 30	31
α_p formula (3)	2	2	2	2
λ_1 formula (7)	1	3	7	10
μ_1 formula (10)	1	1	2	3

A. Compression Artifacts Reduction

1) Test Set Description: Table II presents the seven CIF sequences (352 × 288 pixels) of various contents and characteristics that were used for the benchmark.

Sequences were MPEG-4 encoded, by a Philips proprietary encoder, derived from the VM12 having only single VOP of rectangular size, with block-based DCT coding. Several bitrates between 700 kb/s and 100 kb/s were applied to each sequence. Each GOP holds 50 frames (only I– and P-frames). The proposed algorithm, called OMD (for Optimal MPEG Decoder), is compared to the standard decoder, with and without the post-processing algorithm included in MPEG-4 VM8 [33], i.e., spatial deblocking and deringing filters.

2) Objective Evaluation: At low bitrates, PSNR is definitely unable to assess the visual quality of image sequences, and a-fortiori to rank post-processing methods. Moreover, PSNR deals with image fidelity, and not with image quality. Consequently, we chose two other objective metrics that are expected to predict subjective quality more accurately. The first one, the Generalized Blocking Impairment Metric (GBIM) [43], evaluates the amount of blocking artifacts in a sequence, taking into account luminance masking effect. The second one, the Overall Linear quality Metric (OLqM) [5], measures three impairments due to compression: blocking effect, ringing artifact, and corner outliers. The lower the values of the metrics, the lower the amount of impairments in the sequence. Table III shows the results of the objective evaluation, by averaging the values obtained for each bitrates.

Results provided by both GBIM and OlqM are coherent: they show that the impairment reduction is effective for each post-processed sequence, and that OMD sequences have less artifacts than MPEG-4 post-processing ones.

Objective tests indicate the amount of impairment, and so give a global idea of the visual quality. Nevertheless, these metrics only concentrate on impairments (undesirable features) and not on attributes (desirable features) such as sharpness, contrast, or resolution. This is why subjective tests are required to confirm the feeling that the proposed algorithm improves upon current state of the art in term of visual quality.

3) Subjective Evaluation: A combination of evaluation by advanced experts and subjects with and without expertise in video processing was used. Twelve people were asked to give a score between 0 (worst) and 100 (best), according to perceived visual quality.

Video segments include the original source and the processed versions. The monitor used for experimentations is a Barco professional-grade high definition monitor. The maximum observation angle is 30°, and the room illumination is low. Finally, the viewing distance is in the range of 4H to 6H, four to six times the height of the picture, recommended by the Video Quality Expert Group (VQEG) [47], and compliant with the Recommen-

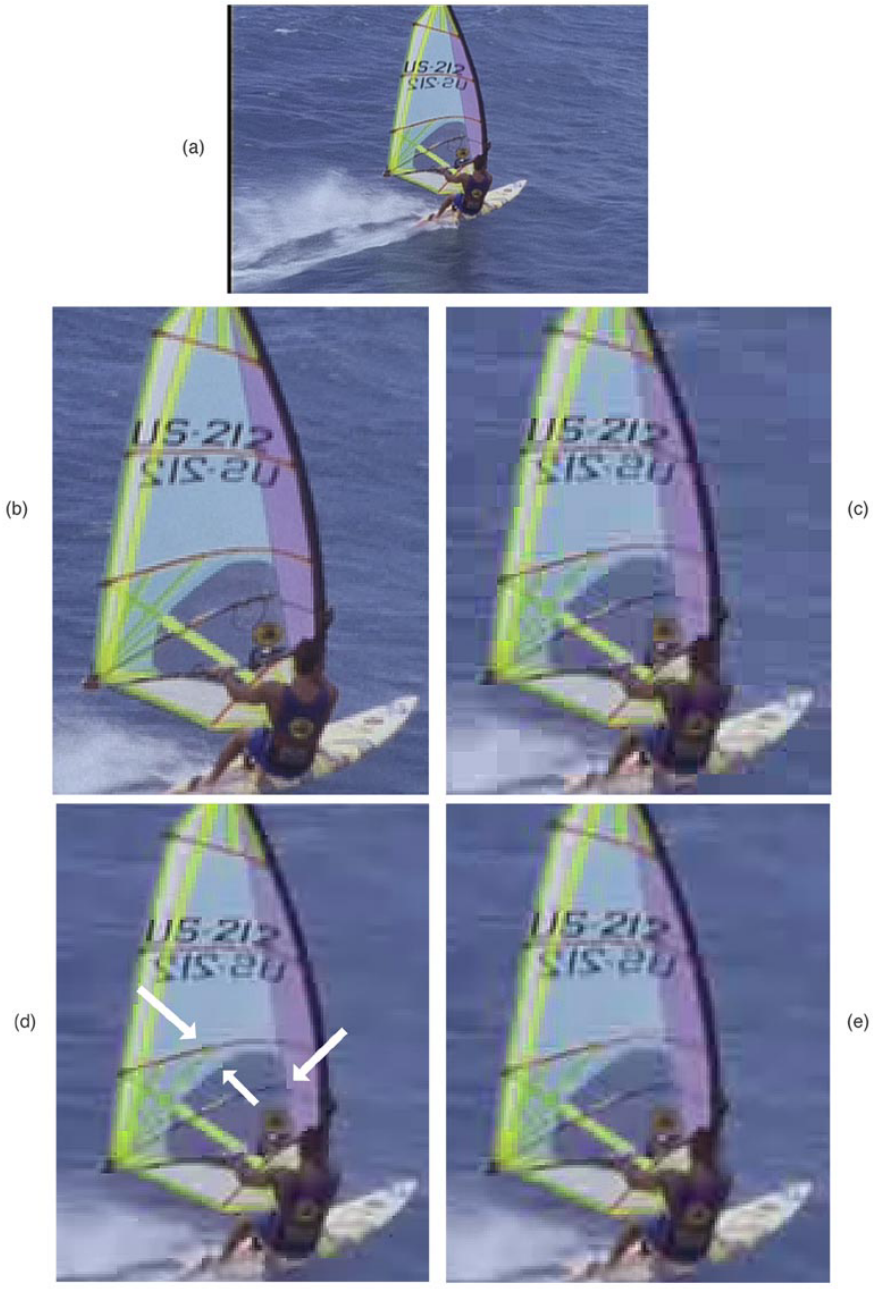


Fig. 5. Decoding results on an I frame from the sequence "Surfing" encoded by MPEG-4 VM12, at 300 kb/s. (a) shows the complete original frame, (b) the original, (c) the MPEG-4 decoded frame, (d) the sequence with MPEG-4 post-processing applied, and (e) the proposed method. Arrows highlight that some blocking artifacts that remain in (d) are removed in (e).

dation ITU-R BT.500-10. Table IV shows the subjective scores obtained.

The scale of subjective scores can be interpreted as follows: if the difference between two scores is smaller or equal to 2, the differences between the two sequences are hardly distinguishable. For a difference above 2, all experts and some naives start seeing the differences, and for a difference above 5, everyone clearly see the difference.

These results show that OMD performs really better on sequences with static background, and as well as MPEG-4 post-processing on moving background sequences. A deeper analysis of the results even shows that OMD performs better on

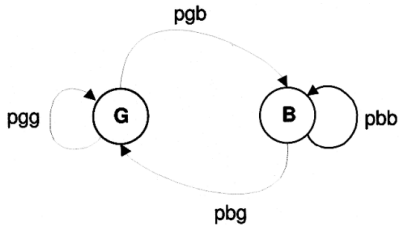


Fig. 6. Model of the internet channel. G stands for "Good" state and B for "Bad" state.

moving background sequences for expert viewers. Naives meet with difficulties for seeing differences between post-processing methods, especially at high bitrates. Most of the time, naïve viewers prefer very smoothed video (like MPEG-4 post-processed) for fast moving sequences.

Fig. 4 presents the results on the reconstructed sequence "Van," encoded by the M-JPEG algorithm. The compression rate is 12.6:1. Such a low compression rate is sufficient to exhibit severe artifacts because the original (uncompressed) sequence is very noisy. Frame (a), corresponding to the original uncompressed 100^{th} image of the sequence, describes the motion of the objects. Frame (b) corresponds to a window extracted from the original, (c) shows the same window, M-JPEG coded and then reconstructed with a standard decoder. As can be observed, lots of artifacts appear in the background and on the objects. Frame (d) is the sequence decoded by the proposed method. The reduction of blocking artifacts has a significant visual impact, and leads to more accurate images. Artifacts were smoothed out, avoiding excessive blurring of the discontinuities. Moreover, noise present on the original sequence was efficiently removed.

The tuning of the parameters is most certainly empirical, but some general rules can be applied. Table V summarizes the values of the main parameters depending on the quantization value q_{fact}. (q_{fact} is the factor that multiplies the quantization matrix, between 1 and 31 for the MPEG-4 codec). Of course q_{fact} does not reflect exactly the amount of blockiness, that depends on the content of the scene too.

Fig. 5 represents the decoding results on an I frame from the sequence "Surfing," encoded by MPEG-4 VM12, at 300 kb/s. This sequence presents large textured areas, and very fast motion. Fig. 5(a) shows the complete original frame, (b) the original, (c) the MPEG-4 decoded frame, (d) the sequence with MPEG-4 post-processing applied, and (e) the proposed method. Arrows highlight some blocking artifacts that remains in (d) but are removed in (e). This points out the limits of the MPEG-4 post-processing method, in tough conditions, i.e. textured areas, fast motions or very low bitrates.

Observing these results, it is important to keep in mind that

- the aim is not fidelity with the original, but final visual quality;
- video quality differs from still image quality.

During subjective tests, we observed that most of the time, viewers (especially naive ones) prefer smoother images when video is played back, depending on their own regions of interest. Here is an advantage of our method: the object-based decoding approach allows adaptive smoothing according to the region types.

B. Transmission Artifacts and Dropout Removal

1) Modeling of the Channel: In order to simulate realistic loss of cells, the Internet behavior was simulated according to the works of J. C. Bolot [2]. The channel is characterized by the round trip delay rtt, and by the unconditional loss probability $ulp = \Pr\{rtt = 0\}$, where $rtt = 0$ corresponds to a lost cell. The probability for a cell to be lost, knowing that the previous cell was lost, called conditional loss probability, is given by $clp = \Pr\{rtt_{n+1} = 0 | rtt_n = 0\}$.

In this experiment, the Internet is simulated by a markovian model (see Fig. 6) that consists in two states: "G" for "good," where all cells are perfectly received, and "B" for "bad," where all of them are lost. With this model, the global rate loss is given by

$$P(B) = \frac{pgb}{pgb - pbb + 1} \tag{23}$$

where $pgb = P(B|G)$ is the probability to move from "G" to "B" state, and pbb the probability to stay in the "B" state.

2) Results: ATM cell losses were simulated on the sequence "Road," used in the COST211 European project. In order to produce realistic losses, we chose $pgb = 0.11$ and $pbb = 0.18$ [2]. The corresponding global loss rate is 11.8%. Sending 150 images of the M-JPEG compressed sequence at 220 kb/s, required to send 28 296 ATM cells of 48 bytes of data each. Among these 28 296 cells, 3962 were lost. Fig. 7 corresponds to the 110th image of the corrupted sequence. In particular, ten slices have been lost on this image, both on the objects and on the background. Fig. 7(c) presents the decoding with the standard method without the cell loss. Fig. 7(a) and (d) present the decoding with standard method with cell losses, and Fig. 7(b) and 7(e) the reconstructed image with the proposed method.

The recovery on the background is very efficient, due to convergence of these pixels to the advanced temporal average, and the motion compensation on the objects. On the objects, losses were recovered too. Nevertheless, this result could be improved using a more accurate motion estimation. It clearly appears that the decoding method takes benefit of the iterative optimization approach, and its object based particularity.

Fig. 8 corresponds to images of sequence "Hall." This sequence was encoded by MPEG-1 algorithm, with a target rate set to 256 kb/s. Three transmission channel errors were simulated: Fig. 8(a) and (c) correspond to a loss of 256 bytes, and Fig. 8(b) was obtained by randomly modifying 128 consecutive bytes of the sequence, that have probably affected P and B frames. In Fig. 8(a) and (c), the defect is detected in the c_k image as a large new object that appears and progressively dissolves on four consecutive images. It has no repercussion on the background estimation in (3). Spatial and temporal characteristics of this dropout allow to remove it easily. In Fig. 8(b), the detected object is not associated with objects from the previous or the next frame. So, it is removed, considered as a dropout. But moreover, it is substituted by the object from the previous frame, because corresponding objects from previous and next frame are associated by the tracking.

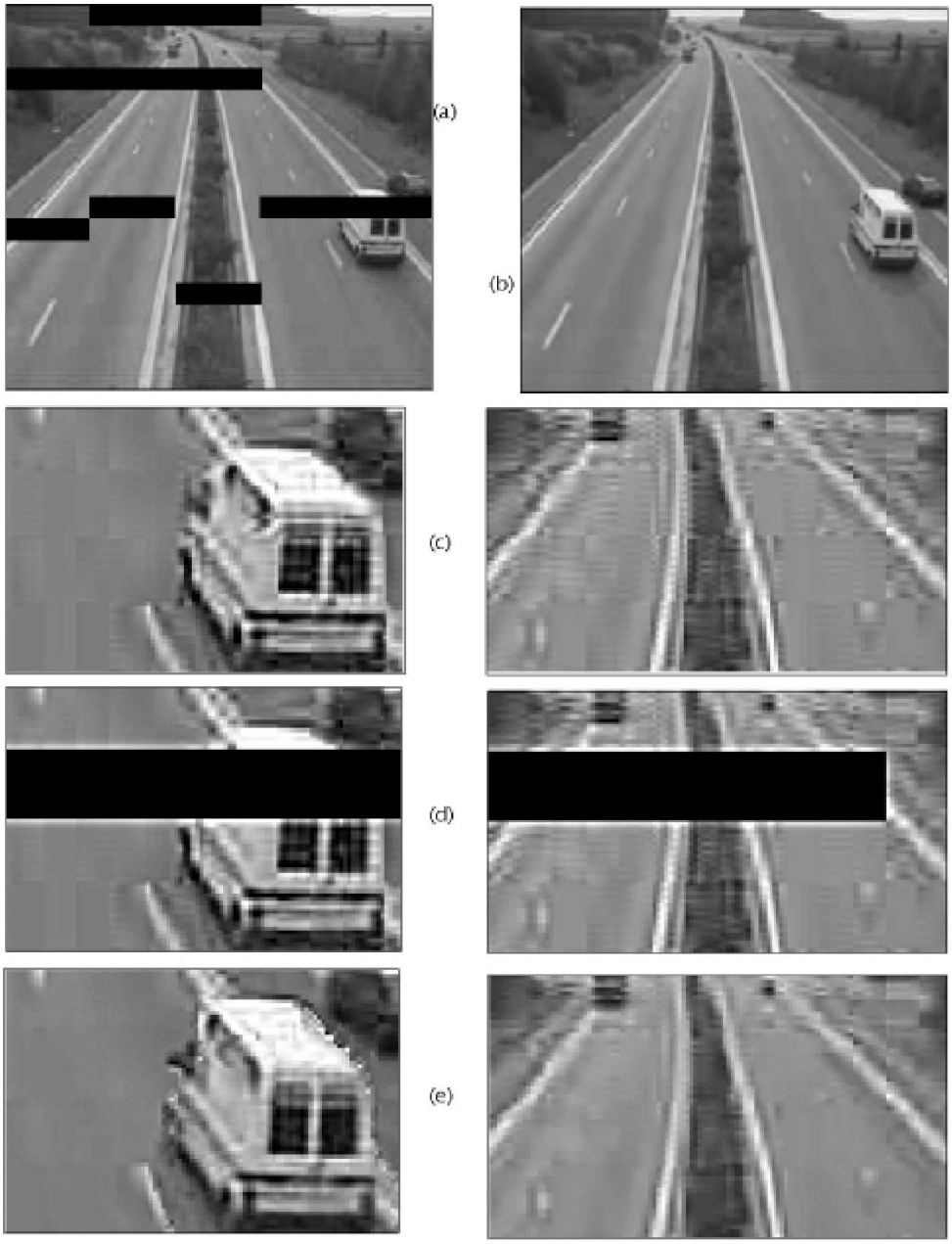


Fig. 7. ATM cell loss for the 100th image of the sequence "Road," decoded by (a) the standard algorithm, and by (b) the proposed method. Extracts of object and background: standard decoder (c) without loss, (d) with loss, and (e) proposed decoder with loss.

C. Complexity Evaluation

To conclude, let give an idea of the complexity of the algorithm. Tests on a Pentium II 450 MHz, with 512 M showed that for the "Hall" sequence, the processing of 100 frames requires 11.2 times more time than the standard decoding. In this sequence, two objects are detected and processed. Computation time depends on the number of objects and the number of iterations. Iterations are stopped when the evolution of two successive estimated image f_k is lower than a threshold. According to Fig. 1, the most time-consuming process are the block 1 (57%) and the block 4 (34%).

Fig. 8. The tenth, 48th, and 59th images of the sequence "Hall" (COST-211 European project) encoded using MPEG-1 (256 kb/s, 25 f/s) with simulated transmission losses. (Left) standard decoding and (right) proposed decoding.

Some simplifications are possible: a tradeoff between decoding precision and complexity is currently investigated. In conclusion, we estimate that a configuration yielding to pleasant results can be obtained for a processing that costs 3.2 times more than the standard decoding. These changes consist for instance in performing the thresholding in wavelets domain as a pre-processing in the decoder.

V. Conclusion

This paper presents a decoding scheme for block-coded video sequences. This efficient new method for improving visual quality differs from existing techniques by tackling simultaneously the problem of blocking effects corresponding to compression artifacts, and the problem of dropouts due to acquisition, transmission and/or storage errors. It performs simultaneously an estimation of the background and a detection of moving objects using motion segmentation. A second step consists in processing each object independently. Experimental results show that our method increases the visual quality of the reconstructed sequence. Compared to standard decoding, annoying temporal effects resulting from DCT blocks are largely reduced.

ACKNOWLEDGMENT

The authors want to acknowledge the anonymous reviewers for their advice, which helped improve the quality of the paper.

REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Apr. 1992.

[2] J. C. Bolot, "Characterizing end-to-end packet delay and loss in the internet," *J. High Speed Networks*, vol. 2, no. 3, pp. 305–323, Dec. 1993.

[3] J. C. Brailean and A. K. Katsaggelos, "Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences," *IEEE Trans. Image Processing*, vol. 4, pp. 1231–1251, Sept. 1995.

[4] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *Int. Conf. Computer Vision*, 1995.

[5] J. E. Caviedes and J. Jung, "No-reference metric for a video quality control loop," in *5th World Multiconf. Systemics, Cybernetics and Informatics, SCI'2001*, Orlando, FL, July 2001.

[6] J. E. Caviedes, C. Miro, and A. Gesnot, "Algorithm and architecture for blocking artifact correction unconstrained by region types," in *Proc. PCS'2001*, Seoul, Korea, Apr. 2001, pp. 89–92.

[7] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 5, pp. 298–311, Feb. 1997.

[8] C. Derviaux, F. X. Coudoux, M. G. Gazalet, and P. Corlay, "Blocking artifact reduction of DCT coded image sequences using a visually adaptive postprocessing," in *Int. Conf. Image Processing*, vol. 1, Lausanne, Switzerland, Sept. 1996, pp. 5–8.

[9] M. Ebbecke, M. B. H. Ali, and A. Dengel, "Real time object detection, tracking and classification in monocular image sequences of road traffic scenes," in *IEEE Int. Conf. Image Processing*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 402–405.

[10] M. Ghanbari, "Cell loss concealment in ATM video codes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 238–247, June 1993.

[11] C. Gu, T. Ebrahimi, and M. Kunt, "Morphological object segmentation and tracking for content based video coding," in *Multimedia Communication and Video Coding*, New York, Nov. 1995.

[12] P. J. Green, "Bayesian reconstruction from emission tomography data using modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 1, pp. 194–202, 1990.

[13] S. S. Hemami and T. H. Y. Meng, "Transform coded image reconstruction exploiting interblock correlation," *IEEE Trans. Image Processing*, vol. 4, pp. 1023–1027, July 1995.

[14] T. Hoshi, T. Komatsu, and T. Saito, "Film blotch removal with a spatiotemporal fuzzy filter based on local image analysis of anisotropic continuity," in *Int. Conf. Image Processing*, vol. 2, Chicago, IL, Oct. 1998, pp. 478–482.

[15] T. Hsung, D. Lun, and W. Siu, "A deblocking technique for JPEG decoded image using wavelets transform modulus maxima representation," in *Int. Conf. Image Processing*, vol. 2, Lausanne, Switzerland, Sept. 1996, pp. 561–564.

[16] International Organization for Standardization ISO/IEC JTC 1/SC 29/WG 11, Coding of Moving Pictures and Audio, Maui, HI, Dec. 1999.

[17] *Joint Committee Draft (CD)*, May 2002.

[18] S. Jehan-Besson, M. Barlaud, and G. Aubert, "Deformable regions driven by an Eulerian accurate minimization method for image and video segmentation, application to face detection in color video sequences," in *Eur. Conf. Computer Vision*, Copenhagen, Denmark, 2002.

[19] ——, "Region-based active contours for video object segmentation with camera compensation," in *Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001.

[20] ——, "An object-based motion method for video coding," in *Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001.

[21] J. Jung, M. Antonini, and M. Barlaud, "Optimal JPEG decoding," in *Int. Conf. Image Processing*, vol. 1, Chicago, IL, Oct. 1998, pp. 410–414.

[22] ——, "Optimal Video Decoder Based on MPEG-Type Standards," French Patent 99/07 443 and U.S. Patent 09 406 673, June 1999.

[23] ——, "An efficient new object-based variational approach for MPEG video decoding," *Ann. Telecommun.*, vol. 55, no. 3–4, pp. 101–107, Mar.-Apr. 2000.

[24] ——, "Removing blocking effects and dropouts in DCT-based video sequences," in *IST/SPIE 12th Int. Symp.*, San Jose, CA, Jan. 2000.

[25] ——, "A new object-based variational approach for MPEG2 data recovery over lossy packet networks," in *Proc. SPIE, Electron. Imag.*, San Jose, CA, Jan. 2001.

[26] A. Kokaram, R. Morris, W. Fitzgerald, and P. Rayner, "Detection of missing data in image sequences," *IEEE Trans. Image Processing*, vol. 4, pp. 1496–1508, Nov. 1995.

[27] ——, "Interpolation of missing data in image sequences," *IEEE Trans. Image Processing*, vol. 4, pp. 1509–1519, Nov. 1995.

[28] P. Kornprobst, R. Deriche, and G. Aubert, "Image sequence restoration: a PDE based coupled method for image restoration and motion segmentation," in *Eur. Conf. Computer Vision*, Freiburg, Germany, June 1998, pp. 548–562.

[29] Y. L. Lee, H. C. Kim, and H. W. Park, "Blocking effect reduction of JPEG images by signal adaptative filtering," *IEEE Trans. Image Processing*, vol. 7, pp. 229–234, Feb. 1998.

[30] M. J. Nadenau and S. K. Mitra, "Blotch and scratch detection in image sequences based on rank ordered differences," in *5th Int.Workshop on Time-Varying Image Processing and Moving Object Recognition*, Florence, Italy, Sept. 1996.

[31] S. Minami and A. Zakhor, "An optimization approach for removing blocking effects in transform coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 74–82, Apr. 1995.

[32] H. Paek, J. W. Park, and S. U. Lee, "Non-iterative post-processing technique for transform coded image sequence," in *Int. Conf. Image Processing*, vol. 3, Washington, DC, Oct. 1995, pp. 208–211.

[33] H. W. Park and Y. L. Lee, "A postprocessing method for reducing quantization effects in low bit-rate moving picture coding," *IEEE Trans. Circuits Syst. Video Technology*, vol. 9, pp. 161–171, Feb. 1999.

[34] B. Ramamurthi and A. Gersho, "Nonlinear space-variant postprocessing of block coded images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 5, Oct. 1986.

[35] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1990.

[36] P. Salama, N. Shroff, and E. J. Delp, "A bayesian approach to error concealment in encoded video streams," in *Int. Conf. Image Processing*, vol. 2, Lausanne, Switzerland, Sept. 1996, pp. 49–52.

[37] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projection onto convex sets," *IEEE Trans. Image Processing*, vol. 4, pp. 470–477, Apr. 1995.

[38] S. Tramini, M. Antonini, M. Barlaud, and G. Aubert, "Quantization noise removal for optimal transform decoding," in *Int. Conf. Image Processing*, vol. 1, Chicago, IL, Oct. 1998, pp. 381–385.

[39] S. Tramini, M. Antonini, and M. Barlaud, "Intraframe image decoding based on a nonlinear variational approach," *Int. J. Imag. Syst. Technol.*, vol. 9, no. 5, Oct. 1998.

[40] J. Vass, K. Palaniappan, and X. Zhuang, "Automatic spatio–temporal video sequence segmentation," in *Int. Conf. Image Processing*, vol. 1, Chicago, IL, Oct. 1998, pp. 958–962.

[41] Y. Wang and Q. F. Zhu, "Error control and concealment for video communication: a review," *Proc. IEEE*, vol. 86, p. 974, May 1998.

[42] J. L. H. Webb, "Postprocessing to reduce blocking artifacts for low bit-rate video coding using chrominance information," in *Int. Conf. Image Processing*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 9–12.

[43] H. R. Wu and M. Yuen, "Quantitative quality metrics for video coding blocking artifacts," in *Proc. Picture Coding Symp. 1*, Melbourne, Australia, 1996, pp. 23–26.

[44] S. F. Yang, "Linear restoration of block-transform coded motion video," *Vis. Commun. Image Process.*, Feb. 1997.

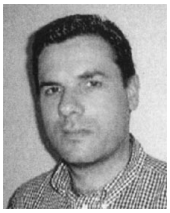
[45] Y. Yang, N. Galatsanos, and A. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 421–432, Dec. 1993.

[46] Y. Yang and N. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," *Trans. Image Processing*, vol. 6, pp. 1345–1357, Oct. 1997.

[47] (1999) Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. [Online] http://www.crc.ca/vqeg

Joël Jung was born in France on June 15, 1971. He received the Ph.D. degree from the University of Nice-Sophia Antipolis, France, in 2000.

From 1996 to 2000, he worked with the I3S/CNRS Laboratory, University of Nice-Sophia Antipolis, on the improvement of video decoders based on the correction of compression and transmission artifacts. He is currently with Philips Research France, Suresnes, and his research interests are video decoding, post-processing, perceptual models, objective quality metrics and low power codecs. He is involved in the SOQUET (System for management Of Quality of service in 3 G nETworks) European project, and contributes to VQEG (Video Quality Expert Group) and MPEG/JVT standards.



Marc Antonini received the Ph.D degree in electrical engineering from the University of Nice-Sophia Antipolis, France, in 1991.

He was a postdoctoral fellow at the Centre National d'Etudes Spatiales, Toulouse, France, in 1991 and 1992. Since 1993, he is working with CNRS at the I3S Laboratory both from CNRS and University of Nice-Sophia Antipolis. His research interests include multidimensional image processing, wavelet analysis, lattice vector quantization, information theory, still image and video coding, joint source/channel coding, inverse problem for decoding, multispectral image coding, multiresolution 3-D mesh coding.

Dr. Antonini is a regular reviewer for several journals (IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON INFORMATION THEORY IEEE TRANSACTIONS ON SIGNAL PROCESSING, and *Electronics Letters*) and participated to the organization of the IEEE Workshop Multimedia and Signal Processing 2001 in Cannes, France). He also participates to several national research and development projects with French industries, and in several international academic collaborations (University of Maryland, College Park, Stanford University, Stanford, CA, Louvain La Neuve, Belgium, and Rio de Janeiro, Brazil).

Michel Barlaud (M'88–SM'01–F'03) received the These d'Etat from the University of Paris XII, France.

He is currently a Professor of Image Processing at the University of Nice-Sophia Antipolis, France, and the leader of the Image Processing group of I3S. His research topics are image and video coding using scan based wavelet transform, inverse problem using half quadratic regularization, and image and video segmentation using region based active contours and PDEs. He is the author of a large number of publications in the area of image and video processing, and the Editor of the book *Wavelets and Image Communication* (Amsterdam, The Netherlands: Elsevier, 1994).

Dr. Barlaud is a regular reviewer for several journals and a member of the technical committees of several scientific conferences. He leads several national research and development projects with French industries, and participates in several international academic collaborations (University Maryland, College Park, Stanford University, Stanford, CA, Louvain La Neuve, Belgium).