



HAL
open science

Estimation du mouvement apparent majoritaire dans une séquence d'images vidéo par accumulation de votes bimodaux sur un histogramme approché

Frédéric Comby

► **To cite this version:**

Frédéric Comby. Estimation du mouvement apparent majoritaire dans une séquence d'images vidéo par accumulation de votes bimodaux sur un histogramme approché. Autre [cs.OH]. Université Montpellier II - Sciences et Techniques du Languedoc, 2001. Français. NNT : . tel-00185943

HAL Id: tel-00185943

<https://theses.hal.science/tel-00185943>

Submitted on 7 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACADEMIE DE MONTPELLIER
UNIVERSITE MONTPELLIER II
– Sciences et Techniques du Languedoc –

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE MONTPELLIER II

Discipline : **Génie Informatique, Automatique et Traitement du signal**

Formation doctorale : **Système Automatiques et Microélectroniques**

Ecole doctorale : **Information, Structures, Systèmes**

présentée et soutenue publiquement

par

Frédéric COMBY

Le 19 décembre 2001

Titre :

**Estimation du mouvement apparent majoritaire dans
une séquence d'images vidéo par accumulation de votes
bimodaux sur un histogramme approché**

JURY

Mme. Marie-José ALDON	Chargé de recherche CNRS au LIRMM	Directeur de thèse
Melle. Isabelle BLOCH	Professeur à l'ENST (Paris)	Rapporteur
M. Patrick BOUTHEMY	Directeur de recherche à l'INRIA (Rennes)	Examineur
M. Didier DUBOIS	Directeur de recherche à l'IRIT (Toulouse)	Rapporteur
M. Michel HABIB	Professeur au LIRMM	Président du jury
M. Olivier STRAUSS	Maître de conférence au LIRMM	Co-directeur de thèse

A Tiphaine, Killian et toute ma famille,

Remerciements

Attention, ce paragraphe est le dernier lisible (hormis les tables des matières) pour les non initiés. Quel que soit l'endroit où vous vous arrêterez, je vous souhaite bon courage pour la lecture.

Bien que généralement dans les films la liste des acteurs et des participants arrive après le mot FIN, c'est au tout début de mon manuscrit que je formulerai les remerciements à toutes les personnes ayant participé, de près ou de loin, à ma thèse.

En tout premier lieu, je souhaite remercier Michel Habib, directeur du LIRMM, Christian Landrault, directeur adjoint et Gaston Cambon, ancien directeur, pour m'avoir accueilli au sein du laboratoire pour mon DEA, ma thèse et qui sait ...

Ensuite, mes remerciements vont à Isabelle Bloch et Didier Dubois qui ont eu la gentillesse d'accepter d'être mes rapporteurs, Michel Habib qui a accepté d'être le président du jury et enfin Patrick Bouthemy pour avoir accepté d'être examinateur.

Je tiens à remercier tout particulièrement Olivier Strauss qui m'a encadré et guidé durant ces trois années de thèse et avec qui il est possible que j'ai nécessairement partagé de bons fou-rires et des moments extra. Mes remerciements vont aussi à Marie-José Aldon, mon directeur de thèse, qui a eu la patience de relire mon manuscrit et de jouer le rôle de filtre passe-rien pour les fêtes d'ortographe.

J'adresse aussi toute ma sympathie aux permanents du département robotique par ordre d'apparition : Alain Liegeois, François Prunet, Alain Fournier, Etienne Dombre (the Boss), André Crosnier, Bruno Jouvencel (mais oui tu finiras Baldur's gate II), René Zapata (un grand merci pour la pré-soutenance), Philippe Fraisse (un jour on la fera cette association sportive au LIRMM), François Pierrot (tu verras cette année rien de cassé), Didier Crestani, Philippe Poignet et David Andreu. N'oublions pas les exilés Pierre Dauchez et Jérôme Vaganay. Aux autres permanents avec qui j'ai été amené à travailler, je pense entre autres à Yves Bertrand, Jan Opderbecke, Jérôme Galy, je renouvelle mes remerciements.

J'adresse un grand merci à toutes les personnes de l'administration et des services techniques qui m'ont grandement simplifié la vie : Anne B. pour sa gentillesse et sa patience pour nos commandes foireuses, Nadine T. pour sa bonne humeur et les moments de décompression pendant la rédaction, Josette D. pour m'avoir raconté toute l'île de la réunion avant mon voyage de noces, Ghislaine T., Martine P., Nicole O. pour leurs

sourires, Philippe T. qui a su maintenir mon PC en vie pendant trois ans (même avec Windows dessus), Freddy notre cuistot chef pour toutes les chasses aux champignons et ses tours pendables, Pascal L. (dit moustache) et tous les autres.

J'adresse ensuite de chaleureux remerciements à tous ceux qui ont fait que ces années sont passées à la vitesse de la lumière : Philippe M. (Nours), Jean-Max D. ("this dedicace is yours" à prononcer en susurrant), Stephan F. (le maître des maîtres) et Sébastien W. (magic guy) pour plein de choses et notamment les parties de jeux de rôles et de battle, Serge B. (mon éternel binôme), la Magic Team de volley : Geovany B. (mon passeur préféré), Christophe C. (mon maître en pétanque), Gilles D. (Blocator), Isabelle V. (c'est l'histoire d'une blonde...), Philippe R. (bon anniversaire), la relève (qui sait) : Jean Damien C. (pour les bons fou-rires), Benoît T. (Tahiti Bob), Julien A. (Marathon man), Sébastien K (Hacker d'artichaut) et Frédéric M. (abla espanol??), Philippe L. (un petit tour avec les souris tic et tac), aux autres : Philippe B. (celui qui nous crève les Taipan), Vincent C. (don't worry...), Jérôme A., Abraham S., Meziane B., Nathalie P. (merci pour tout), Marc S., aux plus jeunes : Micaël M., Fabien L., Sébastien D., Fred S., Ivan E., Andrès V., Amornrit P., Tomas S., Christophe R. (La victime du Lochness), Cedric A. (pour tous p\$!§#n trois ans!!!) et aux anciens encore présents ou déjà partis Arnaud L. (Nono), Christine C. (Woodstock Power), Olivier C. (Mr Seita), mention spéciale à Erwann L. (et oui je l'ai soutenue avant toi) et tous les Wany boys : Laurent T., Denis M., Raphaël B., Manu M. (dans quelques années vous serez tous euros milliardaires), et tous ceux des départements de micro-élec et d'info que j'oublie.

Je voudrai aussi remercier mes parents sans qui rien de tout ça n'aurait pu se passer (ben oui pas de parents, pas de Fred).

Je remercie de tout mon coeur ma femme, Tiphaine, qui a su me supporter pendant toute la thèse (elle avait déjà résisté à l'agrégation, c'est dire si c'est la meilleure) et qui m'a donné le plus beau bébé du monde Killian. Je remercie aussi Killian d'avoir été adorable pendant toute ma fin de thèse (pourvu que ça dure).

Je voudrai enfin remercier tout le monde et j'espère n'oublier personne. Sur ce bonne lecture...

Notations

d	donnée connue précisément
D	donnée imprécise
$(d_i)_{i \in [1, N]}$	ensemble de N données précises
$(D_i)_{i \in [1, N]}$	ensemble de N données imprécises
$\Pi(A)$	possibilité d'un événement A
$N(A)$	nécessité d'un événement A
$\Pi(A; B)$	possibilité de A restreinte à B
$N(A; B)$	nécessité de A restreinte à B
$ A $	cardinal de A
$P(W; (D_i))$	probabilité d'avoir un événement W en connaissant la suite de données (D_i)
$P(W \cap (D_i))$	probabilité d'avoir un événement W et toutes les données de la suite (D_i)
\cap	intersection de deux ensembles (aussi symbole du ET logique)
\cup	union de deux ensembles (aussi symbole du OU logique)
A^c	ensemble complémentaire à A
μ_A	fonction caractéristique de A
$\overline{P(A)}$	probabilité supérieure d'avoir l'événement A
$\underline{P(A)}$	probabilité inférieure d'avoir l'événement A
P_i	pixel de l'image i
\mathcal{P}_i	quantité floue associée à P_i

Les éléments sont notés par des minuscules, et les ensembles par des majuscules.

Table des matières

Table des matières	v
Table des figures	ix
Liste des tableaux	xv
1 Etat de l’art	5
1.1 Mouvement apparent majoritaire	6
1.2 Mise en correspondance de motifs	7
1.2.1 Méthodes de mise en correspondance de motifs	7
1.2.2 Méthodes alternatives	10
1.3 Approches différentielles	13
1.3.1 Principe	13
1.3.2 Améliorations	15
1.4 Estimation du mouvement par mise en correspondance d’indices	17
1.4.1 Extraction d’indices caractéristiques	17
1.4.2 Mise en correspondance et estimation de mouvement	19
1.4.3 Avantages	19
1.4.4 Inconvénients	20
1.5 Mise en correspondance de formes	20
1.6 Conclusion	21
2 Histogrammes approchés	23
2.1 Histogrammes classiques.	24
2.1.1 Généralités	24
2.1.2 Limites des histogrammes classiques	26
2.2 Principe des histogrammes approchés	31

2.2.1	Partitionnement flou de l'espace	31
2.2.2	Accumulation de données précises	33
2.2.3	Accumulation de données imprécises	33
2.3	Restitution du continu	35
2.3.1	Cas général.	35
2.3.2	Cas particuliers pour le sous ensemble W.	38
2.4	Applications des histogrammes approchés	51
2.4.1	Détermination d'un fractile ou de la médiane	52
2.4.2	Détermination du mode d'une distribution	55
2.5	Généralisation des histogrammes approchés en 2-D puis N-D	65
2.5.1	Histogrammes approchés en dimension 2	65
2.5.2	Evaluation de l'erreur de projection	67
2.5.3	Histogrammes approchés en dimension N	69
2.6	Conclusion	70
3	Estimation du mouvement apparent	73
3.1	Principe de base	74
3.1.1	Modèle de mouvement de la caméra	74
3.1.2	Modélisation floue du niveau de gris des pixels	80
3.1.3	Représentation multi-échelles des images	82
3.2	Evaluation de l'éventualité de déplacement d'un pixel en fonction de son niveau de gris	84
3.3	Estimation de l'éventualité de déplacement d'un pixel en fonction d'une transformation imprécise	86
3.3.1	Modèle de mouvement précis	87
3.3.2	Modèle de mouvement imprécis	90
3.4	Estimation de la probabilité d'un mouvement	94
3.5	Estimation du mouvement dominant	95
3.5.1	Introduction	95
3.5.2	Test de robustesse	99
3.6	Comparaison avec les autres méthodes	103
3.7	Processus multi-échelles d'estimation de mouvement	105
3.7.1	Apport de la représentation multi-échelles	105

3.7.2	Influence de l'estimation du mode de la distribution des mouvements sur le PMR	108
3.8	Vers le temps réel ?	110
3.9	Conclusion et perspectives	111
4	Applications	117
4.1	Estimation d'un mouvement réel	118
4.1.1	Effet des variations de profondeur de champ	118
4.1.2	Effet des variations d'intensité globale	121
4.1.3	Effet des occultations et mouvements parasites	123
4.2	Mosaïques d'images	126
4.2.1	Etat de l'art	126
4.2.2	Résultats obtenus	127
4.3	Stabilisation d'images	132
4.3.1	Quel est le principe ?	132
4.3.2	Résultats obtenus	134
4.4	Conclusion	136
	Conclusion générale et perspectives	137

Table des figures

1.1	Définition du motif de référence sur l'image 1 et recherche de ce motif dans l'image 2.	8
1.2	Seule la composante normale du flot optique est estimable.	14
1.3	Cas d'une surface Lambertienne.	16
2.1	Partition classique du segment $[e_{\min}, e_{\max}]$ de l'axe des réels en dix cellules.	24
2.2	Appartenance de la donnée d à la cellule H_k	25
2.3	Histogramme normalisé s'approchant de la densité de probabilité de la distribution à partir de laquelle il est construit.	25
2.4	-10,10	27
2.5	-9,11	28
2.6	Histogrammes représentant la variable aléatoire X (a) Histogramme à 5 cellules (b) Histogramme à 200 cellules.	29
2.7	Représentation des fonctions caractéristiques des 11 cellules formant une partition binaire de l'axe des réels entre e_{\min} et e_{\max}	32
2.8	Représentation des fonctions caractéristiques des 11 cellules formant une partition floue de l'axe des réels entre e_{\min} et e_{\max}	32
2.9	Vote d'une donnée précise sur un histogramme approché.	33
2.10	Accumulation d'une donnée imprécise sur l'accumulateur supérieur. . .	34
2.11	Accumulation d'une donnée imprécise sur l'accumulateur inférieur. . . .	35
2.12	Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'ensemble W	38
2.13	Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle flou W dont la fonction caractéristique est triangulaire, et de granularité 2Δ	40
2.14	Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle flou W^c	41
2.15	Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle binaire W de granularité Δ	43

2.16	Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle binaire W de granularité 2Δ	44
2.17	Partitionnement flou de l'intervalle $[-10, 10]$ en 11 cellules floues.	46
2.18	Reconstruction par excès et par défaut de la distribution avec un sous-ensemble flou W de fonction caractéristique linéaire.	47
2.19	Reconstruction de la distribution avec un sous-ensemble flou W binaire de support Δ	47
2.20	Reconstruction de la distribution avec un sous-ensemble flou W binaire de support 2Δ	48
2.21	Comparaison de la divergence de Kullback-Leibler pour les trois formes de W	49
2.22	Influence de la position du partitionnement sur l'estimation de la distribution.	49
2.23	En trait plein : la divergence de Kullback-Leibler entre l'estimation de $P(W/(D_i))$ avec le transfert classique et $P(x \in W)$, et en trait pointillé la divergence entre l'estimation de $P(W/(D_i))$ avec transfert des deux bornes de $P(H_k/(D_i))$ et $P(x \in W)$	51
2.24	Calcul du rang de w	53
2.25	Convergence de l'estimation de la médiane vers 25 en fonction du nombre de données.	54
2.26	Convergence de l'estimation du fractile d'ordre 0.28 en fonction du nombre de données.	55
2.27	Détermination du mode local pour les cellules $H_{k-1}, H_k, H_{k+1}, H_{k+2}$	57
2.28	Partition floue et accumulateur associé dans le cas de données précises.	58
2.29	Estimation continue du nombre de votes dans W , et histogramme précis.	59
2.30	Evolution de l'estimation de la position du mode en fonction du nombre de données.	59
2.31	Partition floue et accumulateur précis pour une distribution multi-modale.	60
2.32	Evolution de l'estimation de la position du mode en fonction du nombre de données.	61
2.33	Partition floue et accumulateur associé dans le cas de données imprécises.	63
2.34	Evolution de l'estimation du mode en fonction du nombre de données. Le véritable mode de la distribution se situe en 3.4.	64
2.35	Partition floue et accumulateurs associés pour une distribution multi modale dont le mode principal (70% des données) est en 3.4 et le mode secondaire (30% des données) est en -1.2.	64
2.36	Evolution de l'estimation du mode en fonction du nombre de données. Le véritable mode principal de la distribution se situe en 3.4.	65

2.37	Partition floue en deux dimensions sur l'espace $[e_{\min}, e_{\max}] \cdot [f_{\min}, f_{\max}]$.	66
2.38	Projection de la partition 2-D et de son accumulateur associé.	67
2.39	Accumulateur construit sur une partition de 21 par 21 cellules.	68
2.40	Représentation graphique de la distance de Kullback-Lieber entre la distribution d'origine et la distribution projetée.	69
3.1	Composante du mouvement d'une caméra dans l'espace.	74
3.2	Effet des rotations autour des axes X et Y sur la projection des objets.	75
3.3	Similitude entre une rotation autour de l'axe Y et une translation le long de l'axe X	76
3.4	Variables utilisées pour la modélisation du mouvement apparent.	77
3.5	Effet de l'avance de la caméra sur la projection d'un objet sur le plan image.	77
3.6	Effet du zoom sur un pixel de l'image.	78
3.7	Transformation d'un pixel suite à une rotation autour de l'axe optique de la caméra.	79
3.8	Deux classes duales pour représenter le niveau de gris des pixels : pixels blancs et pixels noirs.	81
3.9	Décomposition multi-échelle des images. Les images du haut représentent les valeurs de μ_b et celles du bas les valeurs de μ_n	83
3.10	Déplacement d'un pixel correspondant au mouvement engendré par (T_x, T_y, R_z)	84
3.11	Déplacement d'un pixel P_1 en dehors du cadre de l'image 2.	86
3.12	Quantité floue représentant un pixel.	87
3.13	Déplacement d'un point de l'environnement dans une séquence vidéo.	88
3.14	Détail du déplacement d'un pixel dû au mouvement de la caméra.	88
3.15	Intersection entre les deux quantités floues associées à P'_1 et P_2	90
3.16	Intersection entre la quantité floue englobante associée à P'_1 et celle associée à P_2	90
3.17	Calcul des possibilité et nécessité de P_2 restreinte à P'_1 en projection horizontale et verticale.	91
3.18	Noeud du maillage et son voisinage.	91
3.19	Effet de l'imprécision des paramètres du modèle de mouvement.	92
3.20	Quantité floue englobante issue du calcul d'erreur.	93
3.21	Image source (a) (comprenant un agrandissement de la zone utilisée), et les deux imageries b (cadre foncé) et c (cadre clair) extraites de celle-ci.	96

3.22	Accumulateurs supérieurs et inférieurs associés aux paramètres du mouvement, représentés en coupes, pour différentes rotations R_ϕ	98
3.23	Estimation de T_{hor} (a), T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images.	98
3.24	Erreur commise pour l'estimation de T_{hor} (a), T_{ver} (b) et de la rotation R_ϕ (c) pour les 100 paires d'images.	99
3.25	Image initiale (a) puis la même image avec une diminution de 25% de la luminosité (b).	100
3.26	Estimation de T_{hor} (a) T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images avec une variation de la luminosité de 25% entre les deux images.	100
3.27	Erreur sur l'estimation des paramètres du mouvement pour 100 paires d'images avec modification de la luminosité de 25% pour la seconde image de la paire.	100
3.28	Image d'un dégradé de niveaux de gris (a) et la même image avec une variation de l'intensité de 16% (b).	101
3.29	Estimation de T_{hor} (a) T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images contenant un mouvement parasite.	102
3.30	Erreur sur l'estimation de T_{hor} (a) T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images avec un mouvement parasite affectant 15% de la surface de l'image.	102
3.31	Zone centrale de l'image 1 contenant les pixels ayant le plus de chance de se retrouver dans l'image 2 suite à une transformation.	106
3.32	Taille minimale de l'image à une résolution donnée fonction de la zone de recherche du mouvement.	107
4.1	Images de la séquence maison : (a) Image 1, (b) Image 2.	119
4.2	Superposition des deux images décalées du mouvement estimé ($T_{hor} = 10.7$; $T_{ver} = 1.7$; $R_\phi = 0.5^\circ$)	119
4.3	Images du laboratoire présentant de fortes variations de profondeur : (a) Image 1, (b) Image 2.	120
4.4	Superposition des deux images décalées.	120
4.5	Images du laboratoire présentant de fortes variations de profondeur ainsi qu'une variation de luminosité ambiante : (a) Image 1, (b) Image 2.	121
4.6	Superposition des deux images décallées avec la variation de luminosité.	122
4.7	Images issues d'une séquence prise en extérieur présentant des variations de luminosité ambiante : (a) Image 1, (b) Image 2.	122
4.8	Superposition des deux images du mur avec variation de luminosité.	123

4.9	4 Images issues d'une séquence prise en intérieur avec présence d'un mouvement parasite.	124
4.10	Superposition des quatre images avec effet de transparence pour indiquer la position de la personne tout au long de la séquence.	124
4.11	Séquence présentant un mouvement parasite majoritaire (mouvement du livre).	125
4.12	Mise en correspondance des deux images avec le mouvement du livre comme mouvement dominant.	125
4.13	Images de la séquence des maisons et du champs de fleurs.	128
4.14	Mosaïque des 150 images de la séquence.	129
4.15	Images d'une séquence visualisant la coque du Titanic sous de mauvaises conditions d'éclairage [IFREMER].	129
4.16	Mosaïque des 326 images de la séquence Titanic Wall.	130
4.17	Images de la séquence du champ d'amphores.	131
4.18	Mosaïque des 300 images de la séquence du champ d'amphores.	131
4.19	a) Séquence instable et b) séquence stabilisée.	133
4.20	Images de la séquence des amphores : n°1 (a), n°10 (b), n°20 (c), n°30 (d).	134
4.21	Images stabilisées de la séquence des amphores : n°1 (a), n°10 (b), n°20 (c), n°30 (d).	135

Liste des tableaux

1.1	Distances classiques pour la mise en correspondance d'images	9
2.1	Classes de taille des individus.	26
3.1	Statistiques sur les erreurs commises.	97
3.2	Statistiques sur les erreurs commises : influence d'une variation globale de luminosité de 25%	101
3.3	Statistiques sur les erreurs commises : influence de mouvements parasites affectant 15% de la surface de l'image.	103
3.4	Statistiques sur les erreurs commises : absence de bruit.	114
3.5	Statistiques sur les erreurs commises : variation globale de luminosité de 25%.	114
3.6	Statistiques sur les erreurs commises : présence d'un mouvement parasite.115	

Introduction générale

La perception est une fonction primordiale pour l'être humain. Elle correspond à une opération par laquelle l'esprit se forme une représentation du monde extérieur en interprétant les données que les sens lui fournissent. Dès qu'un individu interagit avec son milieu, c'est-à-dire cherche à se déplacer, saisir un objet, etc., il lui est nécessaire de percevoir son environnement afin d'agir conformément à ses souhaits.

De nos jours, un grand nombre d'applications trop pénibles, ou irréalisables par l'homme, sont effectuées par des processus automatisés. De la même façon qu'un individu doit percevoir l'environnement avec lequel il interagit, un processus automatisé doit lui aussi avoir un moyen de percevoir son espace de travail. Pour cela, on dispose d'une grande quantité de capteurs que l'on adapte à l'application visée. Selon leur nature, ces capteurs permettent d'exprimer une ou plusieurs caractéristiques de leur milieu (son, image, présence d'objets, champ électrique, etc.).

De manière analogue à l'être humain, la vision artificielle est l'un des moyen de percevoir l'environnement qui fournit le plus d'informations. Quoi de plus naturel et intuitif pour caractériser un environnement qu'une représentation graphique? C'est pourquoi un grand nombre de systèmes automatisés font appel à des capteurs fournissant une représentation visuelle de leur milieu. Une caméra vidéo remplit parfaitement ce rôle. Le mode de perception de la caméra se rapproche en effet beaucoup de la vue humaine; l'interprétation des données qu'elle délivre en est grandement simplifiée par rapport aux capteurs de type télémètres ou capteurs à ultra-sons.

Depuis les vingt dernières années, la vidéo connaît de surcroît un essor considérable. Elle a vu son champ d'applications passer du domaine strictement professionnel au domaine grand public, notamment grâce à la baisse significative du prix des appareils d'enregistrement et de restitution d'images et à leur miniaturisation. La richesse des informations contenues dans une image et la vulgarisation de l'emploi des caméras font que le spectre des applications potentielles de ce capteur est très large.

Parmi les applications les plus courantes, on peut citer par exemple les procédés de reconstruction d'environnements tridimensionnels [Fau93] à partir d'une séquence d'images. Les caméras sont aussi employées par des processus de reconnaissance ou d'identification de caractères manuscrits [Bei93], d'expressions du visage [PMS94], ou encore d'empreintes digitales [JBD93]. On les retrouve également dans des applications plus basiques telles que la surveillance de chaînes de montage dans l'industrie. Dans

ce cas, l'objet de leur utilisation peut être, par exemple, de détecter si une pièce est présente ou non sur un convoyeur. Leur aptitude à représenter les scènes de manière intuitive les rend aussi utiles dans le domaine du diagnostic médical [BKS⁺99]. Elles permettent de mémoriser, lors d'une intervention, des vues d'organes et de parties du corps humain. Un procédé de traitement d'images spécialisé permet ensuite d'identifier des pathologies à partir des images capturées.

L'interprétation automatique des images est aussi nécessaire dans les processus d'indexation vidéo. Cette technique consiste à extraire des indices caractéristiques d'une séquence vidéo (motifs, mouvements, etc.). Cette indexation permet de retrouver des séquences d'images complexes à partir de simples indices linguistiques ou par l'exemple [FB00].

Les caméras en tant que capteurs sont aussi très utilisées dans des applications faisant intervenir un asservissement visuel, tel que les procédés de guidage automatique de véhicules [LWKM95]. Dans ce cas, elles servent à repérer des indices visuels tels que les bords de la chaussée ou des balises (naturelles ou artificielles).

On retrouve cette notion de position et de mouvement dans l'utilisation des caméras en tant que capteurs de localisation relative. Dans ce cas, la caméra se comporte comme un capteur de vitesse utilisant les déformations induites sur les images par son mouvement propre. L'analyse de ces déformations permet de retrouver le mouvement de la caméra. De tels procédés nécessitent généralement une phase de calibrage de la caméra (calibrage intrinsèque) et une phase de repérage de la caméra par rapport à son support (calibrage extrinsèque). Le mouvement de l'ensemble caméra/porteur est généralement déterminé par mise en correspondance d'indices extraits de paires d'images successives. On trouve en robotique plusieurs exemples utilisant l'estimation du mouvement propre d'une caméra fixée sur un robot pour en déduire le mouvement du robot [BCSD97]. Des procédés de réalité augmentée [DKe97] font aussi appel à l'estimation du mouvement propre de la caméra pour caractériser et représenter des scènes.

L'utilisation des caméras fait appel également à des techniques de stabilisation d'images [CFR99]. Il s'agit, dans ce cas, d'estimer le mouvement de la caméra afin de compenser les vibrations et tremblements parasites présents sur le film. Ces tremblements surviennent, par exemple, lorsque l'on filme une scène à partir d'une plate-forme en mouvement. L'intérêt de la stabilisation est de rendre plus "confortable" la visualisation des films.

Depuis quelques années, les scientifiques de terrain sont demandeurs de mosaïques d'images [RPF97]. Le principe en est relativement simple. Il s'agit d'effectuer une succession de prises de vue de l'environnement à l'aide de la caméra, et de les assembler sur un même support en respectant la continuité des motifs. On crée ainsi une vue panoramique de l'environnement filmé. Le but de cette technique est d'augmenter considérablement le champ couvert par une seule image sans en réduire la résolution. On emploie généralement les mosaïques d'images dans le but de réaliser une cartographie de grandes zones.

Dans ces trois dernières applications utilisant des caméras en mouvement, il est

nécessaire d'estimer ce que l'on appelle le **mouvement apparent majoritaire**. Ce problème constitue encore de nos jours un des défis majeurs du traitement d'images. Lorsque la caméra se déplace dans son environnement, les images qu'elle acquiert, c'est-à-dire les projections des points de l'environnement sur la rétine, évoluent dans le temps. On emploie le terme de **mouvement apparent** pour désigner le mouvement de la projection des objets sur la rétine de la caméra. On parle aussi de **mouvement projeté**. Du fait qu'une scène 3-D est composée d'une multitude d'objets ne se situant pas tous à la même distance de la caméra, toutes les projections ne semblent pas se déplacer de la même manière sur l'image. En effet, si l'on filme une scène en se déplaçant, les points de l'environnement les plus distants de la caméra se déplaceront moins vite sur l'image que ceux qui en sont plus proches. De ce fait, il existe plusieurs mouvements apparents "visibles" sur la séquence vidéo. Si l'on regroupe tous les points sur l'image se déplaçant à la même vitesse, le mouvement apparent dominant, ou majoritaire, sera le mouvement correspondant au plus grand nombre de pixels de l'image. Nous présentons, dans ce manuscrit, une nouvelle approche permettant d'estimer le mouvement apparent majoritaire dans une séquence vidéo. Le plan de cet ouvrage est détaillé ci-dessous.

Nous présentons tout d'abord un état de l'art sur les techniques habituellement employées pour estimer le mouvement dominant visible dans une séquence d'images, à savoir, les méthodes fondées sur la mise en correspondance de motifs, les méthodes différentielles de calcul du flot optique et les méthodes basées sur une mise en correspondance de primitives. Nous exposons ces méthodes dans leur formulation la plus simple ainsi que quelques unes des améliorations apportées visant à les rendre plus performantes ou plus robustes. De l'étude de ces méthodes, nous dresserons un constat : la problématique de l'estimation de mouvement est souvent liée à des hypothèses trop contraignantes reliant la variation d'intensité des pixels et le mouvement apparent. La violation fréquente de ces hypothèses fait échouer l'estimation ou réduit la robustesse de ces méthodes. C'est pourquoi nous proposons, dans la suite de ce manuscrit, une approche faisant appel à des théories permettant de mieux gérer les problèmes liés à l'incertitude et à la précision. Nous avons utilisé entre autres : la théorie des sous-ensembles flous, celle des possibilités, de la croyance, du transfert d'erreur et des probabilités imprécises.

Puisque nous cherchons à estimer un mouvement apparent **majoritaire**, l'idée d'utiliser une méthode faisant appel à un "vote" ou un "comptage" nous a semblé naturelle (le terme majoritaire faisant penser à une élection). Nous nous sommes donc tournés vers une utilisation des histogrammes. Cependant, il nous est apparu assez rapidement que les histogrammes classiques ne permettaient pas de gérer l'imprécis **et** l'incertain comme nous le souhaitions. Nous avons donc étendu la notion d'histogramme pour définir un nouvel outil que nous appelons "histogramme approché". Cet outil fait l'objet du second chapitre.

Si la scène filmée est relativement éloignée de la caméra (on parlera de scène quasi-plane), et se trouve être majoritairement statique, le mouvement apparent des pixels de l'image dû au déplacement de la caméra se retrouve grandement simplifié. En effet, sous

ces hypothèses, le mouvement apparent sur l'image peut être caractérisé par un modèle de déplacement contenant peu de paramètres. Nous introduisons dans le troisième chapitre le modèle de mouvement apparent dominant que nous avons retenu ainsi que l'algorithme permettant d'en estimer les paramètres. Nous validons ensuite notre approche en estimant le mouvement sur des séquences simulées. Ces séquences nous permettent de tester notre algorithme dans les conditions optimales d'utilisation. Elles nous ont aussi permis d'insérer des perturbations dans les séquences afin d'évaluer la robustesse de la méthode. Nous avons par exemple rajouté des variations de la luminosité ambiante sur la séquence, ou inséré des déplacements parasites. Ces résultats d'estimation sont ensuite comparés avec ceux obtenus par les approches plus classiques présentées dans le premier chapitre.

Nous proposons ensuite une série d'expérimentations sur des séquences d'images réelles visant à mettre en avant la robustesse de notre méthode. Nous illustrons par exemple l'effet d'une variation d'illumination au cours de la séquence, ou encore l'effet du non respect de la planarité de la scène, sur le processus d'estimation de mouvement. Nous illustrons ensuite l'intérêt de notre approche en l'appliquant à deux problématiques : la création de mosaïques d'images et la stabilisation de séquences vidéo.

Finalement une conclusion générale dresse un bilan des apports de notre approche, et fait ressortir les différentes perspectives de ce travail.

Chapitre 1

Etat de l'art

L'estimation de mouvement est un des axes majeurs de la recherche en traitement d'images et vision assistée par ordinateur [MB96]. Le mouvement est en effet un indice très important pour de nombreuses applications comme la segmentation d'images [GB00] [CTS97], la compression de données [THWL98] ou la construction de mosaïques d'images [PRAZ00] [IA98]. De nombreuses méthodes ont été développées visant à estimer le mouvement soit d'objets appartenant à la scène filmée, soit de la caméra. Dans le premier cas, il s'agit d'extraire l'information du mouvement d'un mobile présent dans la séquence vidéo acquise. Cela met en oeuvre des techniques de segmentation pour isoler l'objet puis d'estimation de mouvement pour caractériser la nature de son déplacement. Il s'agit là d'une estimation locale puisque l'objet n'occupe généralement pas la totalité de l'image. Dans le deuxième cas, il s'agit d'estimer le mouvement de la caméra à partir de la séquence vidéo. Sous certaines conditions, le mouvement visible sur la séquence acquise peut être relié au mouvement propre de la caméra. C'est dans ce cadre que s'inscrit notre travail. Nous nous intéressons en fait à l'estimation du mouvement apparent dominant.

Le premier paragraphe explicite la notion de mouvement apparent dominant en définissant chacun des termes. La suite de ce chapitre présente les méthodes habituellement employées pour l'estimation du mouvement apparent. Pour chacune de ces méthodes, nous présentons le principe de base ainsi que les diverses améliorations apportées. Nous avons classé ces méthodes en deux groupes. Les approches dites de "bas niveau" regroupant les méthodes basées sur le traitement du signal et celles dites de "haut niveau" nécessitant une interprétation de l'image. Traitant des approches de bas niveau, le paragraphe trois recense les différentes méthodes apparentées à la corrélation. Le paragraphe quatre présente ensuite les méthodes basées sur un calcul différentiel du flot optique. En ce qui concerne les approches de haut niveau, le paragraphe cinq offre un aperçu des techniques utilisant la mise en correspondance de points caractéristiques de l'environnement. Le paragraphe six traite des méthodes utilisant des mises en correspondance de formes complexes. Nous finirons par une conclusion notant l'intérêt d'une nouvelle approche utilisant une représentation des niveaux de gris moins contraignante.

1.1 Mouvement apparent majoritaire

Une séquence d'images numériques obtenue par une caméra filmant une scène réelle est composée d'une succession de matrices de points appelés pixels. A ces pixels sont associés un ou plusieurs nombres appelés niveau de gris (pour les images en niveaux de gris) et couleur (pour les images en couleurs) variant de façon discrète au cours du temps. Nous nous intéressons ici plus particulièrement aux images en niveaux de gris.

Le niveau de gris (ng) d'un pixel est fonction de la quantité de lumière reçue pendant le temps d'exposition par le capteur (rétine) associée à ce pixel [HM95]. Cet éclairage provient de réflexions et de réfractions de la lumière ambiante sur la scène. Les variations au cours du temps de cette quantité de lumière peuvent avoir de nombreuses causes que nous regroupons en deux catégories.

La première catégorie regroupe les effets liés à des déplacements, soit de la caméra, soit des objets présents dans la scène [CS95]. La deuxième catégorie rassemble les effets qui ne sont pas dûs à des mouvements. On les apparentera à des phénomènes relevant plus des "perturbations". Ainsi, il existe de multiples sources que l'on peut classer dans les perturbations telles que des variations de l'illumination ambiante, une variation de la quantité de lumière réfléchiée par une surface à la suite d'un changement d'orientation, des occlusions, des ombres, etc. Dans cet état de l'art, nous nous focaliserons sur l'étude des méthodes d'estimation du mouvement des objets dans l'image, ou bien de la scène tout entière.

Il faut maintenant faire la distinction entre le mouvement réel des objets (mouvement 3-D) et le mouvement de la projection de ces objets dans l'image. Même si l'image résulte d'une projection des objets 3-D sur le plan image, les mouvements visibles sur la séquence vidéo, ne correspondent pas nécessairement à la projection des mouvements réels dans le plan de l'image. Par exemple considérons une source de lumière se déplaçant au dessus d'une scène statique. Les ombres des objets vont se déplacer sur l'image sans que les objets se déplacent. On est donc dans le cas où aucun objet de la scène ne bouge et pourtant on perçoit un mouvement sur la séquence vidéo. On peut aussi avoir le phénomène inverse. En effet, considérons l'exemple bien connu d'une sphère ne possédant pas de texture sous un éclairage ne provoquant pas d'ombre. Si cette sphère est animée d'un mouvement de rotation sur elle même, le champ de vitesse 3-D de ses points n'est pas identiquement nul. Pourtant, sur la séquence aucun mouvement ne sera visible. Il sera donc impossible de déduire quoi que ce soit du champ de vitesse réel en fonction uniquement des images. Ces deux exemples illustrent le fait que tous les mouvements 3-D ne sont pas indentifiables sur la séquence d'images et vice-versa tous les mouvement visibles sur la séquence ne correspondent pas nécessairement à un déplacement réel des objets 3-D. Ce sont les déplacements "visibles" sur la séquence d'images qui sont appelés **mouvements apparents**.

Dans ce travail nous nous sommes focalisés sur l'étude du mouvement apparent **majoritaire** ou **dominant**. Que signifient ces termes ? Si on considère que l'on peut segmenter l'image en fonction du mouvement, c'est-à-dire regrouper en "zones" tous les points possédant le même mouvement dans l'image, alors le mouvement majoritaire

est celui qui est associé au plus grand nombre de pixels. C'est à ce type de mouvement que nous nous sommes intéressés dans ce travail.

Plusieurs méthodes, largement éprouvées, ont été utilisées pour l'estimation du mouvement apparent (majoritaire ou non). Parmi les plus connues, on note celles apparentées à la corrélation que l'on regroupe sous le nom de méthodes de mise en correspondance de motifs ou d'icônes. On rencontre aussi les approches différentielles souvent appelées méthodes de flot optique. D'autres approches plus haut niveau font appel à une extraction de points caractéristiques ou de primitives complexes pour faire une poursuite de ces attributs tout au long de la séquence d'images. Une comparaison de l'efficacité des différentes approches est proposée dans [JLBB94] dans le cas de mouvements simples sans occlusions ni discontinuités. Cette liste n'est bien sur pas exhaustive. Il existe encore bien d'autres méthodes comme les méthodes utilisant l'énergie ou la phase [MB96] .

REMARQUE 1 Il est important de noter que ces méthodes ne se limitent pas uniquement à l'estimation du mouvement apparent majoritaire. Certaines fournissent au contraire une estimation dense du mouvement apparent. L'information de vitesse est alors disponible en tout point de l'image et donc pour chaque objet se trouvant sur l'image.

Nous avons constaté que peu de travaux en traitement d'images, et plus particulièrement en matière d'estimation de mouvement apparent, utilisent les techniques permettant de traiter l'imprécision des données. Nous avons cependant recensé des méthodes de traitement d'images basées sur des approches géométriques [PG92] [Ros84] ou morphologiques [BM95] [BKG95] faisant appel à la théorie de l'incertain. D'autres méthodes, notamment des techniques de segmentation d'images [Bez81] [KK96], des techniques d'indexation [DLZ95], ou bien de restauration et d'amélioration d'images [KKBMM98] font aussi appel à une modélisation de l'imprécision des données. Le domaine d'application le plus important de cette théorie reste néanmoins la fusion de données [AG92].

1.2 Mise en correspondance de motifs

1.2.1 Méthodes de mise en correspondance de motifs

Dans un premier temps il est nécessaire de faire la différence entre la corrélation au sens mathématique et ce qui est abusivement appelé méthodes apparentées à la corrélation. Dans le monde du traitement du signal, la corrélation entre deux signaux mono-dimensionnels est définie par :

$$C_{1,2}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} X_1(t) X_2(t + \tau) dt \quad (1.1)$$

soit dans le cas discret :

$$C_{1,2}(k) = \frac{1}{N} \sum_{i=0}^N X_1(i) X_2(i+k) \quad (1.2)$$

Dans le milieu du traitement de l'image, un grand nombre de techniques sont regroupées sous le nom de méthodes de corrélation sans pour autant se référer à la formule (1.2). Les méthodes présentées dans ce chapitre sont en fait des méthodes de mise en correspondance de motifs.

L'estimation de mouvement à l'aide de ces méthodes repose sur le principe suivant. Soit à estimer le mouvement entre une image 1 et une image 2. On définit une zone dans la première image que l'on appellera motif de référence. Généralement les motifs sont des fenêtres ou ensembles rectangulaires de pixels de taille réduite par rapport à l'image. (Il est cependant possible de considérer l'image dans son intégralité pour la mise en correspondance.) Les méthodes de mise en correspondance de motifs (pattern matching) visent ensuite à retrouver dans la deuxième image le même motif de référence. Cette recherche s'effectue soit localement dans une petite zone de l'image soit globalement sur toute l'image. La recherche du motif de référence, appelé encore "modèle", dans la seconde image met en oeuvre un calcul de distance entre les deux fenêtres de pixels. Cela est illustré sur la figure 1.1.

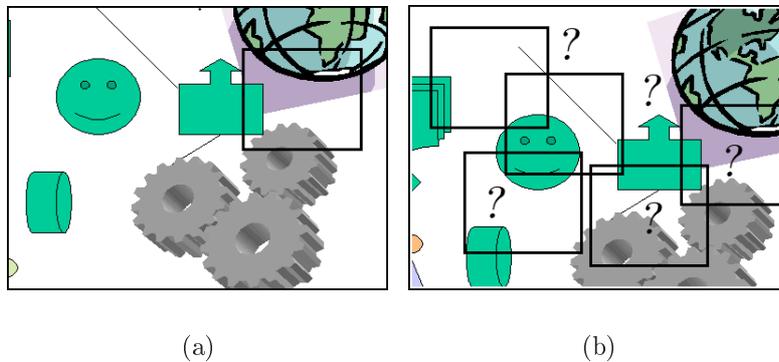


Figure 1.1 – Définition du motif de référence sur l'image 1 et recherche de ce motif dans l'image 2.

Les distances les plus couramment utilisées sont rappelées dans le tableau 1.1. Une étude comparative de ces différentes distances est proposée dans [AG93].

Dans le tableau 1.1, X_1 et X_2 correspondent aux deux images considérées et W définit les bornes du motif de l'image 1. La quantité (u, v) correspond au mouvement dans le plan image dont on cherche à tester la validité. Généralement ces techniques sont employées pour estimer des mouvements constitués de translations comme par exemple le décalage entre deux images d'une paire stéréo [JBC97]. Dans ce cas le déplacement (u, v) n'est pas fonction du pixel considéré. Par contre, si le modèle de mouvement contient une rotation, alors (u, v) dépend de la position du pixel dans l'image. Les formules présentées dans le tableau 1.1 ne sont donc plus valables, il faut remplacer

Somme des écarts absolus	SAD	$\sum_{(i,j) \in W} X_1(i, j) - X_2(i + u, j + v) $
Somme des écarts absolus centrés	ZSAD	$\sum_{(i,j) \in W} (X_1(i, j) - \bar{X}_1) - (X_2(i + u, j + v) - \bar{X}_2) $
Somme des écarts quadratiques	SSD	$\sum_{(i,j) \in W} (X_1(i, j) - X_2(i + u, j + v))^2$
Somme des écarts quadratiques centrés	ZSSD	$\sum_{(i,j) \in W} ((X_1(i, j) - \bar{X}_1) - (X_2(i + u, j + v) - \bar{X}_2))^2$
Intercorrélation normalisée	NCC	$\frac{\sum_{(i,j) \in W} X_1(i,j) \times X_2(i+u,j+v)}{\sqrt{\sum_{(i,j) \in W} X_1^2(i,j) \sum_{(i,j) \in W} X_2^2(i+u,j+v)}}$
Intercorrélation normalisée centrée	ZNCC	$\frac{\sum_{(i,j) \in W} (X_1(i,j) - \bar{X}_1) \times (X_2(i+u,j+v) - \bar{X}_2)}{\sqrt{\sum_{(i,j) \in W} (X_1(i,j) - \bar{X}_1)^2 \sum_{(i,j) \in W} (X_2(i+u,j+v) - \bar{X}_2)^2}}$

Tableau 1.1 – Distances classiques pour la mise en correspondance d'images

(u, v) par $(u(i, j), v(i, j))$.

Lorsque l'on cherche à régulariser le mouvement sur l'image (pour améliorer la précision de l'estimation et supprimer les fausses estimations locales), ce type de méthode implique l'utilisation d'un modèle paramétrique discret du mouvement. Le mouvement recherché correspond à la combinaison de paramètres ayant obtenu la plus petite distance (ou le coefficient de corrélation le plus élevé) entre les motifs issus des images 1 et 2 [Ana89], [Sin92].

Les méthodes de mise en correspondance de motifs présentent l'avantage d'une grande simplicité d'implantation logicielle ou matérielle avec un faible paramétrage [LZ99]. De plus, elles ont fait l'objet de nombreuses études [Ana89] [Sin92] [JBC97] et leurs performances sont bien connues [AG93].

Cependant l'utilisateur de ce type de méthodes est confronté à plusieurs difficultés. Tout d'abord, l'estimation de déplacement nécessite une discrétisation de l'espace des paramètres du mouvement. La précision de l'estimation est alors contrainte par l'échantillonnage de l'espace paramétrique. En effet, sans interpolation, la précision de l'estimation est égale au pas d'échantillonnage de l'espace paramétrique. Pour illustrer ce problème considérons une séquence de 100 images. Supposons que le vrai mouvement entre deux images de la séquence soit une translation de 1.2 pixels. Si le pas d'échantillonnage de l'espace paramétrique est de 1 pixel (ce qui est souvent le cas pour l'estimation de translations), l'estimation donnera au mieux un mouvement de 1 pixel. Au bout de la centième image l'erreur accumulée sera donc de 100×0.2 pixels soit 20 pixels!

Le deuxième problème auquel est confronté l'utilisateur est le type de mouvement que l'on peut estimer avec ce genre de méthodes. Généralement, elles sont employées dans le cas d'estimation de mouvements ne comportant que des translations [JBC97]. En effet, la discrétisation de l'espace des mouvements est implicite lorsqu'il s'agit de translation (le pas d'échantillonnage est pris égal à un multiple entier de pixel). En revanche, elle ne l'est pas lorsque le mouvement à estimer comporte des rotations,

c'est-à-dire des déplacements non-entiers. Doit on prendre un échantillonnage tous les degrés, tous les 5 degrés, ou tous les 0.5 degrés pour l'espace angulaire ? Ce choix est complexe car l'effet d'une rotation sur le mouvement des pixels n'est pas le même pour tous les pixels. En effet un pixel "près" de la projection du centre de rotation sur l'image ne se déplacera pratiquement pas. Par contre un pixel éloigné de ce centre aura un déplacement important. Du fait des rotations, et plus particulièrement des termes en sinus et cosinus, le déplacement des pixels n'est pas entier. Dans ce cas, quels pixels doivent être mis en correspondance pour le calcul de la distance ? Généralement, ces mouvements sont tronqués pour correspondre à des valeurs entières de pixels et ainsi faciliter la mise en correspondance. Il en résulte encore une perte de précision vis-à-vis de l'estimation du mouvement, et un biais considérable lors de l'estimation du mouvement tout au long d'une séquence vidéo.

Comme on peut le voir dans les formules du tableau 1.1, l'estimation du mouvement se fait en fonction d'une distance entre les niveaux de gris des pixels. Le processus d'estimation est donc fortement sensible à la moindre variation parasite de ces niveaux de gris. Ainsi toute variation d'illumination globale ou locale dans la séquence d'images risque de fausser le processus d'estimation. Ces méthodes sont donc très sensibles aux variations parasites des niveaux de gris des pixels.

Il existe encore un problème concernant l'emploi des méthodes de corrélation. Si un motif de l'image 1 disparaît complètement ou partiellement dans l'image 2 suite à une occlusion, le processus d'estimation fournit quand même un extremum du critère de mise en correspondance. Celui-ci correspondra au mouvement le plus probable au sens de la distance utilisée, mais ne représente pas obligatoirement le mouvement du motif. Il est donc nécessaire de fixer arbitrairement un seuil au-delà duquel la mise en correspondance est rejetée car trop peu fiable.

1.2.2 Méthodes alternatives

De nombreuses ont tenté de palier les problèmes des approches classiques. Dans [SLH00] les auteurs montrent que l'emploi de nouvelles métriques permettent d'obtenir de meilleures performances vis-à-vis du bruit sur les niveaux de gris des pixels. Ils proposent par exemple d'utiliser les distances de Cauchy ($D_c = \sum_i \log \left(1 + \frac{(u_i - v_i)^2}{a} \right)$) ou de Kullback ($D_k = \sum_i u_i \log \left(\frac{u_i}{v_i} \right)$). Dans [HKR93] la distance de Hausdorff est utilisée pour faire de la mise en correspondance entre un modèle et une image avec de meilleurs résultats que la corrélation en terme de robustesse vis-à-vis des perturbations sur les niveaux de gris des pixels. Bien que cette distance ne soit pas employée pour de l'estimation de mouvement dans [HKR93], les auteurs assurent qu'il est très simple d'en étendre le principe à une telle application.

Pour s'affranchir des problèmes liés à la dépendance de l'estimation du mouvement par rapport aux fluctuations des niveaux de gris, plusieurs travaux [ZW94], [BN96] [JBC97] proposent d'utiliser des transformations locales non paramétriques des images. Il s'agit en fait de remplacer l'information de niveau de gris des pixels par une fonction

du rang du pixel dans l'image. Ces méthodes impliquent un tri préalable des pixels par rapport à leur niveau de gris.

Transformation Rank

[ZW94] présente la transformation Rank. Son principe est le suivant : autour de chaque pixel de l'image on définit un voisinage, puis on compte le nombre X de pixels du voisinage ayant une intensité plus faible que le pixel central. La valeur du niveau de gris du pixel central est alors remplacée par X . On obtient ainsi un tableau d'entiers allant de 0 à $N-1$, N étant le nombre d'éléments dans le voisinage considéré. Si on considère une portion d'une image définie par a), et un voisinage 3×3 défini autour de chaque pixel, la transformation Rank transforme a) en b) comme illustré dans (1.3).

$$\begin{array}{c}
 \begin{array}{|c|c|c|c|c|}
 \hline
 100 & 180 & 73 & 156 & 111 \\
 \hline
 84 & 123 & 54 & 184 & 191 \\
 \hline
 53 & 49 & 20 & 38 & 66 \\
 \hline
 67 & 29 & 157 & 11 & 5 \\
 \hline
 50 & 30 & 15 & 25 & 32 \\
 \hline
 \end{array}
 \quad
 \begin{array}{|c|c|c|c|c|}
 \hline
 ? & ? & ? & ? & ? \\
 \hline
 ? & 7 & 3 & 7 & ? \\
 \hline
 ? & 2 & 1 & 3 & ? \\
 \hline
 ? & 1 & 8 & 1 & ? \\
 \hline
 ? & ? & ? & ? & ? \\
 \hline
 \end{array}
 \end{array}
 \quad (1.3)$$

Une fois les images transformées, on applique une métrique type SAD ou SSD pour mesurer la similarité entre les deux motifs. Les auteurs montrent que cette transformation rend plus robuste l'estimation de mouvement vis-à-vis des perturbations de luminosité. En effet, la mise en correspondance ne dépend plus directement de la valeur du niveau de gris du pixel mais de son ordre local.

Transformation d'ordre

La transformation proposée dans [BN96] remplace la valeur du niveau de gris des pixels d'un motif par le rang du niveau de gris dans le motif. Ainsi le motif de taille 5×5 dont les niveaux de gris sont donnés dans (1.4) a) est remplacé par le motif de (1.4) b) comportant le rang des pixels.

$$\begin{array}{c}
 \begin{array}{|c|c|c|c|c|}
 \hline
 100 & 180 & 73 & 156 & 111 \\
 \hline
 84 & 123 & 54 & 184 & 191 \\
 \hline
 53 & 49 & 20 & 38 & 66 \\
 \hline
 67 & 29 & 157 & 11 & 5 \\
 \hline
 50 & 30 & 15 & 25 & 32 \\
 \hline
 \end{array}
 \quad
 \begin{array}{|c|c|c|c|c|}
 \hline
 18 & 23 & 16 & 21 & 19 \\
 \hline
 17 & 20 & 13 & 24 & 25 \\
 \hline
 12 & 10 & 4 & 9 & 14 \\
 \hline
 15 & 6 & 22 & 2 & 1 \\
 \hline
 11 & 7 & 3 & 5 & 8 \\
 \hline
 \end{array}
 \end{array}
 \quad (1.4)$$

L'auteur part de l'hypothèse qu'une modification aléatoire ou non des niveaux de gris des pixels d'un motif ne perturbe que très peu le rang relatif de leurs niveaux de gris. En remplaçant la valeur du niveau de gris par son rang dans le voisinage choisi pour la comparaison, on obtient donc un indice beaucoup plus robuste.

Par contre, on ne peut plus utiliser dans ce cas les distances usuelles. On est obligé de comparer les motifs en comparant la permutation des ordres nécessaires pour passer

du motif ordonné de l'image 1 au motif ordonné de l'image 2. Si le motif est inchangé, la permutation est la permutation identité ($1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, \dots, N \rightarrow N$). Sinon, les auteurs définissent une distance pour estimer la similitude de la permutation courante avec la permutation identité. Une étude comparative des performances est proposée entre les méthodes classiques (SSD et NCC) et les deux transformations de rang (Ordre et Rank). Cette transformation permet de s'affranchir partiellement des problèmes de variation d'intensité lumineuse des pixels. En effet, si la totalité des pixels voient leur intensité diminuer ou augmenter de la même manière (modification du facteur gamma), leur rang ne changera pas. Le calcul de similarité restera donc identique. De même, une faible variation du niveau de gris d'un pixel laisse inchangée "l'image des rangs". Par exemple dans le motif présenté en exemple, le niveau de gris du pixel central peut varier de 15 jusqu'à 25 sans pour autant changer de rang.

Transformation Census

La transformation non-paramétrique CENSUS [ZW94] consiste à remplacer le motif des niveaux de gris (ng) par un motif binaire de classification. Cette classification sépare localement les niveaux de gris en deux sous-ensembles : les ng inférieurs au ng de référence (représentés par un 1), et les ng supérieurs au ng de référence (représentés par un 0). La référence est prise au centre du voisinage considéré. Appliquée à l'exemple que nous proposons, la transformation du motif (1.5) a) est (1.5) b) en considérant l'ordre d'apparition des pixels dans le motif lignes par lignes.

$$\begin{array}{l}
 \begin{array}{|c|c|c|c|c|}
 \hline
 100 & 180 & 73 & 156 & 111 \\
 \hline
 84 & 123 & 54 & 184 & 191 \\
 \hline
 53 & 49 & 20 & 38 & 66 \\
 \hline
 67 & 29 & 157 & 11 & 5 \\
 \hline
 50 & 30 & 15 & 25 & 32 \\
 \hline
 \end{array} \\
 a) \\
 \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}
 \hline
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
 \hline
 \end{array} \\
 b)
 \end{array} \tag{1.5}$$

Pour comparer deux motifs, il faut donc effectuer la transformation CENSUS des deux motifs, et comparer les deux chaînes binaires à l'aide d'une distance ensembliste. Dans l'article [ZW94], c'est la distance de Hamming qui est utilisée afin de comparer les deux chaînes binaires. La distance de Hamming compare deux sous-ensembles décrits par leur fonction caractéristique. Elle consiste en une comparaison bit à bit des deux chaînes à l'aide d'un opérateur "ou exclusif", et en la sommation de tous les bits de cette comparaison. La formulation mathématique de la distance de Hamming est rappelée dans (1.6).

$$d_h(I_1, I_2) = \sum_{i=1}^N I_1(i) \oplus I_2(i) \tag{1.6}$$

avec I_1, I_2 les transformées des motifs 1 et 2 et \oplus l'opérateur "ou exclusif". La transformation amenant I_1 en I_2 tel que $d_h(I_1, I_2)$ est minimale sera dite correspondre au

mouvement recherché.

Cette méthode de mise en correspondance est plus robuste vis-à-vis des distorsions radiométriques que les approches classiques type SAD etc. D'après les essais réalisés dans [JBC97], la méthode CENSUS donne de très bons résultats, surtout lorsque les images de la séquence présentent de grandes variations de luminosité.

Cependant, malgré ces améliorations, le problème lié au seuillage du critère de mise en correspondance est toujours présent.

1.3 Approches différentielles

1.3.1 Principe

L'approche différentielle réalise l'estimation du mouvement en mettant en relation les dérivées spatiales et temporelles des images d'une séquence. Cette méthode a tout d'abord été décrite par Horn et Shunck dans [HS81]. Elle a fait l'objet de nombreuses extensions et modifications [Neg98] [FSA01]. Elle repose sur une propriété de conservation de l'illumination des points. Sous les hypothèses de petits mouvements et de différentiabilité de l'image de luminance, cette propriété se traduit généralement par l'équation de contrainte de mouvement au premier ordre :

$$I_x \times u + I_y \times v + I_t = \nabla I \cdot [u, v]^T + I_t = 0 \quad (1.7)$$

où I_x , I_y , I_t , u et v représentent respectivement : les composantes horizontale et verticale du gradient spatial de l'intensité ∇I , le gradient temporel de l'intensité et les composantes horizontale et verticale du vecteur vitesse. On désigne par \cdot le produit scalaire de deux vecteurs. Les quantités I_x , I_y , I_t sont estimées en utilisant uniquement les informations issues de la séquence d'images. Ces estimations sont obtenues de façon classique en convoluant les séquences d'images avec des filtres dérivateurs. En chaque point (i, j) de l'image on obtient une équation de contrainte spatio-temporelle de son déplacement $[u(i, j), v(i, j)]^T$. Une seule contrainte ne permet cependant pas de déterminer les deux composantes u et v du vecteur vitesse. En effet, seule la composante de la vitesse parallèle au gradient de l'image peut être déterminée (*cf.* figure 1.2), le produit scalaire de deux vecteurs perpendiculaires étant nul.

Pour lever l'indétermination, on est obligé de régulariser, c'est-à-dire de supposer une certaine uniformité de répartition du champ de vitesse. Ceci permet d'obtenir l'équation manquante. En fait, généralement, on crée un système redondant d'équations pour lever les indéterminations. D'autres techniques reviennent à ajouter plusieurs contraintes de conservation locales ou globales sur des grandeurs telles que le contraste, la couleur, les courbures, ou les dérivées [Sch86] [LGJ94] [Tis96], et à écrire une équation de type (1.7) pour chacune de ces grandeurs. La technique la plus fréquemment décrite [HS81] consiste à introduire une contrainte de continuité du champ de vitesse. En effet, sous réserve de ne pas se trouver sur les zones de discontinuité de mouvement, les pixels voisins possèdent le même déplacement. Le problème se réduit donc à minimiser une

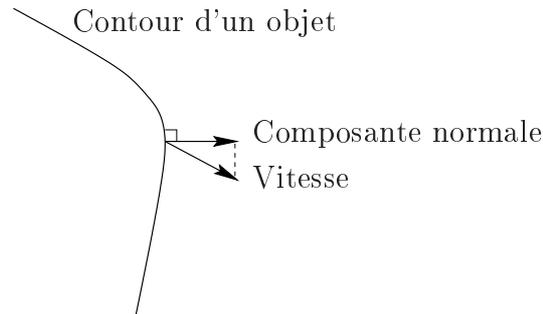


Figure 1.2 – Seule la composante normale du flot optique est estimable.

fonction de coût à deux termes : l'un attaché aux données $C_d(u, v)$ et l'autre au lissage $C_l(u, v)$:

$$E(\mathbf{u}, \mathbf{v}) = C_d(\mathbf{u}, \mathbf{v}) + \lambda C_l(\mathbf{u}, \mathbf{v}) \quad (1.8)$$

avec $\lambda > 0$ et

$$C_d(\mathbf{u}, \mathbf{v}) = \sum_i \sum_j (I_x \times \mathbf{u} + I_y \times \mathbf{v} + I_t)^2 \quad (1.9)$$

$$C_l(\mathbf{u}, \mathbf{v}) = \sum_i \sum_j (u_x)^2 + (u_y)^2 + (v_x)^2 + (v_y)^2 \quad (1.10)$$

où u_x, u_y, v_x, v_y sont les gradients horizontaux et verticaux de la vitesse apparente. Il s'agit donc de trouver le couple (u, v) minimisant $E(u, v)$. La résolution de cette équation multi-contraintes fait généralement l'objet d'un traitement itératif. Deux approches sont alors envisageables, l'approche locale et l'approche globale. L'approche locale réalise une minimisation de $E(u, v)$ en ne prenant en compte que des pixels contenus dans une fenêtre, alors que l'approche globale tient compte de tous les pixels de l'image. Les propriétés de cette méthode sont illustrées dans [HS81]. Les auteurs concluent cependant que cette méthode ne délivre qu'une mesure imprécise du flot optique. Une fois les (u, v) calculés sur l'image, il suffit de chercher les paramètres du modèle de déplacement leur correspondant. Cette estimation est généralement réalisée à l'aide d'un processus de type moindres carrés ou d'une méthode itérative.

REMARQUE 2 *Dans la plupart des méthodes les (u, v) ne sont pas réellement calculés, on préfère estimer directement les paramètres du mouvement qui sont fonction de ces termes.*

Un des principaux avantages des méthodes différentielles est qu'elles ne nécessitent pas de mise en correspondance de points d'une image à l'autre. Cette étape est souvent fastidieuse et source d'erreurs dans les méthodes appariant des points caractéristiques. De plus, contrairement aux méthodes basées sur la corrélation, aucun échantillonnage de l'espace paramétrique du mouvement n'est nécessaire. La précision de l'estimation du mouvement n'est donc pas contrainte par cet échantillonnage.

L'approche différentielle possède cependant quelques inconvénients [KTB87]. Dans un premier temps, pour que l'équation (1.7) soit utilisable, il faut que l'image soit suffisamment texturée. En effet, si on imagine une image unie, les gradients spatiaux et temporel sont nuls, et il est impossible de distinguer un mouvement. En revanche, l'estimation des gradients s'effectue au travers de filtres lisseurs (passe-bas) qui requièrent une faible texture de l'image pour être valides. En effet, si la texture de l'image est trop importante, on aura de fortes variations d'intensité et la valeur estimée de la dérivée sera fautive du fait du lissage. Nous sommes donc en présence de deux hypothèses contradictoires concernant la texture de l'image.

L'hypothèse de conservation de l'illumination globale (deuxième terme de l'équation (1.7) égal à 0) est une hypothèse très restrictive qui est fréquemment violée. En effet si on considère par exemple des scènes en extérieur, le passage d'un nuage devant le soleil peut faire varier l'intensité globale de l'image acquise. De même, un grand nombre de caméras ou de WebCams sont équipées d'un contrôle automatique d'ouverture du diaphragme pour éviter les phénomènes de sur-exposition ou les images trop sombres. La luminosité globale n'est donc pas nécessairement conservée entre deux images d'une séquence vidéo. Il est donc nécessaire de modéliser les variations de l'intensité globale. Ce calcul nécessite l'utilisation d'un modèle de réflectance de la scène ou de la surface filmée [Fau93].

Généralement, dans les approches différentielles, il est fait une hypothèse sous-jacente selon laquelle l'ensemble des points utilisés pour lisser le champ de vitesse appartient à un même objet. Cette hypothèse est évidemment fautive en présence de discontinuités du mouvement apparent. Par discontinuité on entend une frontière entre les projections de différents objets se mouvant indépendamment.

La dernière limitation est liée au fait que l'amplitude des mouvements estimables doit rester faible. En effet, l'estimation des gradients n'est valable que si le déplacement estimé est inférieur à la taille du masque utilisé pour le calcul du gradient.

1.3.2 Améliorations

La littérature présente un certain nombre de techniques permettant de s'affranchir, au moins partiellement, des problèmes causés par la non-vérification des hypothèses de travail dans le cas du flot optique. Il est possible de réduire les incohérences dues aux discontinuités soit en exploitant la coïncidence entre les discontinuités de mouvement et les frontières photométriques, soit en utilisant une segmentation du champ de vitesse estimé en régions homogènes au sens du mouvement [OC96]. Dans [Shu86] une représentation en coordonnées polaires du flot optique est présentée comme possédant de meilleures performances vis-à-vis des discontinuités.

Pour s'affranchir de la contrainte des petits mouvements, la plupart des méthodes s'accordent pour employer une approche pyramidale. Ainsi l'estimation de mouvement se fait à une résolution fortement dégradée (où le mouvement est quasiment nul) puis est rétropropagée de résolution en résolution jusqu'à la résolution la plus fine [MP98], [SK99]. Ce processus permet d'estimer des déplacements de grande échelle sans sortir

des limites d'utilisation des filtres dérivateurs nécessaires à l'estimation des champs locaux. Cependant la dégradation de la résolution des images risque de produire des images quasiment uniformes. Dans ce cas l'estimation des gradients fournit une information biaisée, et l'estimation du mouvement est rendue imprécise.

Une autre source d'erreur vient du fait que l'illumination globale est considérée constante entre deux images successives. Dans [OB95] et [WM95], les auteurs tiennent compte des variations de l'intensité globale. Ainsi l'équation (1.7) s'écrit :

$$I_x \times u + I_y \times v + I_t = \nabla I \times v + I_t = \xi \quad (1.11)$$

avec ξ une variable fonction de la variation d'illumination de la scène. Olivier Faugeras étudie dans [Fau93] le calcul de ξ dans le cas spécifique où la surface de la scène est Lambertienne (c'est-à-dire elle reflète la lumière identiquement suivant toutes les directions). Une telle surface possède en tout point M la propriété suivante :

$$I(m, t) = \rho \mathbf{S}(M, t) \cdot \mathbf{N}(M) \quad (1.12)$$

avec $I(m, t)$ l'intensité reçue par la caméra en m (image du point M) à l'instant t , $\mathbf{S}(M, t)$ le vecteur unitaire dans la direction de la source de lumière, $\mathbf{N}(M)$ la normale en M à la surface et ρ un coefficient constant (voir figure 1.3).

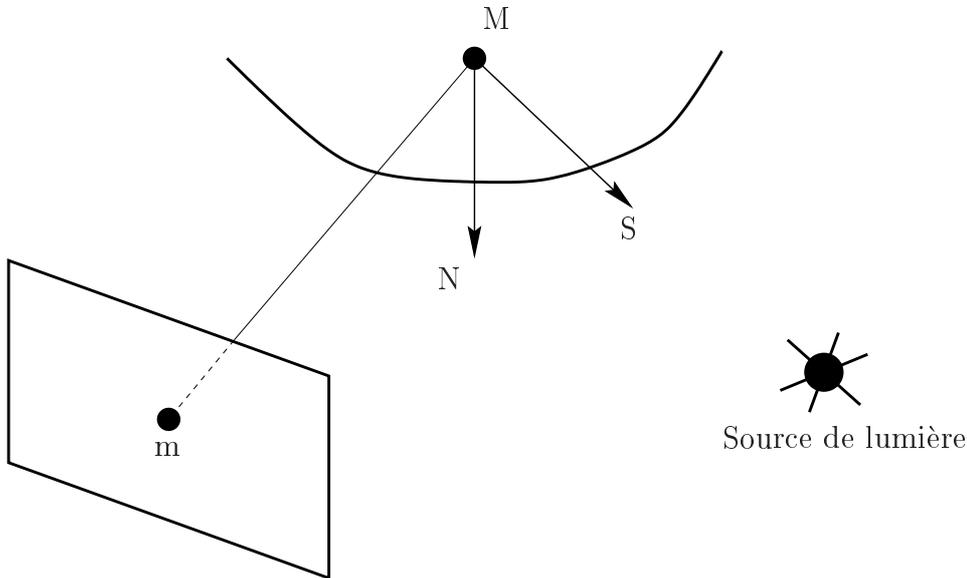


Figure 1.3 – Cas d'une surface Lambertienne.

L'utilisation de cette équation permet de définir le terme ξ dans l'équation (1.11). Une autre approche est présentée dans [Neg98], où l'auteur propose un modèle pour l'évolution de l'illumination globale incluant un facteur multiplicatif et un offset. On a donc $I_g(t + dt) = K(t)I_g(t) + C(t)$, où $I_g(t)$ est l'illumination globale à un instant t .

Finalement, comme l'estimation des deux quantités (u, v) (ou des paramètres du modèle de mouvement) est fortement bruitée [HS81] par l'estimation des gradients,

de nombreuses méthodes ont été développées utilisant des statistiques robustes pour rejeter les estimations aberrantes. Ainsi des méthodes basées sur les moindres carrés robustes ou sur la moindre médiane carrée, sont utilisées pour l'estimation du mouvement [MP98], [WM95], [OB95].

1.4 Estimation du mouvement par mise en correspondance d'indices

Une autre méthode d'estimation de mouvement fait appel à l'extraction d'indices caractéristiques des images suivie d'une mise en correspondance. Ces méthodes sont classées dans les approches haut niveau car elles nécessitent une interprétation de l'information contenue dans l'image. Ces indices correspondent en fait à des éléments remarquables des scènes filmées comme par exemple des lignes, des segments, des coins, des contours, mais aussi des points surbrillants, des points de plus forte courbure etc. L'intérêt de tels indices repose sur le fait que leurs propriétés se conservent tout au long de la séquence. Leur mise en correspondance a donc des chances d'être plus robuste que celle réalisées par les méthodes apparentées à la corrélation.

Généralement le choix des indices dépend de l'application et du type de milieu duquel sont issues les images. Par exemple pour des scènes acquises dans un environnement structuré (construit par l'homme) les lignes droites verticales sont fréquentes et constituent un type d'indice courant. Par contre dans les environnements sous-marins [Sis00], cet indice est peu courant et donc inadapté.

Le principe de fonctionnement des méthodes de mises en correspondance d'indices repose sur deux étapes cruciales : 1) l'extraction des indices caractéristiques et 2) leur mise en correspondance. Nous allons détailler ces deux étapes dans la suite du paragraphe.

1.4.1 Extraction d'indices caractéristiques

On classe généralement les indices caractéristiques en deux groupes.

Le premier groupe est associé aux indices de type "contour". Ils correspondent à de fortes variations photométriques sur l'image. Ils changent très peu au cours d'une séquence, car ils résultent de la projection soit de "motifs décoratifs" fixés sur la scène (tableaux, frises, etc.), soit des contours des objets constituant la scène, soit de leurs textures.

Plusieurs approches sont envisageables pour la détermination des contours. Les plus répandues sont sans doute les méthodes dérivatives [Smi97] [HM95]. Elles sont divisées en trois catégories :

- Méthodes utilisant le calcul du gradient : elles consistent à déterminer les gradients directionnels de l'image en chaque point de l'image. La norme du gradient

est ensuite calculée. Le maximum local de la norme du gradient définit alors les points du contour.

- Méthodes basées sur le calcul des dérivées secondes : elles reprennent le calcul de la norme du gradient. La dérivée seconde est ensuite calculée dans la direction du gradient. Ensuite il convient de rechercher les passages par zéro de la dérivée seconde pour déterminer les points du contour.
- Méthodes reposant sur le calcul du Laplacien : elles sont basées sur le calcul du Laplacien de l'image lissée et sur la recherche de ses passages par zéro.

Le gradient et le laplacien sont obtenus en convoluant des masques de lissage et dérivation avec les images à traiter. Les masques les plus utilisés sont les masques de Sobel et Prewitt. Canny, et par la suite Deriche, Shen et Castan ont défini des filtres d'estimation du gradient optimisant trois critères de détection (fiabilité, précision et caractère optimal du filtre) [HM95]. Il existe d'autres approches faisant appel à des critères morphologiques ou surfaciques pour déterminer les contours d'une image. Celles-ci ne seront pas abordées dans ce manuscrit, cependant le lecteur peut se référer aux livres suivants [Ser82], [CC89].

Une fois les contours déterminés, l'image en niveaux de gris se réduit à une image binaire (noir et blanc). Parmi tous ces contours des indices tels que des droites ou des segments sont alors recherchés et mis en correspondance entre deux images consécutives de la séquence vidéo [Fau93] [Zha95]. Puisque l'image ne contient plus que les contours l'identification de droites est grandement simplifiée, réduisant ainsi la complexité des algorithmes par rapport à ceux traitant l'image brute.

La seconde catégorie d'indices concerne les points d'intérêt. Ces points ont été utilisés pour palier les défauts des détecteurs de points de contours à savoir : l'imprécision de la localisation (les extrémités des contours ne sont pas fiables). Contrairement aux points de contours, ces points correspondent à des caractéristiques bidimensionnelles telles que la plus forte courbure, les coins, les jonctions entre segments, etc. Généralement, il s'agit d'extraire les intersections de segments en "L", "T", "Y", ou "X". Ces points sont beaucoup moins nombreux que les points de contours mais ils sont nettement plus fiables.

Différents types d'extraction de ces points sont envisageables. Schmid [Sch96] distingue trois approches :

- La première approche s'appuie sur une extraction de contours et consiste à chercher parmi les points trouvés ceux dont la courbure est la plus forte. Les principales méthodes utilisant ce principe sont décrites dans [AB86] et [DF90].
- La deuxième approche exploite des modèles théoriques des points à extraire. Il s'agit de comparer tous les points de l'image avec ces modèles et d'en extraire ceux qui sont le plus ressemblant. Les travaux présentés dans [Roh92] et [DF90] sont parmi les plus connus.
- La troisième approche concerne les méthodes basées sur le traitement du signal. Il s'agit d'appliquer en chaque point de l'image un opérateur permettant la détection des points d'intérêts. Parmi les plus connus on note ceux définis par Beaudet (DET) basé sur le Hessien de l'image [Bea78], ou celui de Harris et Stephens

employant les dérivées premières de l'image [HS88], ou bien encore celui de Smith et Brady (SUSAN) [SB97]. Pour plus de détails sur ces opérateurs, le lecteur peut se reporter aux documents suivants : [Smi97] [ZT97] [HM95] et [Sis00].

1.4.2 Mise en correspondance et estimation de mouvement

Une fois les indices caractéristiques déterminés, il s'agit de les mettre en correspondance entre les images afin d'estimer le mouvement apparent. On distingue deux approches pour la mise en correspondance des points caractéristiques : les approches globales et les approches locales [MB96] [Bro92]. Les approches locales sortent du cadre de notre étude car seules les approches globales permettent d'estimer un mouvement majoritaire. Elles visent à mettre en correspondance un ensemble d'indices de l'image 1 avec un ensemble d'indices de l'image 2 par l'intermédiaire d'un modèle de mouvement. Pour ceci, une unique transformation est recherchée pour l'ensemble des indices de sorte qu'ils soient mis en correspondance de la meilleure façon possible. Ceci est souvent réalisé par une régression à l'aide des moindres carrés ou une distance statistique telle que la distance de Hausdorff. D'autres théories telles que la relaxation itérative de marquages ou le raisonnement à base de graphes sont aussi employés [BT80] [MN84]. Le nombre de points mis en correspondance doit alors être suffisamment élevé pour permettre l'identification de tous les paramètres du modèle de mouvement. Plusieurs applications de ces méthodes sont présentées dans [Bro92] dans des domaines tels que l'imagerie satellitaire ou médicale.

1.4.3 Avantages

Les points caractéristiques sont généralement choisis en fonction de leur invariance aux changements d'illumination de la scène filmée conférant ainsi une certaine robustesse au processus d'estimation de mouvement vis-à-vis des perturbations radiométriques. Par exemple, un coin dans une image restera un coin même si la luminosité diminue.

Ces points sont aussi choisis pour leurs attributs invariants aux transformations géométriques (translations, rotations, changement d'échelle, etc.) et aux déformations dues aux changements de perspective. Le modèle de mouvement identifiable dépend donc de la stabilité de ces points vis-à-vis des transformations.

L'extraction de points caractéristiques permet aussi de réduire considérablement le volume de données à traiter dans le processus d'estimation de mouvement. Au lieu de traiter tous les pixels, il suffira de traiter un ensemble réduit de points caractéristiques (avec ou sans leur voisinage). Le temps de calcul de ces algorithmes est donc réduit par rapport aux méthodes traitant l'image brute.

1.4.4 Inconvénients

Le fait que seuls quelques points ou les contours soient utilisés pour l'estimation du mouvement est aussi un inconvénient. L'estimation du champ de vitesse est éparse et n'est calculée qu'en des points particuliers, là où d'autres méthodes fournissent une estimation plus dense. Dans l'optique de l'estimation du mouvement dominant celle-ci se fait donc sur un nombre de mesures réduit. Ce genre de méthodes pose donc un problème. En effet, il n'existe aucune garantie que les points caractéristiques sélectionnés appartiennent à la partie de l'image associée au mouvement majoritaire. Prenons l'exemple d'un chien se déplaçant devant un mur. Le chien n'occupe qu'une petite partie de l'image par rapport au mur, cependant il y a de fortes chances pour qu'une majorité des points particuliers soient situés sur le chien. Le mouvement estimé correspondra donc à celui du chien et non du mur, alors que ce dernier est majoritaire sur l'image. Ces méthodes risquent donc d'être mises en défaut pour l'estimation du mouvement dominant.

Il est donc nécessaire de prendre un grand nombre de points pour être certain que la majorité seront associés à ce mouvement majoritaire. Ceci revient à sur-contraindre le système et risque de dégrader la précision du processus d'estimation de mouvement. En effet plus le nombre de points caractéristiques est élevé, plus on augmente la probabilité de faire de faux appariements.

De même, pour être robuste vis-à-vis des occlusions (points qui disparaissent entre deux images), il est également nécessaire de sélectionner un grand nombre de points. Ainsi on diminue l'influence des points particuliers qui disparaissent. Cependant, là encore cela se fait au détriment de la fiabilité des appariements.

1.5 Mise en correspondance de formes

Ces méthodes visent à mettre en correspondance des formes complexes pour en estimer le mouvement. La limite entre mise en correspondance de formes et mise en correspondance de points caractéristiques est très floue. On pourra parler de mise en correspondance de forme à partir du moment où on représente les données par un modèle. Par exemple dans [KN93] les auteurs extraient de l'image des contours fermés. Ceux-ci sont représentés par leur centre de gravité et leur longueur. Ils estiment ensuite les vitesses de déplacement du contour à l'aide de trois grandeurs : deux vitesses dans le plan image représentant l'évolution de la position du centre de gravité et une troisième vitesse relative à l'éloignement du contour (perpendiculaire au plan image). Cette dernière est calculée en fonction de la variation de la longueur du contour.

Dans le livre de Faugeras [Fau93], une méthode est présentée pour estimer le mouvement de la caméra à partir de six mises en correspondance de lignes dans trois images successives. Il n'est cependant pas précisé comment mettre en correspondance ces lignes. Cette estimation utilise une approche géométrique pour la modélisation et la mise en équation suivie d'une étape de minimisation pour déterminer les valeurs opti-

males des transformations. La transformation estimée correspond en fait au mouvement dans l'espace de la caméra et non au mouvement dans l'image.

Généralement, les méthodes de mise en correspondance de formes se contentent de déterminer si oui ou non l'objet dont on possède le modèle est présent dans la scène perçue. La position et l'orientation de l'objet peuvent être déterminées [CS96] [PFT94]. En fonction des variations de position et d'orientation on peut ensuite déduire le mouvement des objets. Cependant ces méthodes sont rarement utilisées pour l'estimation de mouvement.

1.6 Conclusion

Dans ce chapitre, nous avons proposé un état de l'art des principales techniques d'estimation de mouvement apparent majoritaire. Ces méthodes se divisent en deux classes : les approches de bas niveau et de haut niveau. Dans les approches bas niveau le traitement s'effectue directement sur le niveau de gris des pixels. Les approches haut niveau requièrent quant à elles un pré-traitement des images comme une extraction de contours ou de points caractéristiques.

Cette étude nous a permis de dégager les avantages mais aussi les inconvénients des principales méthodes. Il en ressort plusieurs constatations :

- Les méthodes apparentées à la corrélation nécessitent une discrétisation de l'espace des paramètres du mouvement. Celles-ci sont bien adaptées pour l'estimation de mouvements composés principalement de translations. Les mouvements introduisant des rotations n'engendrant pas des déplacements entiers dans l'espace image, et la discrétisation de l'espace angulaire n'est pas implicite. De plus la discrétisation de l'espace paramétrique influe sur la précision de l'estimation. La précision d'estimation de chaque paramètre est imposée par le pas d'échantillonnage de l'espace paramétrique. Un autre problème lié aux méthodes dites de corrélation est que l'on peut toujours trouver un maximum de correspondance entre les deux motifs, même si ceux-ci sont totalement différents. Pour éviter les faux appariements, il est nécessaire de fixer arbitrairement un seuil en dessous duquel l'estimation sera rejetée. La nature arbitraire de ce seuil dégrade la robustesse de la méthode et il semblerait préférable qu'il soit imposé par les données.
- Les méthodes utilisant des approches différentielles sont parmi les plus utilisées. Elle permettent une estimation dense du champ de vitesses apparentes. Cependant la formulation de l'équation de contrainte du flot optique utilisant les gradients de l'image ne suffit pas à déterminer totalement le mouvement apparent. Il est nécessaire de rajouter des contraintes de régularisation. De cette approche résultent plusieurs problèmes. D'une part, l'emploi de filtres pour calculer les gradients de l'image impose que le déplacement entre deux images soit petit et constitue une source d'erreurs non négligeable. D'autre part il résulte de l'emploi des filtres dérivateurs une hypothèse contradictoire quant à la texture de l'image.

Celle-ci doit être faible pour rester dans le domaine de validité des filtres et en même temps suffisante pour que l'équation de contrainte soit exploitable.

- Les méthodes faisant appel à l'extraction d'indices caractéristiques présentent comme avantage de réduire considérablement le volume de données à traiter, et donc d'en accélérer le traitement. Les points extraits possèdent généralement des caractéristiques invariantes aux transformations ou aux changements d'illumination, ce qui rend ces méthodes d'estimation plus robustes vis-à-vis de ces perturbations. Cependant, le fait que seuls quelques points soient utilisés rend le processus d'estimation sensible aux occlusions (disparition de points). Il est donc nécessaire de sélectionner un nombre important de points. Ceci se fait aux dépens de la fiabilité de la mise en correspondance des points. En effet plus le nombre de points à mettre en correspondance est important plus le risque d'appariements multiples est élevé. Le problème majeur de ces méthodes est donc d'obtenir une mise en correspondance fiable des différents indices.
- Pour les méthodes de mise en correspondance de formes complexes, le problème repose aussi sur l'appariement. Nous n'avons trouvé que peu d'utilisations de ces méthodes pour estimer un mouvement.

De cet état de l'art nous pouvons déduire que chaque méthode possède ses propres avantages et limitations. De nombreuses techniques ont été développées ultérieurement pour palier les lacunes de chaque méthode. Généralement leurs limitations proviennent d'hypothèses trop contraignantes vis-à-vis du niveau de gris des pixels.

Des approches basées sur la fusion de données ont été employées pour tirer le meilleur parti de plusieurs méthodes. Dans [PRH00], l'auteur propose de fusionner l'estimation de mouvement provenant d'une méthode différentielle avec celle provenant d'une méthode basée sur la corrélation.

Plutôt que de tenter d'améliorer les performances des méthodes déjà existantes, nous avons préféré employer une toute nouvelle approche. L'idée est de minimiser l'influence des hypothèses émises sur les variations des niveaux de gris des pixels d'une séquence d'images dans le processus d'estimation afin de le rendre plus robuste. Nous n'emploierons pas d'approche différentielle qui restreindrait l'espace des mouvements à estimer. Nous utiliserons une approche faisant appel aux théories de l'incertain.

Chapitre 2

Histogrammes approchés

Ce chapitre introduit la notion d'histogramme approché (H.A.) qui généralise le principe d'histogramme classique. Les histogrammes sont généralement employés en statistique pour visualiser l'allure d'une distribution de données réelles. Ils sont par contre rarement utilisés pour effectuer des calculs statistiques. Nous montrons au cours de ce chapitre que les H.A. permettent d'estimer un certain nombre de paramètres caractéristiques d'une distribution réelle tels que les modes, les fractiles, la moyenne, etc. Hormis cette possibilité de calcul, une des principales innovations des H.A. est qu'ils offrent la possibilité de prendre en compte l'imprécision des données de la distribution. Cet outil a été développé dans le cadre de la thèse en tant qu'outil mathématique nous permettant de réaliser l'estimation du mouvement apparent majoritaire dans une séquence d'images.

Le premier paragraphe de ce chapitre propose un bref aperçu des histogrammes classiques ainsi que les limites inhérentes à cet outil. Le deuxième paragraphe présente la notion d'histogramme approché pour des données mono-dimensionnelles. Il y sera explicité les différents éléments composant un histogramme approché ainsi que leur mode d'utilisation. Nous verrons qu'un des principaux avantages de ces histogrammes est de permettre une restitution de l'aspect continu de la distribution de données, et cela malgré la discrétisation de l'espace de représentation. Ce point est discuté dans le troisième paragraphe. Le paragraphe suivant présente par l'intermédiaire de deux exemples, la capacité des H.A. à servir d'outil statistique. Les deux exemples sont l'estimation d'un fractile d'ordre α et celle du mode d'une distribution. Le dernier paragraphe propose une extension des H.A. à des données bi-dimensionnelles puis N -dimensionnelles. Nous présentons les limites de la généralisation à des dimensions élevées, ainsi que les solutions envisagées pour pallier ce problème. Finalement, une conclusion fera la synthèse de l'apport des histogrammes approchés ainsi que de leurs limites, et tentera d'ouvrir des perspectives pour en améliorer le principe.

2.1 Histogrammes classiques.

2.1.1 Généralités

Les histogrammes classiques sont généralement utilisés en statistique pour avoir une représentation grossière des données. Ils permettent de visualiser simplement les caractéristiques de celles-ci, telles que la position moyenne des données, l'allure de leur répartition ainsi que leur étalement. Néanmoins, cet outil est très peu utilisé pour effectuer des calculs statistiques, comme l'estimation du mode d'une distribution.

Utiliser un histogramme usuel pour la classification consiste à diviser l'espace de représentation des données à traiter en un certain nombre de classes et à compter le nombre de données appartenant à chaque classe.

Dans le cas de données discrètes, les classes sont implicitement imposées par les valeurs possibles des données. Dans le cas de données réelles - qui nous intéresse ici - ces classes sont généralement des intervalles de \mathbb{R}^n avec n la dimension des données. En vue de simplifier notre propos, nous limiterons le cadre de l'exposé au cas $n = 1$, une généralisation étant proposée dans le paragraphe 2.5.

L'espace des réels est généralement subdivisé en intervalles réguliers de même précision, comme représenté sur la figure 2.1. Cette division de l'espace est appelée **partition classique de l'espace**.

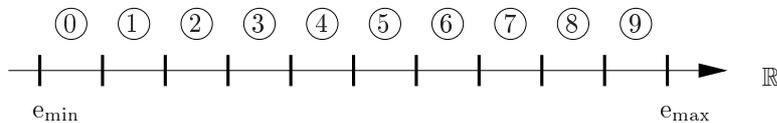


Figure 2.1 – Partition classique du segment $[e_{\min}, e_{\max}]$ de l'axe des réels en dix cellules.

La subdivision de \mathbb{R} a une influence majeure sur l'aptitude de l'histogramme à représenter les données [WW91]. Un partitionnement trop grossier aura tendance à masquer les détails de l'allure de la distribution de donnée en regroupant dans une même classe des données très différentes. Un partitionnement trop fin aura tendance à ne plus représenter l'allure générale de la distribution de données. En effet, si les classes sont trop fines, le nombre de données appartenant à celles-ci restera très faible, toutes les classes se ressembleront donc.

”Le nombre de classes retenues est un compromis raisonnable entre trop de détails et pas assez de renseignements [WW91].”

Construire un histogramme à partir de données réelles consiste à accumuler le nombre de données appartenant à chacun des intervalles de la partition. Si χ_k est la fonction caractéristique de l'intervalle C_k et $(d_i)_{i \in [1, N]}$ sont N données réelles, alors un accumulateur Acc_k est associé à chaque intervalle C_k de la partition. Sa valeur est définie par :

$$Acc_k = \sum_{i=1}^N \chi_k(d_i) \quad (2.1)$$

avec

$$\chi_k(d_i) = \begin{cases} 1 & \text{si } e_{\min} + (k - 1/2)\Delta < d_i < e_{\min} + (k + 1/2)\Delta \\ 0 & \text{sinon} \end{cases} \quad (2.2)$$

où d_i est une donnée issue de $(d_i)_{i \in [1, N]}$ et :

$$\Delta = \frac{e_{\max} - e_{\min}}{p} \quad (2.3)$$

avec p le nombre d'intervalles de la partition.

Ce processus d'accumulation est illustré sur la figure 2.2. L'accumulateur de l'intervalle C_k est alors augmenté de 1 à chaque fois qu'une donnée d appartient à C_k . C'est-à-dire chaque fois que $\chi_k(d) = 1$. On dira dans ce cas que la donnée d **vote** pour C_k .

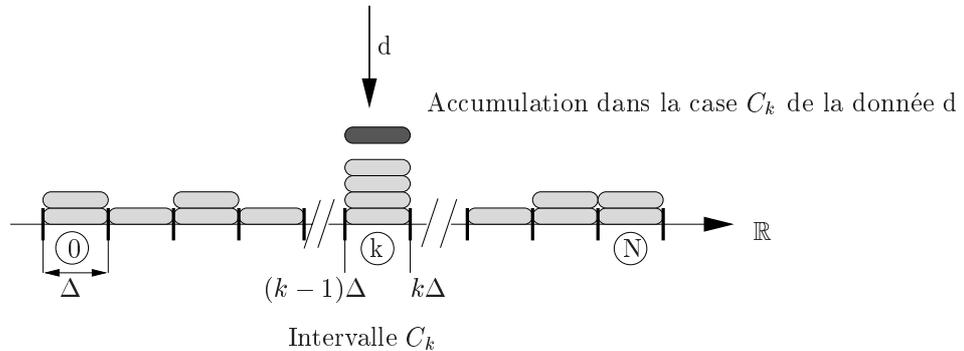


Figure 2.2 – Appartenance de la donnée d à la cellule H_k

Si la répartition des données réelles $(d_i)_{i \in [1, N]}$ suit une densité de probabilité, alors l'allure générale de l'histogramme se rapproche de cette distribution de probabilité sous-jacente, à mesure que N croît et que la taille des classes diminue [WW91].

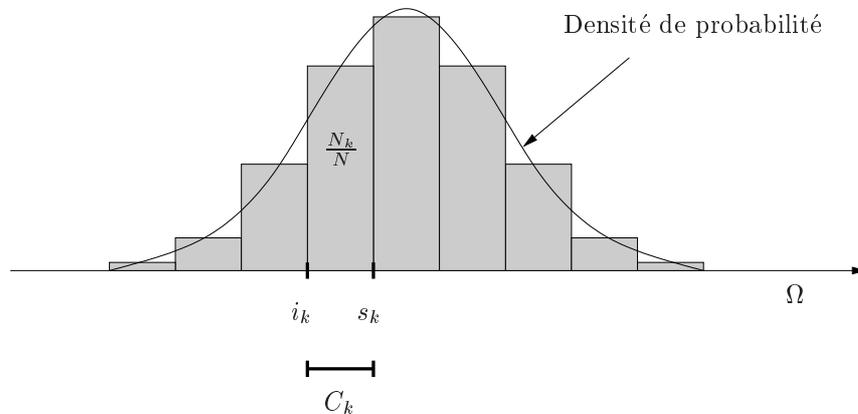


Figure 2.3 – Histogramme normalisé s'approchant de la densité de probabilité de la distribution à partir de laquelle il est construit.

Intervalle de taille	Classe de taille
]1m10, 1m30]	Très petit
]1m30, 1m50]	Petit
]1m50, 1m70]	Moyen
]1m70, 1m90]	Grand
]1m90, 2m10]	Très grand

Tableau 2.1 – Classes de taille des individus.

Sur la figure 2.3 N_k représente le nombre de données d associées à l'intervalle C_k et N le nombre total de données. On a une propriété remarquable des histogrammes : le rapport $\frac{N_k}{N}$ tend vers l'intégrale de la fonction de densité de probabilité sur l'intervalle C_k . Ceci s'écrit :

$$\lim_{N \rightarrow \infty} \frac{N_k}{N} = \int_{i_k}^{s_k} p(x) dx = P(C_k \cap (d_i)) \quad (2.4)$$

d'où

$$N \times P(C_k \cap (d_i)) \simeq N_k = Nb_H(C_k; (d_i)) \quad (2.5)$$

où p est la fonction de densité de probabilité de la distribution sous-jacente à la variable $(d_i)_{i \in [1, N]}$, $P(C_k \cap (d_i))$ est la probabilité d'avoir un élément dans l'intervalle C_k , $Nb_H(C_k; (d_i))$ est le nombre de données dans l'intervalle C_k connaissant les $(d_i)_{i \in [1, N]}$ représentées sur l'histogramme H . Aux vues de cette propriété, on peut penser que les histogrammes classiques sont appropriés aux calculs statistiques. Néanmoins, nous allons voir que certains problèmes liés au partitionnement prohibent l'emploi des histogrammes classiques pour de tels calculs.

2.1.2 Limites des histogrammes classiques

Les principaux problèmes associés aux histogrammes classiques sont au nombre de quatre : le problème de granularité, le caractère arbitraire du partitionnement, la dualité entre la précision de l'histogramme et sa fiabilité, et la non-prise en compte de l'imprécision des données.

Pour clarifier cet exposé, nous illustrerons chacun de ces problèmes à partir, entre autres, de l'exemple suivant : on souhaite classer un groupe d'individus relativement à leur taille. L'espace des tailles est partitionné en cinq classes qui sont définies dans le tableau 2.1.

Le problème associé à la **granularité** [Paw98], [PS98] est une conséquence de la représentation des données en classes d'équivalence. Il limite la précision de la représentativité des données pour l'histogramme en faisant perdre leur répartition dans chaque classe. Il n'y a en effet aucun moyen de différencier deux données d accumulées dans le même intervalle.

Pour illustrer ce problème, considérons deux personnes l'une de 1m501 et l'autre de 1m70. La première se situe tout juste à la limite basse de la classe des personnes

de taille moyenne alors que la deuxième se situe à la limite haute de cette même classe. Cependant, dans le processus classique de construction d'histogrammes par accumulation, ces deux personnes seront regroupées dans la même classe. On perd ainsi l'information sur leur position relative au sein de la classe.

Le deuxième problème est lié au caractère **arbitraire du partitionnement**. En effet, le partitionnement peut avoir une influence considérable sur la représentation des données. Une personne de 1m70 se situe dans la classe des personnes de taille moyenne. Si on considère maintenant une personne de 1m701, elle se situera dans la classe de personnes de grande taille. Ces deux personnes sont affectées à deux classes différentes alors que seul 1mm les sépare. Un déplacement de 1mm des bornes des classes aurait changé la classification de ces deux individus. Il convient bien de remarquer à quel point l'arbitrarité du partitionnement peut avoir une influence considérable sur la représentation des données.

L'exemple suivant illustre encore ce problème. Soient $(d_i)_{i \in [1,2000]}$ 2000 données réelles tirées aléatoirement d'une distribution gaussienne de moyenne 2.5 et d'écart type 1. Nous avons construit deux histogrammes pour représenter les données (d_i) . Les deux sont constitués de 11 intervalles. L'un est construit sur la portion $[-10, 10]$ de l'axe des réels (Figure 2.4), l'autre sur la portion $[-9, 11]$ (Figure 2.5). Il apparaît clairement que la représentation graphique des données $(d_i)_{i \in [1,2000]}$ dépend fortement du partitionnement.

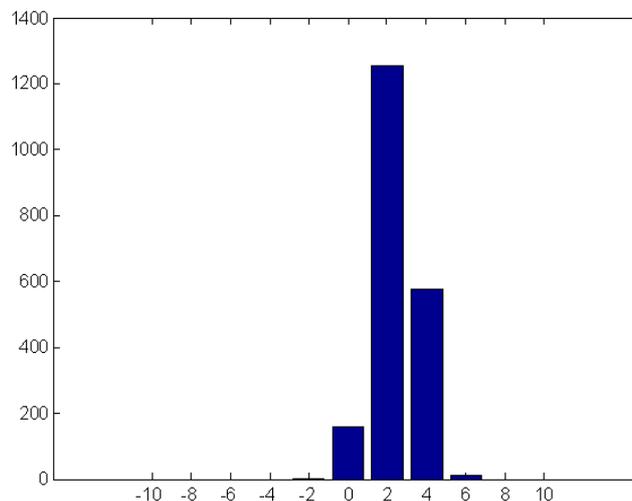


Figure 2.4 – Histogramme des données (d_i) construit à l'aide d'une partition de 11 cellules sur l'espace $[-10, 10]$.

En effet, sur la figure 2.4 les données les plus fréquentes sont celles comprises dans l'intervalle $[1, 3]$. Sur la figure 2.5, les données les plus fréquentes sont comprises dans l'intervalle $[2, 4]$. De plus la symétrie de l'histogramme n'est pas conservée lors du changement de partition.

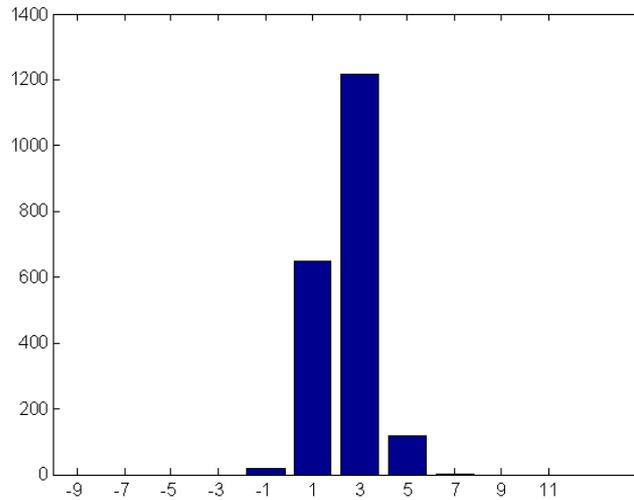


Figure 2.5 – Histogramme des données (d_i) construit à l'aide d'une partition de 11 cellules sur l'espace $[-9,11]$.

Une autre problématique des histogrammes classiques est la **dualité** entre **précision** et **fiabilité**. Comme présenté dans [WW91], le centre d'une classe est généralement utilisé pour représenter l'ensemble des données dans cette classe. Plus la cellule représentant la classe est large, plus la classe regroupera de données "éloignées", et moins la précision de l'histogramme sera élevée. La précision de l'histogramme est donc directement reliée à la largeur des intervalles ou cellules le constituant. On peut alors penser à diminuer la taille des cellules pour améliorer la précision de l'histogramme. Cependant, la fiabilité de l'information contenue dans une cellule (i.e. $P(Ck \cap (d_i))$) est directement reliée au nombre de votes accumulés dans cette cellule. Ainsi, le nombre de données N étant fixe dans une expérience, plus le nombre de cellules de la partition augmente, plus le nombre de données accumulées dans chaque cellule diminue. Ceci entraîne une diminution de la fiabilité de l'information contenue dans chaque cellule. Ce problème est illustré sur la figure 2.6 où $(d_i)_{i \in [1,200]}$ est une variable aléatoire gaussienne centrée sur 0 d'écart type un. On souhaite par exemple déterminer le mode de ces données (i.e. la valeur la plus fréquemment prise par ces données). Nous avons construit pour ce faire deux histogrammes sur l'intervalle $[-3, +3]$. Pour le premier et le second histogramme, la partition de l'espace $[-3, +3]$ contient respectivement cinq et deux cents intervalles.

Sur le premier histogramme, la localisation du mode est aisée et fiable. Près de cent valeurs sur les deux cents utilisées ont voté pour l'intervalle $[-0.5, +0.5]$. On peut donc affirmer avec une bonne fiabilité que le mode de la distribution se situe dans cet intervalle. Cette fiabilité peut être corroborée par d'autres expérimentations qui donneront un histogramme similaire. Cependant la position réelle du mode dans l'intervalle est inconnue. Une précision de ± 0.5 peut en effet être perçue comme insuffisante. Si on souhaite une meilleure précision dans la localisation du mode, il est naturel de prendre un partitionnement plus fin de la droite réelle. Cependant, dans ce cas, la fiabilité de la

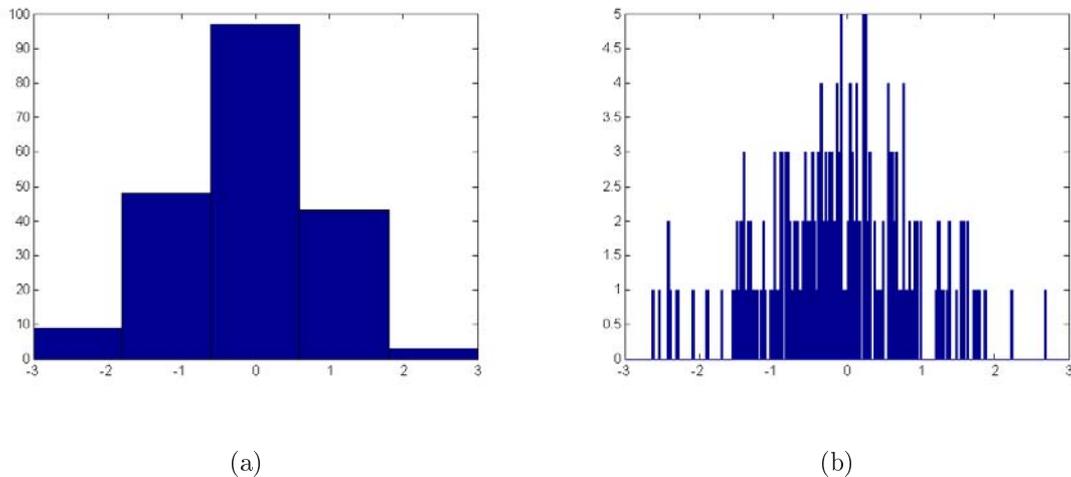


Figure 2.6 – Histogrammes représentant la variable aléatoire X (a) Histogramme à 5 cellules (b) Histogramme à 200 cellules.

localisation du mode est mauvaise. Sur la figure 2.6 (b) les intervalles $[-0.09, -0.06]$, $[0.18, 0.21]$, et $[0.21, 0.24]$ sont candidats au statut de mode avec une faible probabilité (5 votes face à 200 votes au total soit 2.5% des votes). Une itération de l'expérience aurait certainement donné d'autres intervalles.

Il apparaît clairement que la construction d'un histogramme pour représenter des données $(d_i)_{i \in [1, 200]}$ n'est pas simple. On se trouve confronté au problème suivant : pour choisir de manière adéquate le partitionnement de l'histogramme, il faudrait connaître un certain nombre de paramètres statistiques de la distribution des données à traiter. Cela peut limiter l'emploi des histogramme pour des autres applications qu'une visualisation des données.

Enfin, un des problèmes qui nous intéresse le plus dans le cadre de notre application est la **difficulté à prendre en compte l'imprécision des données** [Vie96]. Une donnée imprécise est une donnée dont on ne connaît pas la valeur exacte mais seulement un ensemble de valeurs pouvant être prises par cette donnée (généralement il s'agit d'un intervalle). La distribution exacte (d_i) est alors remplacée par la distribution imprécise (D_i) des intervalles de définition de chaque donnée.

La procédure d'accumulation est alors modifiée. En effet, il se peut que l'ensemble D_i intersecte plusieurs intervalles de la partition. Dans ce cas comment la donnée vote elle pour un intervalle C_k de l'histogramme ?

Plusieurs approches sont envisageables. La plus intuitive consiste à faire voter une donnée D_i proportionnellement au rapport de la taille de l'intersection entre C_k et D_i sur le cardinal de D_i . Cela s'écrit :

$$Vote_{C_k}(D_i) = \frac{|C_k \cap D_i|}{|D_i|} \quad (2.6)$$

où $|X|$ est la taille de l'intervalle X . Le nombre total de données N sera conservé puisque $\sum_{k=1}^p \text{Vote}_{C_k}(D_i) = 1$, avec p le nombre de classes de l'histogramme. Cependant, l'hypothèse sous-jacente à cette méthode prenant en compte l'imprécision des données est que la densité de probabilité de la donnée est uniforme dans son intervalle de définition. Soit cette hypothèse est dictée par une connaissance sur les données, soit elle rajoute une contrainte a priori sur les données. Un des effets pervers d'une telle méthode est de "défavoriser" a priori les données les moins précises. Or ce sont justement les données les moins précises dont on est le plus sûr [WW91].

Si on supprime la normalisation par le cardinal de l'intervalle représentant D_i , on supprime toute hypothèse a priori sur la distribution de probabilité de D_i dans son intervalle de définition. Cependant la somme de tous les votes se rapportant à D_i ($\sum_{k=1}^p \text{Vote}_{C_k}(D_i)$) est supérieure à un. On aura donc une valeur par excès de l'accumulateur associé à la partition de l'histogramme. Dans ce cas, les valeurs les moins précises risquent d'occulter les valeurs plus précises.

Une autre approche encore plus extrême consiste à considérer qu'un ensemble D_i vote pour un intervalle C_k de la partition si son intersection avec celui-ci n'est pas vide, ce qui se traduit par :

$$\begin{aligned} \text{Vote}_{C_k}(D_i)_{Excès} &= 1 \text{ ssi } C_k \cap D_i \neq \emptyset \\ &= 0 \text{ sinon} \end{aligned} \quad (2.7)$$

On a donc là encore une approximation par excès du nombre de données dans les intervalles de la partition de l'espace. La méthode duale est définie par : une donnée D_i vote pour un intervalle C_k de la partition si l'intersection entre C_k et D_i vaut D_i . C'est-à-dire si l'ensemble D_i est entièrement inclus dans l'intervalle C_k de la partition. Cela se traduit par :

$$\begin{aligned} \text{Vote}_{C_k}(D_i)_{Défaut} &= 1 \text{ ssi } C_k \cap D_i = D_i \\ &= 0 \text{ sinon} \end{aligned} \quad (2.8)$$

Dans ce cas, on a une approximation par défaut du nombre de données dans les intervalles de la partition de l'histogramme. La méthode utilisant une approximation par excès et par défaut du nombre de données dans les cellules de la partition est la façon la plus neutre de prendre en compte l'imprécision des données. Elle ne suppose aucune hypothèse a priori sur la répartition des données. Cependant elle ne délivre qu'un encadrement grossier du nombre de données dans chaque intervalle. Une donnée ne votant que d'une quantité 0 ou 1 pour un intervalle C_k , il n'y a aucune nuance en fonction de l'intersection entre D_i et C_k . Par ailleurs, si l'imprécision des données n'est pas en adéquation avec la partition, il se peut que l'accumulateur par défaut soit nul en tout point. Dans ce cas, aucune information n'est disponible du vote par défaut et la partition doit être transformée pour mieux s'adapter aux données.

Toutes ces limitations et surtout la mauvaise prise en compte de l'imprécision des données nous ont conduit à nous intéresser à l'apport de la théorie du flou pour les histogrammes. Le paragraphe suivant propose une extension de la notion d'histogramme

permettant de réduire l'influence de la discrétisation sur la construction d'un histogramme. De plus cette extension permet de prendre en compte l'imprécision des données sans pour autant présupposer une distribution a priori. Il sera aussi illustré au travers d'exemples comment cette extension des histogrammes permet de passer d'un simple outil de visualisation à un outil de calcul.

2.2 Principe des histogrammes approchés

Le principe des histogrammes approchés est basé sur la théorie des sous ensembles flous grossiers [DP90], [Paw98]. Leur but est de dissocier la granularité de l'histogramme de la précision des informations que l'on en extrait - comme par exemple la localisation du mode de la distribution des données accumulées dans l'histogramme. La première partie de ce paragraphe présente comment, en remplaçant la partition classique de l'histogramme par une partition floue, on s'affranchit partiellement des problèmes liés à l'arbitrarité du partitionnement. La deuxième partie présente le processus d'accumulation lorsque les données sont précises. Enfin, le dernier paragraphe reprend le processus d'accumulation lorsque les données sont imprécises.

2.2.1 Partitionnement flou de l'espace

Pour atténuer les effets dus à l'aspect arbitraire du partitionnement de l'espace propre aux histogrammes, les histogrammes approchés sont construits sur une partition floue. Certaines propriétés sont nécessaires pour qu'une famille de sous ensembles soit une partition floue [Bez81], [Wan97], [ZALE99].

Soit Φ une portion de l'espace Ω . Soit U un ensemble de N sous ensembles flous $(H_k)_{k \in [1, P]}$ sur l'intervalle Φ . Soit μ_k la fonction d'appartenance du sous ensemble flou H_k . U est une partition floue si et seulement si :

$$\forall \omega \in \Phi, \forall k \in [1, P], \mu_k(\omega) \in [0, 1] \quad (2.9)$$

$$\forall \omega \in \Phi, \exists k / \mu_k(\omega) \neq 0 \quad (2.10)$$

$$\forall (i, j) \in [1, P] \text{ si } \Pi(H_i; H_j) = 1 \text{ alors } i = j \quad (2.11)$$

$$\forall \omega \in \Phi, \sum_{k=1}^P \mu_k(\omega) = 1 \quad (2.12)$$

avec :

$$\Pi(H_i; H_j) = \sup_{\Omega} (\min(\mu_i, \mu_j))$$

L'équation (2.10) traduit le fait qu'il n'y a pas de discontinuité dans le partitionnement. L'équation (2.11) exprime la discernabilité entre deux cellules de la partition. C'est à dire que deux cellules ne peuvent représenter identiquement une même variable. Finalement, l'équation (2.12) est une équation de normalisation qui n'est pas indispensable mais utile si l'on désire effectuer des calculs statistiques.

Pour simplifier le processus d'accumulation, mais aussi pour rester le plus neutre vis-à-vis de la partition, nous avons choisi d'utiliser des fonctions d'appartenance linéaires. Pour conserver une homogénéité entre la partition floue et la partition binaire, l'aire de la fonction caractéristique binaire doit être la même que celle de la fonction caractéristique linéaire. La base de la cellule binaire étant de Δ , la base de la cellule triangulaire est de 2Δ . La partition binaire de la figure 2.7 est changée en partition floue comme sur la figure 2.8.

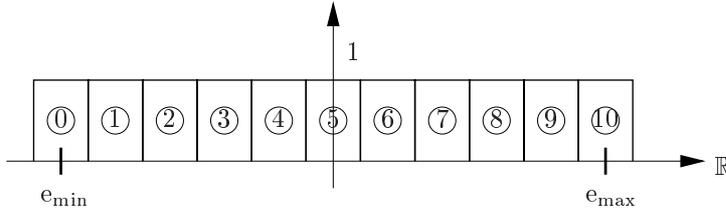


Figure 2.7 – Représentation des fonctions caractéristiques des 11 cellules formant une partition binaire de l'axe des réels entre e_{\min} et e_{\max} .

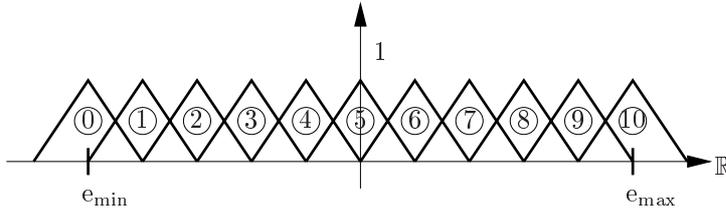


Figure 2.8 – Représentation des fonctions caractéristiques des 11 cellules formant une partition floue de l'axe des réels entre e_{\min} et e_{\max} .

L'équation de la fonction caractéristique de l'intervalle (H_k) de la partition s'écrit alors :

$$\mu_k(x) = \left\{ \begin{array}{l} \frac{1}{\Delta}(x - e_{\min} - (k-1)\Delta) \text{ si } x \in [e_{\min} + (k-1)\Delta, e_{\min} + k\Delta] \\ 1 - \frac{1}{\Delta}(x - e_{\min} - k\Delta) \text{ si } x \in [e_{\min} + k\Delta, e_{\min} + (k+1)\Delta] \\ 0 \text{ sinon} \end{array} \right\} \quad (2.13)$$

On notera m_k la position du centre du $k^{\text{ième}}$ intervalle de la partition. C'est à dire :

$$m_k = e_{\min} + k\Delta \quad (2.14)$$

On parlera de cellule floue pour désigner la représentation graphique du sous ensemble flou (H_k) et sa fonction caractéristique.

2.2.2 Accumulation de données précises

Dans le cas où les données $(D_i)_{i \in [1, N]}$ sont précises, leur domaine de définition se réduit à un seul singleton $D_i = \{d_i\}$ avec $d_i \in \mathbb{R}$. L'extension de la formule (2.1) est :

$$Acc_k = \sum_{i=1}^N \mu_k(d_i) \quad (2.15)$$

où μ_k est la fonction caractéristique du sous ensemble flou H_k de la partition. Ce type de vote équivaut à une répartition de l'influence de la donnée sur deux cellules adjacentes de la partition. Ceci est illustré sur la figure 2.9. où Acc_k (resp. Acc_{k+1}) sont

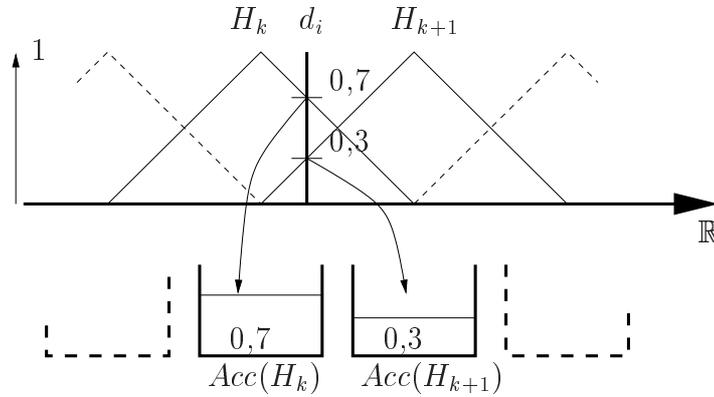


Figure 2.9 – Vote d'une donnée précise sur un histogramme approché.

les accumulateurs associés à la cellule H_k (resp. H_{k+1}). Comme pour les histogrammes classiques, on retrouve $\sum_{k=1}^P Acc_k = N$, avec N le nombre de données votant dans l'histogramme. Dans le cas de données précises, le fait d'utiliser un histogramme approché n'a pas ajouté d'imprécision sur le nombre de votants.

2.2.3 Accumulation de données imprécises

Si les données $(D_i)_{i \in [1, N]}$ sont imprécises, alors celles-ci sont représentées par des quantités floues. Dans ce cas l'extension de la formule (2.1) est nettement moins triviale. En effet il n'est plus possible de déterminer à quel point l'ensemble D_i appartient à une cellule floue de la partition comme lorsque les données sont précises. Il existe au contraire tout un ensemble continu de valeurs de compatibilités entre les deux ensembles D_i et H_k . L'imprécision de la donnée D_i rend imprécise la mesure de la compatibilité entre D_i et H_k . Nous avons choisi de représenter cet ensemble de compatibilités par un intervalle ordinaire, c'est-à-dire par sa borne supérieure et sa borne inférieure. Ces deux bornes sont données par la possibilité de H_k restreinte à D_i notée $\Pi(H_k; D_i)$ pour la borne supérieure et la nécessité de H_k restreinte à D_i notée $N(H_k; D_i)$ pour la borne inférieure. Leurs expressions, issues de la théorie des possibilités [DP88], sont données par les formules (2.16) et (2.17)

$$\Pi(H_k; D_i) = \sup_{\Omega} (\min(\mu_{H_k}(x), \mu_{D_i}(x))) \quad (2.16)$$

$$N(H_k; D_i) = \inf_{\Omega} (\max(\mu_{H_k}(x), 1 - \mu_{D_i}(x))) \quad (2.17)$$

où Ω est l'axe des réels, μ_{D_i} est la fonction d'appartenance à la donnée imprécise D_i , et μ_{H_k} est la fonction d'appartenance à la cellule H_k de la partition floue. La mesure de possibilité restreinte définit à quel point la donnée D_i intersecte la cellule H_k , tandis que la mesure de nécessité restreinte définit à quel point la donnée D_i est incluse dans la cellule H_k . La valeur de la possibilité restreinte correspond à une mesure optimiste de la compatibilité entre la donnée et la cellule de la partition. La valeur de la nécessité conjointe, quant à elle, correspond à une mesure pessimiste de la compatibilité entre la donnée et la cellule de la partition.

La généralisation de (2.15) au cas imprécis nous amène à remplacer l'accumulateur précis Acc_k par un accumulateur imprécis défini par sa borne inférieure \underline{Acc}_k et sa borne supérieure \overline{Acc}_k . La valeur de la nécessité restreinte sera accumulée dans l'accumulateur inférieur (ou accumulateur par défaut), et la valeur de la possibilité restreinte sera accumulée dans l'accumulateur supérieur (ou accumulateur par excès). Ceci s'écrit :

$$\underline{Acc}_k = \sum_{i=1}^N N(H_k; D_i) \quad (2.18)$$

$$\overline{Acc}_k = \sum_{i=1}^N \Pi(H_k; D_i) \quad (2.19)$$

Les figures 2.10 et 2.11 illustrent le calcul de la possibilité et de la nécessité conjointe, c'est-à-dire comment la donnée imprécise D_i va voter sur l'histogramme approché.

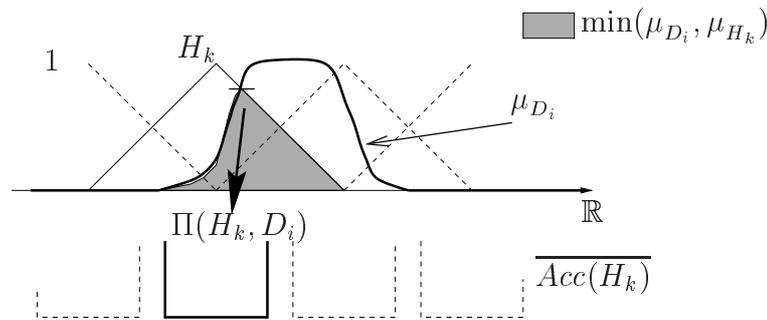


Figure 2.10 – Accumulation d'une donnée imprécise sur l'accumulateur supérieur.

REMARQUE 3 Dans le cas où les données sont précises, les mesures inférieures et supérieures de compatibilité entre la donnée et la cellule de la partition sont identiques. En effet, on a $\mu_D(x)$ qui est nul partout sauf en un point x_0 (là où se situe la donnée) où elle vaut 1. Dans ce cas, les formules (2.16) et (2.17) deviennent :

$$\begin{aligned} \Pi(H_k; D_i) &= \min(\mu_{H_k}(x_0), 1) = \mu_{H_k}(x_0) \\ N(H_k; D_i) &= \max(\mu_{H_k}(x_0), 0) = \mu_{H_k}(x_0) \end{aligned} \quad (2.20)$$

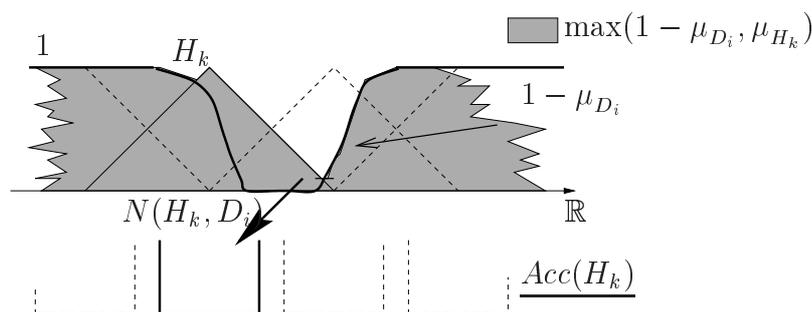


Figure 2.11 – Accumulation d’une donnée imprécise sur l’accumulateur inférieur.

Puisque les valeurs $\Pi(H_k; D_i)$ et $N(H_k; D_i)$ sont identiques, les deux accumulateurs sont confondus. On voit bien ici que l’écart entre les deux accumulateurs est relié à la précision des données. Si les données sont précises, cet écart est nul.

Un des intérêts des histogrammes approchés repose sur le fait que l’écart entre les deux accumulateurs est fonction de l’imprécision sur les données. Plus l’imprécision sur les données est grande plus l’écart entre le vote optimiste et le vote pessimiste sera grand, et plus l’écart entre les deux accumulateurs sera élevé.

2.3 Restitution du continu

On a vu dans le paragraphe 2.1 qu’un histogramme classique ne donnait qu’une approximation grossière de la densité de probabilité sous-jacente aux données accumulées sur celui-ci. La finesse de celle-ci est imposée par la précision des intervalles de la partition. La dualité précision/fiabilité des histogrammes classiques exclut une diminution abusive de la taille des classes sous peine de réduire leur fiabilité.

Contrairement aux histogrammes classiques, les histogrammes approchés permettent de restituer une approximation de la densité de probabilité de manière continue. Dans un premier temps nous traiterons le cas le plus général consistant à reconstituer une approximation de la densité de probabilité d’une variable aléatoire à partir de données imprécises. Nous illustrerons les propriétés de restitution du continu par des exemples. Il sera aussi mis en évidence que l’influence de la position des bornes du partitionnement est considérablement réduite par rapport au cas des histogrammes classiques.

2.3.1 Cas général.

La densité de probabilité d’une distribution est étroitement liée au nombre de données accumulées dans l’histogramme construit sur des données issues de cette distribution [WW91]. Soit $(D_i)_{i \in [1, N]}$ N données tirées aléatoirement de la distribution dont on doit estimer l’allure de la densité de probabilité. Il est mentionné dans le paragraphe 2.1 (cf. formule (2.4)) que l’intégrale de la densité de probabilité sur un intervalle est

proche du rapport $\frac{N_k}{N}$ avec N_k le nombre de votes accumulés dans l'intervalle C_k . La formule (2.21) donne une approximation de la probabilité d'un élément X_k , noté $P(X_k)$, si celui-ci est compris entre les bornes $[i_k, s_k]$. L'égalité est exacte lorsque la densité de probabilité est uniforme sur l'intervalle $[i_k, s_k]$. Dans le cas d'une densité de probabilité quelconque, l'allure de la densité de probabilité est conservée.

$$P(X_k) \simeq \frac{1}{s_k - i_k} \int_{i_k}^{s_k} p(x) dx \quad (2.21)$$

En utilisant les équations (2.5) et (2.21) on peut écrire :

$$P(X_k) \simeq \frac{1}{\Delta} \times \frac{Nb(C_k; (d_i))}{N} \quad (2.22)$$

où Δ est la largeur de l'intervalle C_k .

Dans le cas où les données sont imprécises, on n'a pas connaissance du nombre exact de données accumulées dans un intervalle. En effet, les accumulateurs inférieurs et supérieurs de l'histogramme approché ne fournissent qu'une valeur par défaut et par excès de ce nombre. La formule (2.22) n'est donc plus utilisable. La valeur de la probabilité d'un événement W en connaissant les données $(D_i)_{i \in [1, N]}$ (notée $P(W; (D_i))$) ne pourra donc être connue que par un intervalle de définition délimité par deux bornes. Il est donc nécessaire de trouver l'encadrement le plus fin possible pour caractériser $P(W; (D_i))$. La formule de Bayes généralisée [Wal91] fournit un encadrement de la valeur de cette probabilité avec plusieurs niveaux de précision. Elle s'écrit :

$$\begin{aligned} \frac{\underline{P}(W \cap (D_i))}{\underline{P}(W \cap (D_i)) + \overline{P}(W^c \cap (D_i))} &\leq \underline{P}(W; (D_i)) \leq P(W; (D_i)) \\ &\leq \overline{P}(W; (D_i)) \leq \frac{\overline{P}(W \cap (D_i))}{\overline{P}(W \cap (D_i)) + \underline{P}(W^c \cap (D_i))} \end{aligned} \quad (2.23)$$

où $\underline{P}(W; (D_i))$ (resp. $\overline{P}(W; (D_i))$) est la probabilité inférieure (resp. supérieure) d'avoir l'événement W en connaissant la répartition des données (D_i) , $\underline{P}(W \cap (D_i))$ (resp. $\overline{P}(W \cap (D_i))$) est la probabilité inférieure (resp. supérieure) d'avoir l'événement W et toutes les données (D_i) , et W^c est l'événement contraire de W .

La caractérisation de N données sur un histogramme approché ne permet pas de fournir les probabilités inférieures et supérieures exactes. Il est donc nécessaire de considérer le deuxième niveau de l'encadrement de $P(W; (D_i))$ à savoir :

$$\frac{\underline{P}(W \cap (D_i))}{\underline{P}(W \cap (D_i)) + \overline{P}(W^c \cap (D_i))} \leq P(W; (D_i)) \leq \frac{\overline{P}(W \cap (D_i))}{\overline{P}(W \cap (D_i)) + \underline{P}(W^c \cap (D_i))} \quad (2.24)$$

Le rapport des probabilités correspond à une normalisation. De plus, il apparaît clairement dans cette formulation, que la probabilité d'avoir un événement W décroît à mesure que la probabilité d'avoir son événement contraire augmente. Avec la représentation des N données imprécises $(D_i)_{i \in [1, N]}$ sur un histogramme approché, la valeur de la probabilité (inférieure ou supérieure) d'avoir l'événement W et toutes les données (D_i)

- notée $\underline{P}(W \cap (D_i))$ - n'est pas connue de manière rigoureuse. Néanmoins, en faisant une analogie avec les histogramme classiques, cf. formule (2.22), ces probabilités peuvent être assimilées, à un facteur près, aux nombres inférieur et supérieur de données correspondant à l'événement W . Ce qui s'écrit :

$$\begin{aligned}\underline{P}(W \cap (D_i)) &\simeq \frac{1}{K} \underline{Nb}_H(W \cap (D_i)) \\ \overline{P}(W \cap (D_i)) &\simeq \frac{1}{K} \overline{Nb}_H(W \cap (D_i))\end{aligned}\tag{2.25}$$

avec K facteur de normalisation inconnu et $\underline{Nb}_H(W \cap (D_i))$ (resp. $\overline{Nb}_H(W \cap (D_i))$) le nombre inférieur (resp. supérieur) de données appartenant à l'intersection entre W et les intervalles flous (H_k) de l'histogramme approché sur lequel les données sont accumulées. En utilisant l'équation (2.25), le facteur K inconnu disparaît et on a :

$$\begin{aligned}\frac{\underline{Nb}_H(W \cap (D_i))}{\underline{Nb}_H(W \cap (D_i)) + \overline{Nb}_H(W^c \cap (D_i))} &\leq \\ &P(W; (D_i)) \\ &\leq \frac{\overline{Nb}_H(W \cap (D_i))}{\overline{Nb}_H(W \cap (D_i)) + \underline{Nb}_H(W^c \cap (D_i))}\end{aligned}\tag{2.26}$$

Le fait de chercher de manière continue l'allure de la densité de probabilité, implique que l'on connaisse les nombres supérieurs et inférieurs de données ayant voté pour l'intervalle W quelle que soit la position de celui-ci. C'est-à-dire, connaître ces nombres même pour des positions de W intermédiaires situées entre les valeurs discrètes des intervalles (H_k) de la partition floue.

La technique employée, pour estimer continûment ces nombres de votes, est inspirée du modèle de transfert de croyance (transferable belief model ou TBM) mis au point par Philippe Smets dans [Sme90] et [Sme94] et reprise par Thierry Denoeux dans [Den97]. Le principe repose sur le transfert de la croyance affectée à des éléments dont la masse de croyance est connue vers un élément quelconque sur lequel on veut prendre une décision. Ce transfert fait intervenir la notion de probabilité pignistique $BetP$ (du latin pignus qui signifie pari).

$$BetP(\{B\}) = \sum_{A \subset \mathfrak{R}} m(A) \frac{|\{B\} \cap A|}{|A|}\tag{2.27}$$

où \mathfrak{R} est l'espace considéré, $m(A)$ est la masse de croyance qui correspond à la confiance attribuée exclusivement à l'ensemble $A \subset \mathfrak{R}$ [Sha76]. $|A|$ signifie cardinal de A , c'est-à-dire nombre d'éléments dans A .

Revenons à notre problème qui est d'estimer le nombre de votes qu'aurait recueilli un intervalle W inclus dans le support de la partition floue à partir des nombres de votes accumulés dans les intervalles de la partition floue. Plusieurs méthodes permettent

une approximation de ce nombre de votes (méthode linéaire, quadratique etc...) La méthode suggérée par Philippe Smets consiste à transférer le nombre de votes dans les cellules de la partition vers l'intervalle W au prorata du recouvrement des intervalles W et (H_k) . En utilisant une analogie avec la probabilité pignistique, une approximation raisonnable des nombres inférieurs et supérieurs de votes contenus dans W -connaissant la répartition des données $(D_i)_{i \in [1, N]}$ sur la partition floue constituée des (H_k) - est obtenue en utilisant l'équation :

$$\underline{Nb}_H(W \cap (D_i)) = \sum_{k=0}^p \underline{Acc}_k \frac{|W \cap H_k|}{|H_k|} \quad (2.28)$$

$$\overline{Nb}_H(W \cap (D_i)) = \sum_{k=0}^p \overline{Acc}_k \frac{|W \cap H_k|}{|H_k|} \quad (2.29)$$

$$\underline{Nb}_H(W^c \cap (D_i)) = \sum_{k=0}^p \underline{Acc}_k \frac{|W^c \cap H_k|}{|H_k|} \quad (2.30)$$

$$\overline{Nb}_H(W^c \cap (D_i)) = \sum_{k=0}^p \overline{Acc}_k \frac{|W^c \cap H_k|}{|H_k|} \quad (2.31)$$

avec p le nombre de cellules de la partition, \underline{Acc}_k (resp. \overline{Acc}_k) le nombre inférieur (resp. supérieur) de données accumulées dans l'intervalle flou (H_k) et $|X|$ l'aire de la fonction caractéristique de l'ensemble X . Le principe du transfert de nombre de votes est illustré sur la figure 2.12.

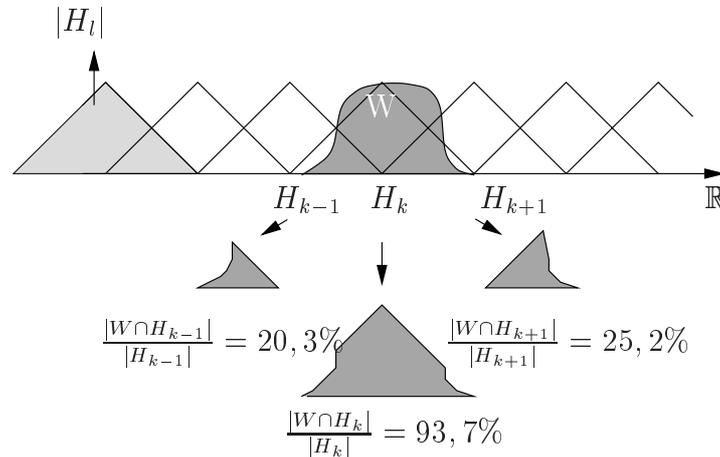


Figure 2.12 – Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'ensemble W .

2.3.2 Cas particuliers pour le sous ensemble W .

Afin d'effectuer le calcul de la reconstruction de la densité de probabilité, il reste encore une question en suspens : l'allure de l'intervalle W ? Quelle forme de W peut

on prendre sans trop rajouter d'a priori : un intervalle flou identique à ceux de la partition floue, un intervalle binaire de même aire que ceux de la partition, autre chose ? Nous avons testé trois types d'intervalles. Dans un premier temps l'intervalle W a été pris identique aux intervalles de la partition floue. Ceci correspond à l'approche la plus naturelle. Elle revient à considérer un élément de la partition se positionnant n'importe où, pour lequel on essaye d'estimer le nombre de votes accumulés. Dans un second temps, nous avons essayé un intervalle binaire. Deux choix sont possibles pour conserver une certaine cohérence entre le processus de vote et le processus d'estimation du vote :

- Soit on conserve la granularité de l'intervalle, c'est-à-dire la taille de son support ($\Delta' = \Delta$).
- Soit on conserve la précision de l'intervalle, c'est-à-dire l'aire de la fonction caractéristique de cet intervalle ($\int \mu(x) dx = \int \chi(x) dx$).

Les formules de (2.28) à (2.31) seront détaillées pour les trois choix de W et nous comparerons leur efficacité sur un exemple significatif issu de nombreuses simulations.

Cas de l'ensemble W de forme triangulaire

Explicitons dans le cas de l'intervalle flou triangulaire le calcul du nombre de données qu'aurait reçu W s'il avait été un élément de la partition. Soit m_W la position du mode de l'intervalle flou W sur la partition floue. La fonction caractéristique de W est la suivante :

$$\mu_W(x) = \begin{cases} \frac{1}{\Delta}(x - m_W + \Delta) & \text{si } x \in [(m_W - \Delta, m_W)] \\ 1 - \frac{1}{\Delta}(m_W - x) & \text{si } x \in [m_W, m_W + \Delta] \\ 0 & \text{sinon} \end{cases} \quad (2.32)$$

Le transfert du nombre de votes contenu dans la partition vers l'intervalle W est illustré sur la figure 2.13 :

avec :

$$\begin{aligned} Aire(H_k) &= \Delta \\ Aire_{H_{k-1} \cap W} &= \frac{1}{4}(1 - \alpha)^2 \\ Aire_{H_k \cap W} &= \frac{1}{4}(2 - \alpha)^2 \\ Aire_{H_{k+1} \cap W} &= \frac{1}{4}(1 + \alpha)^2 \\ Aire_{H_{k+2} \cap W} &= \frac{1}{4}\alpha^2 \\ Aire_{H_i \cap W} &= 0 \forall i \neq \{k - 1, k, k + 1, k + 2\} \end{aligned} \quad (2.33)$$

et où $\alpha\Delta$ est le décalage entre m_W et le mode du deuxième intervalle intersecté par W . Les équations (2.28) et (2.29) peuvent donc se réécrire :

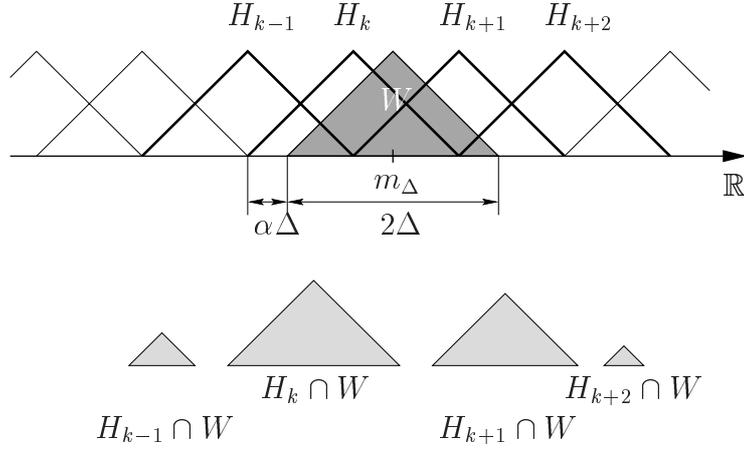


Figure 2.13 – Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle flou W dont la fonction caractéristique est triangulaire, et de granularité 2Δ .

$$\underline{Nb}_H(W \cap (D_i)) = \frac{1}{4} \left(\frac{\underline{Acc}_{k-1}(1-\alpha)^2 + \underline{Acc}_k(2-\alpha)^2}{+\underline{Acc}_{k+1}(1+\alpha)^2 + \underline{Acc}_{k+2}\alpha^2} \right) \quad (2.34)$$

$$\overline{Nb}_H(W \cap (D_i)) = \frac{1}{4} \left(\frac{\overline{Acc}_{k-1}(1-\alpha)^2 + \overline{Acc}_k(2-\alpha)^2}{+\overline{Acc}_{k+1}(1+\alpha)^2 + \overline{Acc}_{k+2}\alpha^2} \right) \quad (2.35)$$

Cette représentation souffre d'un défaut de normalisation. Si toutes les cellules H_k intersectant W ont accumulé le même nombre de données ($\underline{Acc}_{k-1} = \underline{Acc}_k = \underline{Acc}_{k+1} = \underline{Acc}_{k+2} = A$), il semble légitime de penser que $\underline{Nb}_H(W \cap (D_i))$ sera égal à A quelle que soit la position de W sur l'intervalle considéré (c'est à dire $\forall \alpha \in [0, 1]$). Or il n'en est rien, la valeur de $\underline{Nb}_H(W \cap (D_i))$ dépend de α ($\underline{Nb}_H(W \cap (D_i)) = \frac{1}{4}A(6 - 4\alpha + 4\alpha^2)$). Il est donc nécessaire de normaliser les formules (2.34) et (2.35) par la somme de toutes les aires d'intersection. Les formules (2.34) et (2.35) deviennent :

$$\begin{aligned} & \frac{\underline{Nb}_H(W \cap (D_i))}{\underline{Acc}_{k-1}(1-\alpha)^2 + \underline{Acc}_k(2-\alpha)^2 + \underline{Acc}_{k+1}(1+\alpha)^2 + \underline{Acc}_{k+2}\alpha^2} \quad (2.36) \\ &= \frac{\underline{Nb}_H(W \cap (D_i))}{6 - 4\alpha + 4\alpha^2} \end{aligned}$$

$$\begin{aligned} & \frac{\overline{Nb}_H(W \cap (D_i))}{\overline{Acc}_{k-1}(1-\alpha)^2 + \overline{Acc}_k(2-\alpha)^2 + \overline{Acc}_{k+1}(1+\alpha)^2 + \overline{Acc}_{k+2}\alpha^2} \quad (2.37) \\ &= \frac{\overline{Nb}_H(W \cap (D_i))}{6 - 4\alpha + 4\alpha^2} \end{aligned}$$

La méthode d'estimation du nombre de votes par excès et par défaut pour l'ensemble complémentaire W^c est illustré sur la figure 2.14.

Les équations (2.30) et (2.31) peuvent être détaillées en :

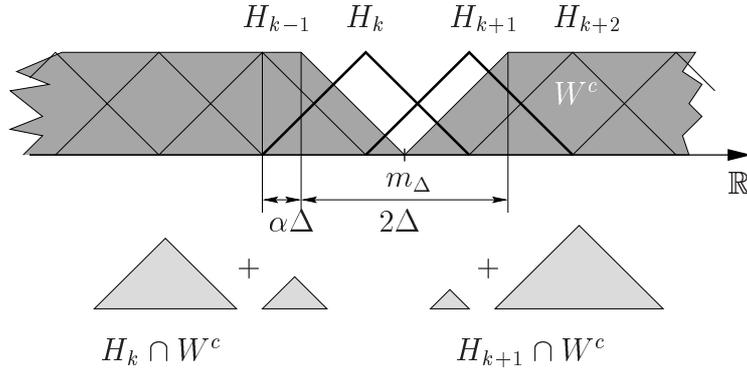


Figure 2.14 – Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle flou W^c .

$$\begin{aligned} \underline{Nb}_H(W^c \cap (D_i)) &= \sum_{k=0}^p \underline{Acc}_k + \\ &\quad \frac{1}{4}((1-\alpha)^2 + (1+\alpha)^2 - 4)\underline{Acc}_k + \frac{1}{4}((2-\alpha)^2 + \alpha^2 - 4)\underline{Acc}_{k+1} \\ \overline{Nb}_H(W^c \cap (D_i)) &= \sum_{k=0}^p \overline{Acc}_k + \\ &\quad \frac{1}{4}((1-\alpha)^2 + (1+\alpha)^2 - 4)\overline{Acc}_k + \frac{1}{4}((2-\alpha)^2 + \alpha^2 - 4)\overline{Acc}_{k+1} \end{aligned}$$

soit en simplifiant :

$$\underline{Nb}_H(W^c \cap (D_i)) = \sum_{k=0}^p \underline{Acc}_k + \left(\frac{\alpha^2}{2} - \frac{1}{2}\right)\underline{Acc}_k + \left(\frac{\alpha^2}{2} - \alpha\right)\underline{Acc}_{k+1} \quad (2.38)$$

$$\overline{Nb}_H(W^c \cap (D_i)) = \sum_{k=0}^p \overline{Acc}_k + \left(\frac{\alpha^2}{2} + \frac{1}{2}\right)\overline{Acc}_k + \left(\frac{\alpha^2}{2} - \alpha\right)\overline{Acc}_{k+1} \quad (2.39)$$

Comme pour les formules (2.34) et (2.35), lorsque tous les accumulateurs sont identiques ($Acc_0 = Acc_1 = \dots = Acc_p = A$), le nombre de votes pour W^c n'est pas indépendant de α . Une solution consiste alors à normaliser les équations (2.38) et (2.39). Cette normalisation est obtenue en remarquant que le nombre $\underline{Nb}_H(W^c \cap (D_i))$ doit être égal à $(p + \frac{1}{2})A$. On obtient alors :

$$\begin{aligned} \underline{Nb}_H(W^c \cap (D_i)) &= & (2.40) \\ (2p+1) \frac{\sum_{k=0}^p \underline{Acc}_k + \frac{1}{2}(\alpha^2 + 1) \underline{Acc}_k + \left(\frac{\alpha^2}{2} - \alpha\right) \underline{Acc}_{k+1}}{2(p + \alpha^2 - \alpha + \frac{3}{2})} \end{aligned}$$

$$\begin{aligned} \overline{Nb}_H(W^c \cap (D_i)) &= & (2.41) \\ (2p+1) \frac{\sum_{k=0}^p \overline{Acc}_k + \frac{1}{2}(\alpha^2 + 1) \overline{Acc}_k + \left(\frac{\alpha^2}{2} - \alpha\right) \overline{Acc}_{k+1}}{2(p + \alpha^2 - \alpha + \frac{3}{2})} \end{aligned}$$

On peut alors réécrire l'équation (2.26) en fonction des équations (2.36), (2.37), (2.40) et (2.41). Ceci permet d'exprimer un encadrement de $P(W; (D_i))$ en fonction des valeurs des accumulateurs imprécis et de la position de l'intervalle W paramétrée par α .

Cas de l'ensemble W de forme binaire

Dans un second temps, nous avons pensé qu'utiliser un intervalle flou pour l'estimation du nombre de votes $Nb_H(W \cap (D_i))$ rajoute inutilement une fuzzyfication à ce processus. Nous avons donc essayé d'estimer le nombre de votes avec un intervalle W binaire de même précision que les éléments de la partition floue. Pour que cet intervalle W conserve la même précision (aire de la fonction caractéristique) que celle de l'intervalle triangulaire, il faut que la largeur de celui-ci soit deux fois moins importante. la fonction caractéristique de celui-ci est alors la suivante :

$$\mu_W(x) = \begin{cases} 1 & \text{si } x \in [(m_W - \frac{\Delta}{2}, m_W + \frac{\Delta}{2}]] \\ 0 & \text{sinon} \end{cases} \quad (2.42)$$

Le transfert du nombre de votes de la partition vers W est illustré sur la figure 2.15. avec :

$$\begin{aligned} Aire(H_k) &= \Delta \\ Aire_{H_{k-1} \cap W} &= \frac{1}{2}(1 - \alpha)^2 \\ Aire_{H_k \cap W} &= \frac{1}{2} - \alpha^2 + \alpha \\ Aire_{H_{k+1} \cap W} &= \frac{1}{2}\alpha^2 \\ Aire_{H_i \cap W} &= 0 \forall i \neq \{k-1, k, k+1\} \end{aligned} \quad (2.43)$$

Les équations (2.28) et (2.29) peuvent alors s'écrire :

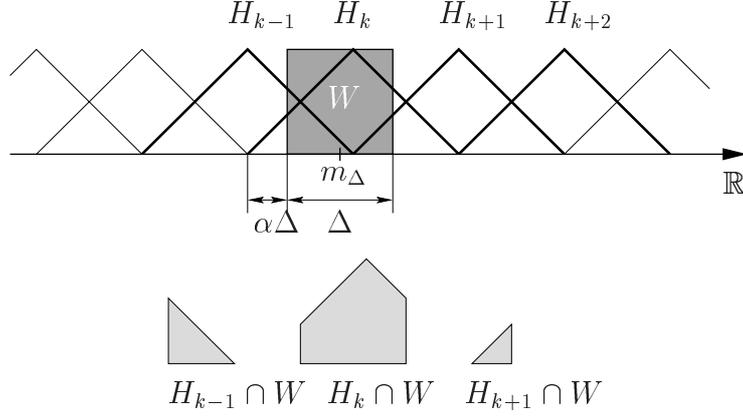


Figure 2.15 – Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle binaire W de granularité Δ .

$$\underline{Nb}_H(W \cap (D_i)) = \frac{1}{2} [\underline{Acc}_{k-1}(1 - \alpha)^2 + \underline{Acc}_k(1 + 2\alpha - 2\alpha^2) + \underline{Acc}_{k+1}\alpha^2] \quad (2.44)$$

$$\overline{Nb}_H(W \cap (D_i)) = \frac{1}{2} [\overline{Acc}_{k-1}(1 - \alpha)^2 + \overline{Acc}_k(1 + 2\alpha - 2\alpha^2) + \overline{Acc}_{k+1}\alpha^2] \quad (2.45)$$

Les équations (2.30) et (2.31) sont elles aussi changées en :

$$\underline{Nb}_H(W^c \cap (D_i)) = \sum_{i=0}^p Acc_k - \frac{1}{2} (\underline{Acc}_{k-1}(1 - \alpha)^2 + \underline{Acc}_k(1 + 2\alpha - 2\alpha^2) + \underline{Acc}_{k+1}\alpha^2)$$

$$\overline{Nb}_H(W^c \cap (D_i)) = \sum_{i=0}^p Acc_k - \frac{1}{2} (\overline{Acc}_{k-1}(1 - \alpha)^2 + \overline{Acc}_k(1 + 2\alpha - 2\alpha^2) + \overline{Acc}_{k+1}\alpha^2)$$

soit :

$$\underline{Nb}_H(W^c \cap (D_i)) = \sum_{i=0}^p Acc_k - \underline{Nb}_H(W \cap (D_i)) \quad (2.46)$$

$$\overline{Nb}_H(W^c \cap (D_i)) = \sum_{i=0}^p Acc_k - \overline{Nb}_H(W \cap (D_i)) \quad (2.47)$$

Contrairement à la méthode précédente d'estimation du nombre de votes, celle-ci ne souffre pas d'un défaut de normalisation. Il n'est donc pas nécessaire de normaliser les équations (2.44)-(2.47).

L'utilisation des formules (2.26), (2.44)-(2.47) permet de définir un encadrement de la probabilité d'avoir un élément de la distribution (D_i) dans n'importe quel sous-ensemble W de Φ . Cet encadrement s'écrit :

$$\frac{\underline{Nb}_H(W \cap (D_i))}{\underline{Nb}_H(W \cap (D_i)) - \overline{Nb}_H(W \cap (D_i)) + \sum_{k=0}^p \overline{Acc}_k} \leq P(W; (D_i)) \leq \frac{\overline{Nb}_H(W \cap (D_i))}{\overline{Nb}_H(W \cap (D_i)) - \underline{Nb}_H(W \cap (D_i)) + \sum_{k=0}^p \underline{Acc}_k} \quad (2.48)$$

Cas de l'ensemble W de forme binaire étendu

Le dernier cas présenté est celui où l'intervalle binaire W conserve le même support que les intervalles flous de la partition, c'est à dire 2Δ . La fonction caractéristique de W est alors la suivante :

$$\mu_W(x) = \begin{cases} 1 & \text{si } x \in [(m_W - \Delta, m_W + \Delta)] \\ 0 & \text{sinon} \end{cases} \quad (2.49)$$

La précision de cet intervalle est moins grande que dans le cas précédent. Cependant, le nombre d'intervalles flous de la partition couverts par W est plus important. On peut donc s'attendre à une plus grande fiabilité de l'estimation du nombre de votes.

Le transfert du nombre de votes de la partition vers l'intervalle W est illustré sur la figure 2.16 avec :

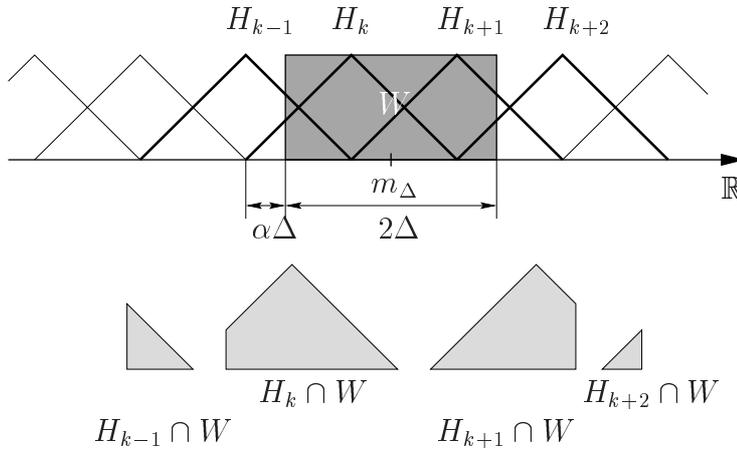


Figure 2.16 – Transfert du nombre de votes contenus dans les cellules de l'histogramme vers l'intervalle binaire W de granularité 2Δ .

$$\begin{aligned}
Aire(H_k) &= \Delta \\
Aire_{H_{k-1} \cap W} &= \left(\frac{1}{2} (1 - \alpha)^2 \right) \Delta \\
Aire_{H_k \cap W} &= \left(1 - \frac{\alpha^2}{2} \right) \Delta \\
Aire_{H_{k+1} \cap W} &= \left(1 - \frac{1}{2} (1 - \alpha)^2 \right) \Delta \\
Aire_{H_{k+2} \cap W} &= \left(\frac{1}{2} \alpha^2 \right) \Delta \\
Aire_{H_i \cap W} &= 0 \forall i \neq \{k-1, k, k+1, k+2\}
\end{aligned} \tag{2.50}$$

Les formules (2.28) et (2.29) peuvent alors s'écrire :

$$\begin{aligned}
\underline{Nb}_H(W \cap (D_i)) &= \frac{1}{2} \underline{Acc}_{k-1} (1 - \alpha)^2 + \underline{Acc}_k \left(1 - \frac{\alpha^2}{2} \right) \\
&\quad + \underline{Acc}_{k+1} \left(1 - \frac{1}{2} (1 - \alpha)^2 \right) + \underline{Acc}_{k+2} \frac{\alpha^2}{2}
\end{aligned} \tag{2.51}$$

$$\begin{aligned}
\overline{Nb}_H(W \cap (D_i)) &= \frac{1}{2} \overline{Acc}_{k-1} (1 - \alpha)^2 + \overline{Acc}_k \left(1 - \frac{\alpha^2}{2} \right) \\
&\quad + \overline{Acc}_{k+1} \left(1 - \frac{1}{2} (1 - \alpha)^2 \right) + \overline{Acc}_{k+2} \frac{\alpha^2}{2}
\end{aligned} \tag{2.52}$$

Cette estimation de $Nb_H(W \cap (D_i))$ n'est pas sujette au problème de normalisation rencontré dans la première méthode. De plus, on peut simplifier l'écriture des nombres de votes par excès et par défaut pour l'ensemble W^c de la même manière que dans le cas d'un sous ensemble W binaire de support Δ (formules (2.46) et (2.47)). On a alors :

$$\underline{Nb}_H(W^c \cap (D_i)) = \sum_{k=0}^p \underline{Acc}_k - \underline{Nb}_H(W \cap (D_i)) \tag{2.53}$$

$$\overline{Nb}_H(W^c \cap (D_i)) = \sum_{k=0}^p \overline{Acc}_k - \overline{Nb}_H(W \cap (D_i)) \tag{2.54}$$

La formule (2.48) reste valable en utilisant les valeurs de $\underline{Nb}_H(W \cap (D_i))$, $\overline{Nb}_H(W \cap (D_i))$, $\underline{Nb}_H(W^c \cap (D_i))$ et $\overline{Nb}_H(W^c \cap (D_i))$ définies dans les formules (2.51)-(2.54).

Nous allons maintenant comparer l'efficacité des ces trois méthodes d'estimation. Soit une variable aléatoire réelle gaussienne centrée sur 0 et d'écart type 1. La loi de cette densité de probabilité s'écrit :

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (2.55)$$

Supposons que l'on tire au hasard 2000 valeurs imprécises de cette variable. Les valeurs imprécises sont obtenues par troncature de données réelles. Ces 2000 valeurs $(D_i)_{i \in [1,2000]}$ sont accumulées sur un histogramme approché construit sur une portion de l'axe des réels entre -10 et 10. La partition construite sur cet intervalle est constituée de 11 sous ensemble flous comme illustré sur la figure 2.17.

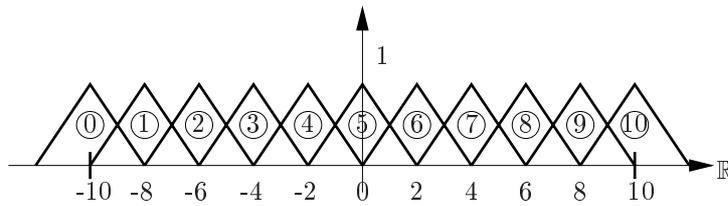


Figure 2.17 – Partitionnement flou de l'intervalle $[-10, 10]$ en 11 cellules floues.

Soit m_W la position du mode du sous ensemble flou W . Nous avons fait évoluer la position de m_W de -10 à +10 par pas de 0,02. Pour chacune de ces positions, et pour chacune des fonctions caractéristiques de W , nous avons estimé les valeurs des bornes de l'encadrement de $P(W; (D_i))$ définies dans la formule (2.26). La moyenne de ces deux bornes nous donne alors une estimation de la valeur de $P(W; (D_i))$. Pour comparer avec la véritable distribution de probabilité, nous avons calculé une approximation de $P(x \in W)$ en fonction de la loi gaussienne (équation (2.55)). Cette probabilité peut s'écrire :

$$P(x \in W) = \int_{B_i}^{B_s} p(x)\mu(x)dx \quad (2.56)$$

où B_i et B_s sont respectivement les bornes inférieure et supérieure du sous ensemble flou W et $p(x)$ est donné par l'équation (2.55). $\mu(x)$ est la fonction caractéristique de W . La formule (2.56) peut être approximée par :

$$P(x \in W) \simeq \Delta_R \times \sum_{k=0}^s p(B_i + k\Delta_R)\mu(B_i + k\Delta_R) \quad (2.57)$$

Les figures 2.18-2.20 illustrent la reconstruction de la probabilité $P(W; (D_i))$ ainsi que le calcul de $P(x \in W)$.

Sur les trois figures 2.18-2.20, la courbe en pointillés correspond à l'estimation de $P(x \in W)$, la courbe la plus haute correspond à l'estimation supérieure de $P(W; (D_i))$ et la courbe la plus basse correspond à l'estimation inférieure de $P(W; (D_i))$. On peut remarquer que l'écart type des deux reconstructions de $P(W; (D_i))$ est plus important que celui de $P(x \in W)$. Cependant, l'allure générale est respectée. Ceci s'explique très probablement par le fait que les données $(D_i)_{i \in [1, N]}$ sont imprécises ce qui a tendance

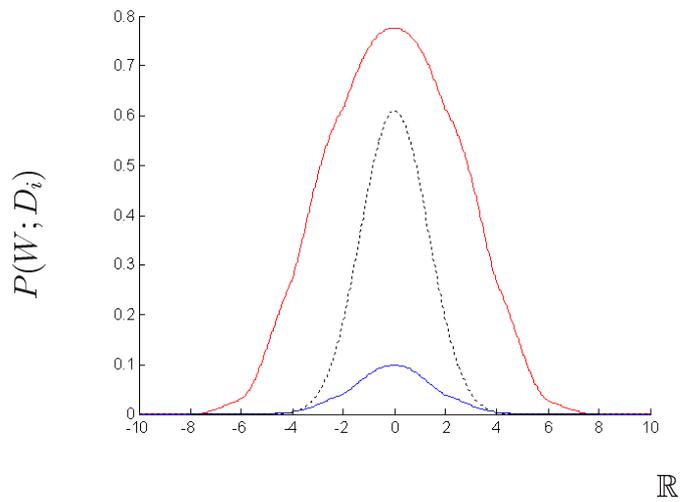


Figure 2.18 – Reconstruction par excès et par défaut de la distribution avec un sous-ensemble flou W de fonction caractéristique linéaire.

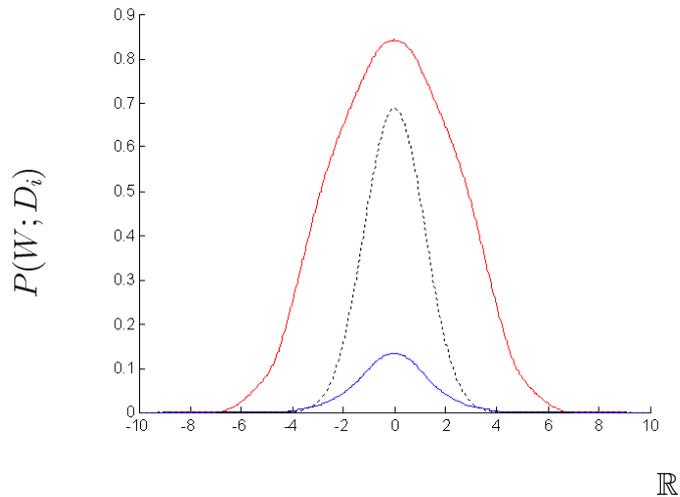


Figure 2.19 – Reconstruction de la distribution avec un sous-ensemble flou W binaire de support Δ .

à répartir l'influence des votes sur plusieurs intervalles de l'histogramme approché. De plus, le partitionnement flou a tendance à répartir le vote des données sur une zone plus grande. Comme on le voit sur la figure 2.18, l'allure de la reconstruction à l'aide du sous-ensemble flou W de forme triangulaire est moins satisfaisante. En effet, la courbe présente des distorsions, et ce principalement au niveau des modes de la partition floue.

Pour comparer les trois reconstruction, nous avons effectué une mesure de distance statistique (distance de Kullback-Leibler) entre chaque distribution reconstituée et la

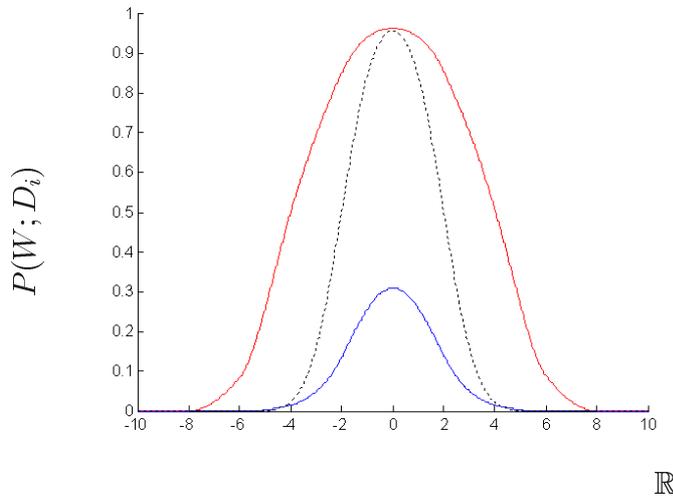


Figure 2.20 – Reconstruction de la distribution avec un sous-ensemble flou W binaire de support 2Δ .

distribution originale. Il est nécessaire de normaliser les différentes courbes de reconstruction. La formule de la distance de Kullback-Leibler entre deux distributions de probabilités discrètes P_1 et P_2 est donnée par l'équation suivante :

$$I_{KLdiv}(P_1, P_2) = \sum_{E \in \varepsilon} P_1(E) \log_2 \left(\frac{P_1(E)}{P_2(E)} \right) \quad (2.58)$$

avec \log_2 le logarithme en base 2, P_1 la distribution originale, P_2 la distribution approximée et E un événement de l'ensemble des événements ε . Cette distance a été calculée pour les trois types de sous ensembles W , et en faisant varier le nombre N de données. L'évolution de cette distance est représentée sur la figure 2.21.

Comme on peut le voir sur la figure 2.21, la valeur de la divergence se stabilise à partir de $N = 300$. La reconstruction avec un sous ensemble flou binaire de granularité 2Δ semble donner de meilleurs résultats puisque sa divergence par rapport à la courbe de $P(x \in W)$ est la plus faible quel que soit le nombre de données N . Nous garderons donc la forme binaire de granularité 2Δ pour le sous ensemble de reconstruction W .

Nous allons maintenant tester l'influence du choix de la position du partitionnement. Nous avons montré dans le paragraphe 2.1.2 que les histogrammes classiques sont fortement dépendants de la position de la partition. Qu'en est-il pour les histogrammes approchés? Nous avons pris le même ensemble de données $(D_i)_{i \in [1, N]}$ et nous avons décalé la portion de l'axe des réels couverte par la partition. Dans un premier temps, nous avons effectué une partition de 11 sous ensembles flous sur l'intervalle $[-10, 10]$, puis sur l'intervalle $[-9.5, 10.5]$ et enfin sur l'intervalle $[-9, 11]$. Pour visualiser l'influence du positionnement de la partition, la figure 2.22 présente les trois reconstructions lorsque le nombre de données. $N = 2000$.

Comme on peut le voir sur la figure 2.22, l'allure des trois reconstructions est

Figure 2.21 – Comparaison de la divergence de Kullback-Leibler pour les trois formes de W .

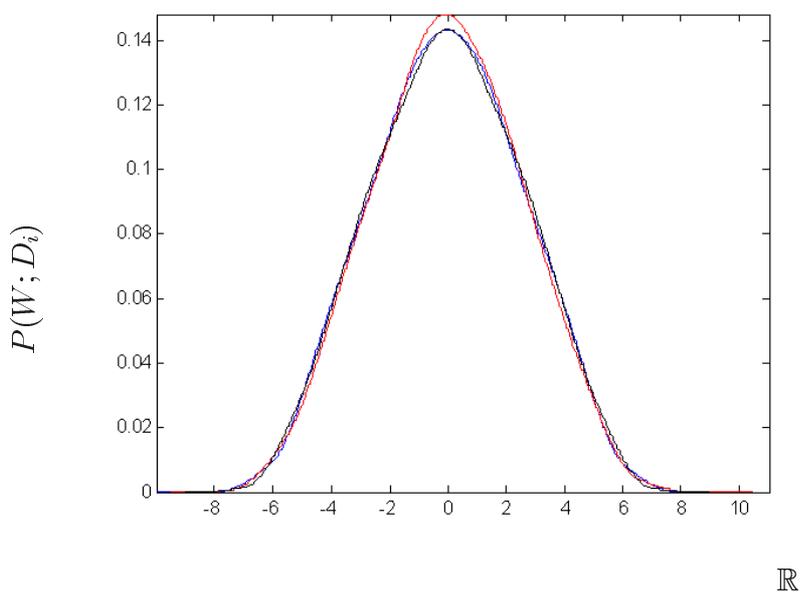
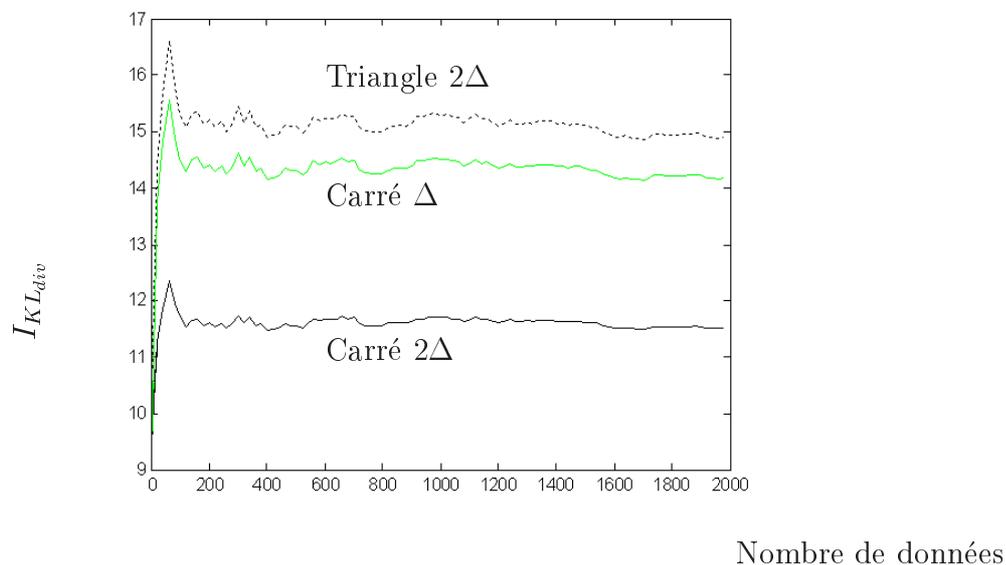


Figure 2.22 – Influence de la position du partitionnement sur l'estimation de la distribution.

semblable. On peut donc en déduire qu'avec les histogrammes approchés, la position de la partition n'a que peu d'effets sur la représentation de données.

L'approche que nous venons de présenter consiste en l'estimation de $P(W; (D_i))$

à partir du transfert pignistique des accumulateurs inférieurs et supérieurs, suivi du calcul des deux bornes de $P(W; (D_i))$ (respectivement $\underline{P}(W; (D_i))$ et $\overline{P}(W; (D_i))$) à partir de l'estimation des votes pour le sous ensemble W .

Une autre approche consiste à inverser cette procédure, c'est-à-dire, calculer l'encadrement de $P(H_k / (D_i))$ à savoir : $[\underline{P}(H_k / (D_i)), \overline{P}(H_k / (D_i))]$, et d'estimer $\underline{P}(W; (D_i))$ et $\overline{P}(W; (D_i))$ par transfert pignistique de $\underline{P}(H_k / (D_i))$ et $\overline{P}(H_k / (D_i))$.

L'encadrement de $P(H_k / (D_i))$ est défini par :

$$\begin{aligned}\underline{P}(H_k; (D_i)) &= \frac{\underline{Acc}_k}{\underline{Acc}_k + \sum_{i \neq k} \underline{Acc}_i} \\ \overline{P}(H_k; (D_i)) &= \frac{\overline{Acc}_k}{\overline{Acc}_k + \sum_{i \neq k} \overline{Acc}_i}\end{aligned}\quad (2.59)$$

Les bornes définissant l'encadrement de $P(W; (D_i))$ sont déterminées en effectuant un transfert pignistique des équations (2.59) vers l'intervalle W . Ceci s'écrit :

$$\sum_{k=0}^p \underline{P}(H_k; (D_i)) \frac{|W \cap H_k|}{|H_k|} \leq P(W; (D_i)) \leq \sum_{k=0}^p \overline{P}(H_k; (D_i)) \frac{|W \cap H_k|}{|H_k|} \quad (2.60)$$

soit pour un intervalle binaire W de granularité 2Δ :

$$\begin{aligned}& \frac{1}{2}(\underline{P}(H_{k-1}; (D_i)) (1 - \alpha)^2 + \underline{P}(H_k; (D_i)) (1 - \frac{\alpha^2}{2}) + \\ & \underline{P}(H_{k+1}; (D_i)) (1 - \frac{1}{2}(1 - \alpha)^2) + \underline{P}(H_{k+2}; (D_i)) \frac{\alpha^2}{2}) \\ \leq & P(W; (D_i)) \leq \\ & \frac{1}{2}(\overline{P}(H_{k-1}; (D_i)) (1 - \alpha)^2 + \overline{P}(H_k; (D_i)) (1 - \frac{\alpha^2}{2}) + \\ & \overline{P}(H_{k+1}; (D_i)) (1 - \frac{1}{2}(1 - \alpha)^2) + \overline{P}(H_{k+2}; (D_i)) \frac{\alpha^2}{2})\end{aligned}\quad (2.61)$$

Reprenons notre exemple de $(D_i)_{i \in [1, 2000]}$, 2000 données réelles imprécises issues d'une distribution de probabilité de loi gaussienne centrée sur 0 et d'écart type 1. Pour vérifier que cette seconde approche est valable, nous avons réalisé un grand nombre d'expériences. L'exemple que nous présentons ici est tout à fait représentatif de ces expériences. On peut voir sur la figure 2.23 deux divergences de Kullback-Leibler. La première compare les reconstructions de $P(W; (D_i))$ obtenues d'une part avec le transfert classique des nombres de votes (formule (2.48)) et d'autre part avec la loi gaussienne (formule (2.57)). La deuxième divergence compare la reconstruction utilisant le transfert pignistique des deux bornes de $P(H_k / (D_i))$ (formule (2.61)) avec celle utilisant la loi gaussienne. Ces deux comparaisons sont illustrées sur la figure 2.23. Pour estimer la distance entre les distributions de $P(x \in W)$ et de $P(W; (H_k))$, nous

avons considéré le centre de l'intervalle défini par $I_{KLdiv}(P(x \in W), \underline{P}(W; (H_k)))$ et $I_{KLdiv}(P(x \in W), \overline{P}(W; (H_k)))$ soit :

$$I_{KLVdiv} = \frac{1}{2} (I_{KLdiv}(P(x \in W), \underline{P}(W; (H_k))) + I_{KLdiv}(P(x \in W), \overline{P}(W; (H_k)))) \quad (2.62)$$

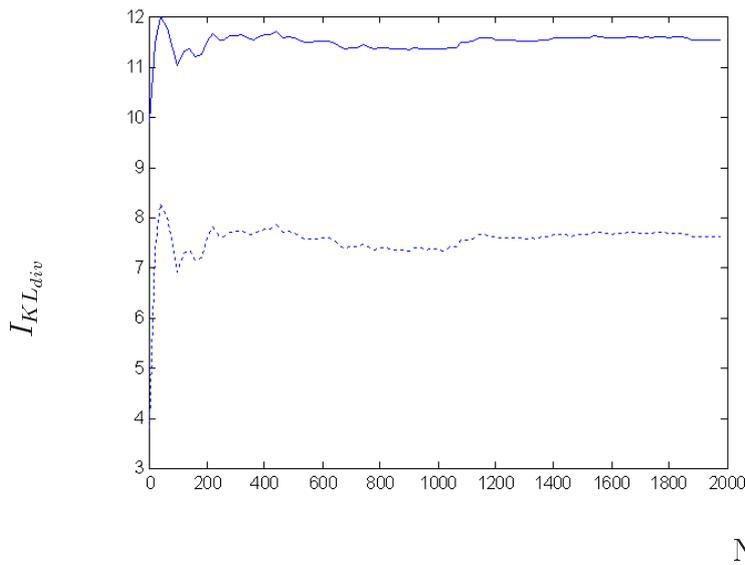


Figure 2.23 – En trait plein : la divergence de Kullback-Leibler entre l'estimation de $P(W/(D_i))$ avec le transfert classique et $P(x \in W)$, et en trait pointillé la divergence entre l'estimation de $P(W/(D_i))$ avec transfert des deux bornes de $P(H_k/(D_i))$ et $P(x \in W)$.

Contre toute attente, c'est le transfert des deux bornes de $P(H_k; (D_i))$ (formule (2.61)) qui obtient les résultats les plus proches de la courbe de probabilité théorique, au sens de la distance de Kullback-Leibler. Cette propriété a été vérifiée sur un grand nombre d'autres exemples de distribution de données mono et multi-modales.

Cette deuxième approche semble donc convenir pour l'estimation de $P(W; (D_i))$. Elle fournit une meilleure approximation de $P(W; (D_i))$ que la méthode utilisant le transfert classique des votes.

2.4 Applications des histogrammes approchés

Dans ce paragraphe, nous allons illustrer au travers de deux exemples l'aptitude des histogrammes approchés à servir d'outil de calcul statistique. Dans un premier temps, nous traiterons le cas de l'estimation d'un fractile d'une distribution. Puis nous aborderons le cas de l'estimation du mode d'une distribution.

2.4.1 Détermination d'un fractile ou de la médiane

Une application des histogrammes approchés est de servir d'outil complémentaire aux statistiques de rang pour déterminer la médiane ou un fractile d'ordre α ($\alpha \in [0, 1]$) d'une distribution de probabilité [SL99]. Le fractile empirique f_α d'ordre α de la distribution $(x_n)_{n \in [1, N]}$ est défini par $f_\alpha = x_{\varsigma(\alpha N)}$; où N est le nombre de données, et ς est une permutation triant les x_n par ordre croissant.

REMARQUE 4 *La médiane correspond en fait au fractile d'ordre 0,5.*

Avec une représentation des données par un histogramme classique, la recherche du fractile f_α revient à trouver les cellules C_i et C_s vérifiant $f_\alpha \in]m_i - \frac{\Delta}{2}, m_s + \frac{\Delta}{2}[$ avec m_k le centre de la classe C_k ($m_k = e_{\min} + k\Delta$), e_{\min} la borne inférieure du partitionnement et :

$$\sum_{k=1}^i Acc_k < \alpha N < \sum_{k=1}^s Acc_k \quad (2.63)$$

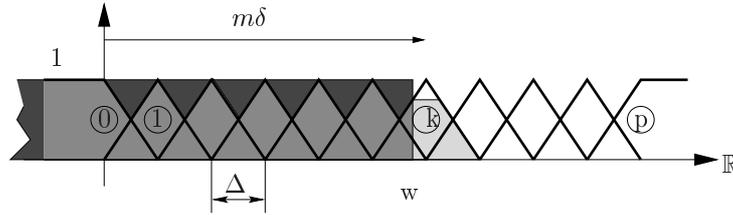
Généralement $s = i+1$ sauf dans le cas où certains intervalles de l'histogramme n'ont accumulé aucune donnée (histogramme contenant des classes creuses). On supposera pour simplifier l'écriture des formules que les histogrammes ne contiennent pas de classe creuse.

Comme on peut le voir sur la formule (2.63), la précision de l'estimation de α est limitée par l'échantillonnage de l'histogramme. Il est malgré tout possible de trouver une valeur "probable" de f_α , plus précise, en utilisant une interpolation linéaire [WW91]. Cependant, utiliser une interpolation linéaire c'est admettre a priori l'hypothèse que les données sont réparties uniformément sur les intervalles de l'histogramme.

Avec les histogrammes approchés, les bornes de l'encadrement de f_α sont définies en généralisant l'équation (2.63). L'encadrement de f_α revient en fait à trouver deux éléments fictifs w_i (resp. w_s) qui soient le plus grand des éléments de la distribution nécessairement plus petit que f_α (resp. le plus petit des éléments de la distribution nécessairement plus grand que f_α). Cette recherche fait intervenir la notion de rang nécessaire qui est introduite dans [DP88]. Dans le cas des données précises ($\overline{Acc}_k = \underline{Acc}_k = Acc_k$), le rang $r(w)$ d'un élément w de la distribution est alors compris entre deux bornes qui sont le rang par défaut $r^-(w)$ et le rang par excès $r^+(w)$ définis par :

$$r^-(w) = \sum_{k=0}^p Acc_k \cdot N(H_k < w) < r(w) < r^+(w) = \sum_{k=0}^p Acc_k \cdot \Pi(H_k < w) \quad (2.64)$$

où p est le nombre de cellules de l'histogramme, H_k est la $k^{ième}$ cellule de la partition floue, Acc_k est l'accumulateur associé à l'intervalle H_k de la partition et N est le nombre de données ayant voté. La figure 2.24 illustre le calcul du rang d'un élément w , avec 2Δ la granularité de la partition floue, et $p+1$ le nombre d'intervalles de la partition. En séparant en deux la double inégalité, on obtient les équations suivantes.

Figure 2.24 – Calcul du rang de w .

$$r^-(w) = \sum_{k=0}^p Acc_k \cdot N(H_k < w) < r(w) \quad (2.65)$$

$$r(w) < r^+(w) = \sum_{k=0}^p Acc_k \cdot \Pi(H_k < w) \quad (2.66)$$

avec :

$$\Pi(H_k < w) = \sup_{\mathbb{R}}(\min(\mu_{H_k}, \mu_w)) \quad (2.67)$$

$$N(H_k < w) = \inf_{\mathbb{R}}(\max(1 - \mu_{H_k}, \mu_w)) \quad (2.68)$$

et :

$$\begin{aligned} \mu_w &= 1 \forall x \in \mathbb{R}/x < w \\ &= 0 \text{ sinon} \end{aligned} \quad (2.69)$$

On peut alors déterminer un encadrement de l'élément fictif w , ayant le rang $r(w) = \alpha N$, en inversant les deux équations (2.65) et (2.66), puis en explicitant les sommes et le calcul des possibilités et nécessités. On obtient alors deux éléments w^- et w^+ . w^+ est l'élément fictif au-delà duquel aucun élément x ne peut avoir le rang $r(w)$. w^- est l'élément fictif en dessous duquel aucun élément ne peut avoir le rang $r(w)$.

$$w^- = \Delta \left[m + \frac{1}{Acc_{m-1}} \left(\alpha N - \sum_{k=0}^{m-1} Acc_k \right) \right] \quad (2.70)$$

$$w^+ = \Delta \left[m + \frac{1}{Acc_m} \left(\alpha N - \sum_{k=0}^m Acc_k \right) \right] \quad (2.71)$$

avec $e_{\min} + m\Delta$ la position du mode du $m^{\text{ième}}$ intervalle de la partition. L'intervalle H_m de la partition est le premier tel que $\Pi(H_k < w) < 1$.

Dans le cas de données imprécises, les équations (2.65) et (2.66) sont modifiées comme suit :

$$r^-(w) = \sum_{k=0}^p \underline{Acc}_k \cdot N(H_k < w) < r(w) \quad (2.72)$$

$$r(w) < r^+(w) = \sum_{k=0}^p \overline{Acc}_k \cdot \Pi(H_k < w) \quad (2.73)$$

où \overline{Acc} (resp. \underline{Acc}) est l'accumulateur supérieur (resp. inférieur) associé à la partition floue. Ici encore on peut développer ces équations pour obtenir :

$$\Delta \left[m^- + \frac{1}{\underline{Acc}_{m^- - 1}} \left(\alpha N - \sum_{k=0}^{m^- - 1} \underline{Acc}_k \right) \right] < w < \Delta \left[m^+ + \frac{1}{\overline{Acc}_{m^+}} \left(\alpha N - \sum_{k=0}^{m^+} \overline{Acc}_k \right) \right] \quad (2.74)$$

L'expérience que nous avons réalisée, consiste à estimer la médiane et le fractile d'ordre 0.28 de la distribution uniforme des premiers chiffres du tirage du loto de 1976 à 1998. La médiane théorique de cette distribution est 25. Puisqu'il s'agit d'un tirage aléatoire de données issues d'une distribution uniforme de nombres entre 1 et 49, le fractile d'ordre 0.28, quant à lui, est donné par : $f_{0.28} = 0.28 \times (49 - 1) + 1 = 14.44$.

Le fractile d'ordre α sera alors pris comme la moyenne entre w^+ et w^- déterminés à l'aide des formules (2.70) et (2.71).

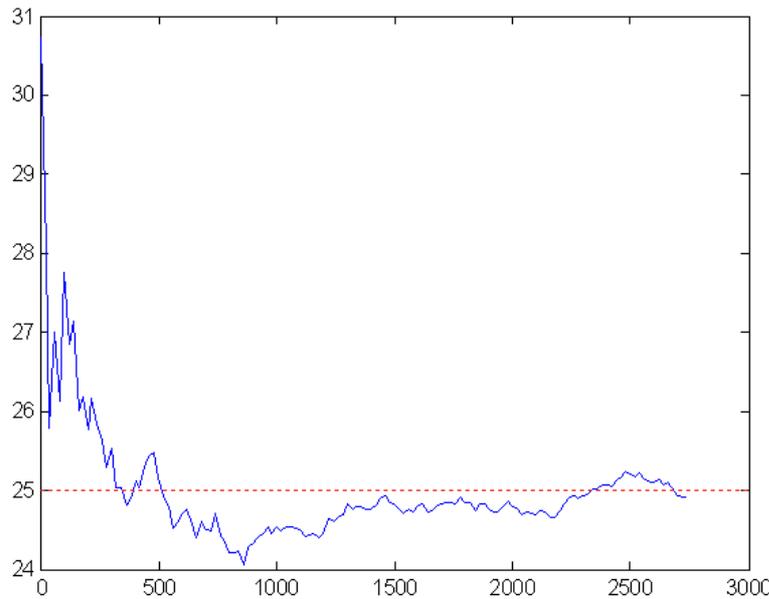


Figure 2.25 – Convergence de l'estimation de la médiane vers 25 en fonction du nombre de données.

La figure 2.25 illustre la convergence de l'estimateur de médiane à l'aide des histogrammes approchés vers la médiane théorique. Cette valeur est prise comme étant le centre de l'intervalle défini par $[w^-, w^+]$.

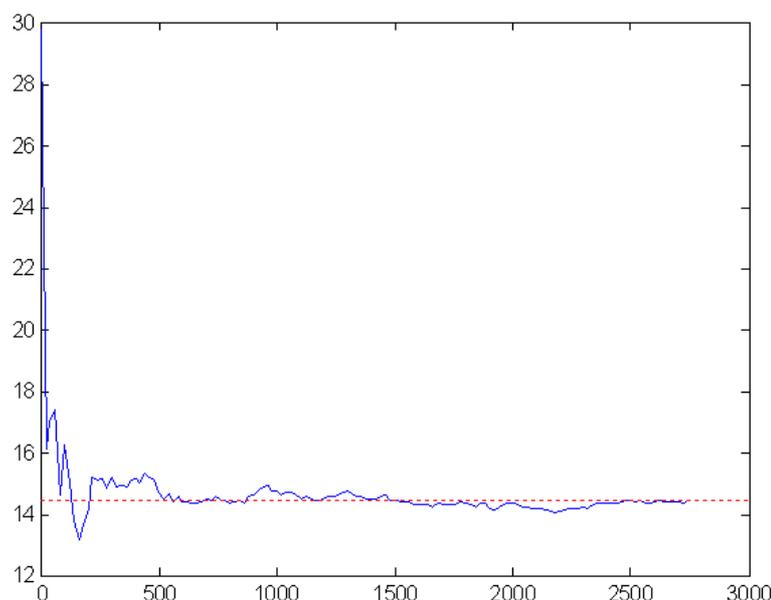


Figure 2.26 – Convergence de l'estimation du fractile d'ordre 0.28 en fonction du nombre de données.

La figure 2.26, quant à elle, illustre la convergence de l'estimation du fractile d'ordre 0.28. Le fait que le fractile ne soit pas entier (14.44) ne perturbe pas l'estimation.

Nous avons donc illustré le fait que les histogrammes approchés pouvaient servir d'outil de calcul pour la détermination d'un fractile d'une distribution. L'utilisation des histogrammes approchés possède plusieurs avantages. Tout d'abord, il est possible d'obtenir des fractiles d'ordre non entier. Des essais pour illustrer les bonnes propriétés et la robustesse de l'estimation des fractiles et de la médiane ont été réalisés dans [SL99]. L'estimateur basé sur les histogrammes approchés y est comparé aux estimateurs empirique et aux estimateurs de rang. Des essais restent cependant à effectuer lorsque les données sont imprécises.

2.4.2 Détermination du mode d'une distribution

Ce travail a plus particulièrement été développé pendant les trois années de ma thèse. La recherche du mode d'une distribution revient à chercher la valeur la plus fréquente prise par celle-ci. Dans le cas d'une distribution symétrique mono-modale, le mode, la médiane et la moyenne sont confondus. La valeur la plus fréquente de la distribution peut donc être estimée à l'aide d'un simple calcul de moyenne. Cependant,

lorsque la distribution n'est plus symétrique, ou bien possède plusieurs modes, les estimateurs de moyenne et de médiane ne donnent plus la valeur la plus fréquente [WW91]. Seul un estimateur spécifique du mode peut le permettre.

Si les données issues d'une distribution sont représentées par un histogramme classique, le mode appartient généralement à la cellule de l'histogramme ayant reçu le plus grand nombre de votes. La précision de localisation de ce mode dépend alors de la discrétisation de l'histogramme. La fiabilité de cette localisation dépend du pourcentage de votes recueilli par cette cellule.

Nous montrons maintenant que l'utilisation des histogrammes approchés permet une meilleure localisation du mode. Soit $(D_i)_{i \in [1, N]}$ une distribution de N données dont on veut estimer le mode. Ces données sont accumulées dans un histogramme approché. Estimer le mode de la distribution sous-jacente aux données $(D_i)_{i \in [1, N]}$ revient à chercher la position d'un sous-ensemble flou W qui aurait reçu un maximum de votes s'il avait été un élément de la partition de l'histogramme approché. Or, nous avons montré dans le paragraphe 2.3 que l'on pouvait obtenir un encadrement de ce nombre de votes (formule (2.26)) de manière continue. Pour déterminer un mode, il suffit donc de trouver l'encadrement du plus grand nombre de votes (localement ou globalement). L'estimation du mode se fait alors de manière continue [SCA00].

REMARQUE 5 Dans l'application de ma thèse, il s'agit de trouver un mouvement dominant (ou mouvement majoritaire des pixels), c'est-à-dire le mouvement le plus fréquent dans l'image. Si l'on envisage plusieurs déplacements, et que l'on fait voter les pixels de l'image pour chaque déplacement possible; le mouvement majoritaire sera celui ayant reçu le plus de votes. Ceci correspond au mode de la distribution sous-jacente à l'ensemble des mouvements.

Cas des données précises.

Dans un premier temps, nous nous intéressons au cas des données précises. Il faut donc estimer la position d'un sous ensemble flou W recevant un maximum de votes. W est pris comme étant un intervalle binaire de granularité 2Δ (nous avons vu au paragraphe 2.3 que c'est ce qui donne le meilleur résultat).

Dans le cas où les données sont précises, on peut estimer $P(W; (D_i))$ de façon précise :

$$P(W; (D_i)) = \frac{Nb_H(W \cap (D_i))}{\sum_{k=0}^p Acc_k} = \frac{Nb_H(W \cap (D_i))}{N} \quad (2.75)$$

Nous cherchons en fait parmi tous les sous-ensembles W de granularité 2Δ de \mathbb{R} , celui (\widehat{W}) ayant la plus grande probabilité de représenter l'ensemble des mesures. C'est-à-dire l'ensemble qui correspond au plus grand nombre d'éléments de la distribution. Rechercher \widehat{W} revient donc à évaluer $P(W; (D_i))$ pour chaque sous-ensemble W de

granularité 2Δ et à sélectionner \widehat{W} tel que :

$$\forall W \subset \Phi / |W| = 2\Delta, P(W; (D_i)) \leq P(\widehat{W} / (D_i)) \quad (2.76)$$

Du fait du partitionnement de l'espace réel, cette procédure peut être grandement simplifiée en remarquant quelques propriétés locales des transferts pignistiques.

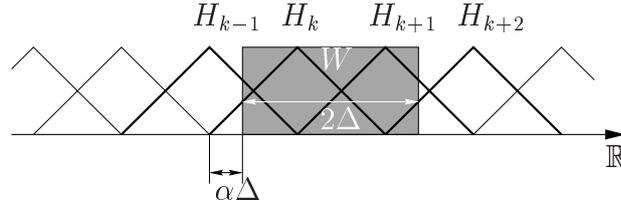


Figure 2.27 – Détermination du mode local pour les cellules $H_{k-1}, H_k, H_{k+1}, H_{k+2}$.

Sur la figure 2.27 est représenté un intervalle W de granularité 2Δ . On peut noter qu'il n'intersecte que quatre cellules de la partition respectivement $(H_{k-1}, H_k, H_{k+1}, H_{k+2})$. Si on note α la distance algébrique entre la borne inférieure de W et le mode de la première cellule intersectée (dans le sens des réels croissants), alors $P(W; (D_i))$ s'écrit sous la forme :

$$P(W; (D_i)) = \frac{1}{2N} [Acc_{k-1}(1-\alpha)^2 + Acc_k(1-\frac{\alpha^2}{2}) + Acc_{k+1}(1-\frac{1}{2}(1-\alpha)^2) + Acc_{k+2}\frac{\alpha^2}{2}] \quad (2.77)$$

Ainsi, pour tous les sous-ensembles W tels que $|W| = 2\Delta$ inclus dans l'intervalle $[m_{k-1}, m_{k+2}]$ il existe un unique W dont la probabilité est extremum. Sa position est fonction de α . Elle est obtenue en dérivant l'équation (2.77) par rapport à α et en résolvant :

$$\frac{dP(W; (D_i))}{d\alpha} = 0 \quad (2.78)$$

On trouve ainsi une valeur de α maximisant (ou minimisant) $P(W; (D_i))$. Cette solution doit respecter l'ensemble de définition de α qui est $[0, 1]$. Si elle existe, la solution de (2.78) est de la forme :

$$\widehat{\alpha} = \frac{Acc_{k-1} - Acc_{k+1}}{Acc_{k-1} - Acc_k - Acc_{k+1} + Acc_{k+2}} \quad (2.79)$$

Cette valeur de α correspond à la position de l'extremum local de la distribution sur l'intervalle $[m_k, m_{k+1}]$. Du fait de la géométrie du problème, la formule de recherche de $\widehat{\alpha}$ donnant un extremum local est valable pour chaque groupe de quatre cellules contiguës de la partition. Cette procédure est alors répétée pour chaque intervalle, et la position du mode global est donnée par $e_{\min} + (m_{Max} + \widehat{\alpha}_{Max})\Delta$ où $\widehat{\alpha}_{Max}$ est la valeur de α correspondant au plus grand des maxima locaux du nombre de votes contenus dans W , et m_{Max} le mode de la cellule de la partition associée au calcul de $\widehat{\alpha}_{Max}$.

L'exemple suivant illustre la détermination du mode dominant d'une distribution de données précises. Soit 2000 nombres réels précis issus d'une distribution aléatoire gaussienne centrée sur 3,4 et d'écart type unitaire. Ces données sont représentées sur un histogramme approché. Celui-ci est calculé sur l'intervalle $[-10, 10]$ qui est divisé en 11 cellules floues. La figure 2.28 présente la partition floue de l'intervalle considéré, ainsi que l'accumulateur associé à celle-ci.

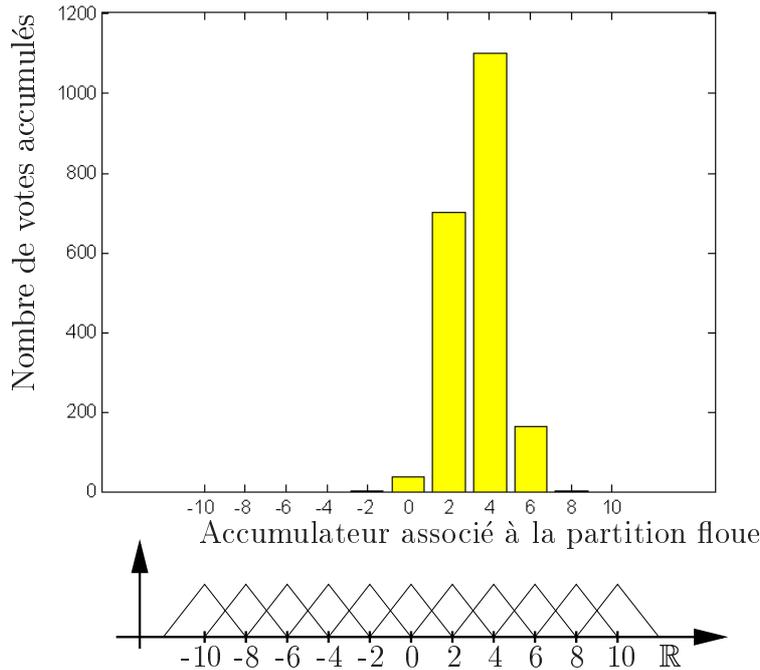


Figure 2.28 – Partition floue et accumulateur associé dans le cas de données précises.

La figure 2.29 illustre la reconstruction de $P(W; (D_i))$ et l'estimation du mode de la distribution. Nous avons représenté, à titre indicatif, l'histogramme normalisé, c'est-à-dire la suite des $(Acc_k)/N$. Le mode trouvé par une méthode d'estimation classique est l'intervalle $[3, 5]$. Notre méthode permet d'estimer une valeur plus précise de ce mode (ici 3.329).

Ces essais ont été effectués un certain nombre de fois avec des échantillons de différentes tailles. Sur la figure 2.30 est représentée l'estimation du mode de cette même distribution en fonction de la taille des échantillons.

A partir de peu de données, l'écart entre l'estimation du mode et sa valeur réelle est inférieur à la granularité de l'histogramme. On peut voir sur la figure 2.30 qu'un biais subsiste (dans le cas présenté il s'agit d'un biais par défaut). Ce biais peut s'expliquer aisément par l'hypothèse de monotonie de $P(X)$ dans une cellule, cette hypothèse étant justement fautive près du mode.

Considérons maintenant le cas où la distribution est contaminée par une autre distribution. Soit la distribution donnée dans l'exemple précédent contaminée à 30 % par une distribution gaussienne centrée sur -1.2 et d'écart type 1. La figure 2.31 illustre

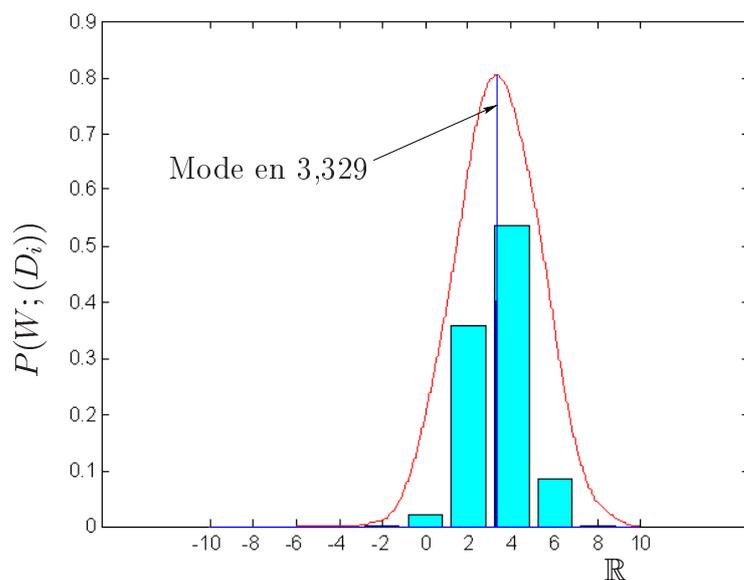
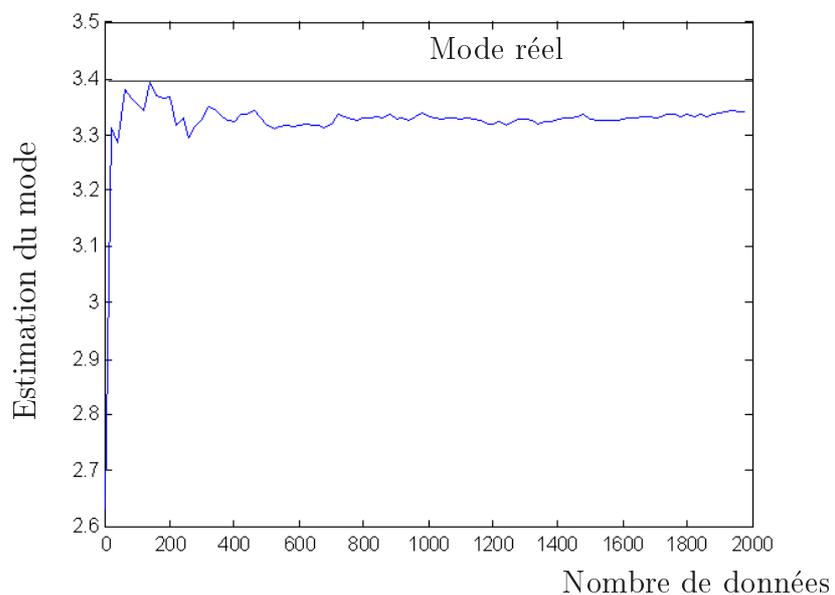
Figure 2.29 – Estimation continue du nombre de votes dans W , et histogramme précis.

Figure 2.30 – Evolution de l'estimation de la position du mode en fonction du nombre de données.

la partition floue et l'accumulateur associé normalisé. L'estimation de $P(W; (D_i))$ est aussi représentée. La position du mode principal est indiquée par un trait vertical et se situe en 3.156.

L'estimation du mode en fonction du nombre de données est présentée sur la figure 2.32. Ici aussi, l'écart entre le mode estimé et le vrai mode devient très rapidement

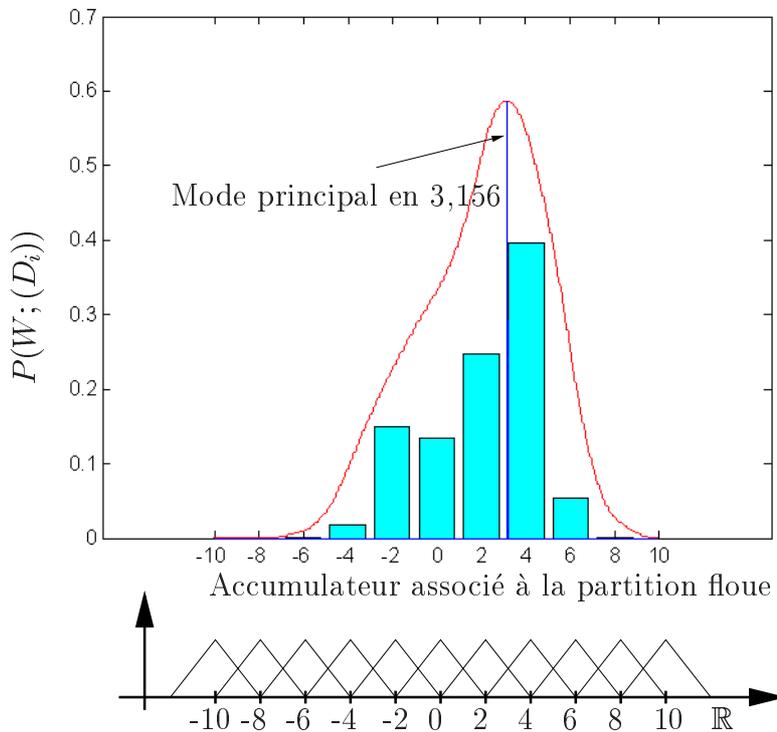


Figure 2.31 – Partition floue et accumulateur précis pour une distribution multi-modale.

inférieur à la granularité de l'histogramme.

La figure 2.32 illustre la convergence vers 3.18 de la position du mode à mesure que le nombre de données augmente. Certes, la précision de localisation du mode est moins bonne, mais elle reste meilleure que dans le cas de l'estimation avec un histogramme classique. En effet, l'estimation du mode avec une méthode classique aurait fourni l'intervalle $[3, 5]$ comme résultat.

Cas des données imprécises.

Lorsque les données sont imprécises, les deux accumulateurs supérieur et inférieur ne sont plus confondus. La simplification faite dans la formule (2.75) n'est plus possible. Il faut donc considérer l'encadrement de la valeur de $P(W; (D_i))$ par ses bornes inférieure et supérieure. Cet encadrement a été estimé à l'aide de deux approches présentées dans le paragraphe 2.3.

La première, utilisant le transfert pignistique du nombre de votes, définit les bornes de $P(W; (D_i))$ à l'aide des formules (2.26), (2.51)-(2.54). Cet encadrement (formule (2.80)) est paramétré par α et peut s'écrire :

$$\frac{Num_{\text{inf}}}{Den_{\text{inf}}} \leq P(W; (D_i)) \leq \frac{Num_{\text{sup}}}{Den_{\text{sup}}} \quad (2.80)$$

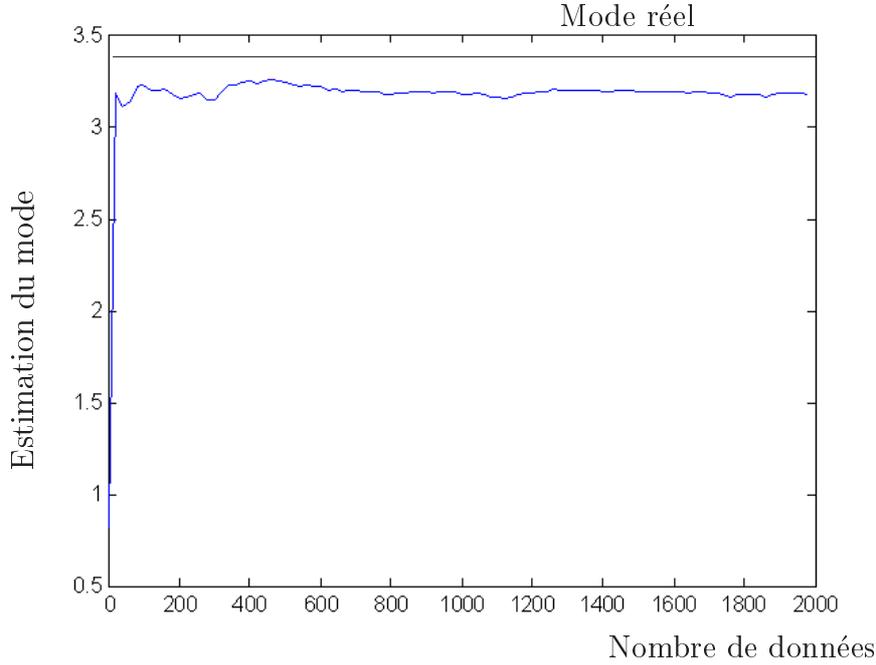


Figure 2.32 – Evolution de l'estimation de la position du mode en fonction du nombre de données.

avec :

$$Num_{\text{inf}} = \frac{1}{2}(\underline{Acc}_{k-1}(1-\alpha)^2 + \underline{Acc}_k(1-\frac{\alpha^2}{2}) + \underline{Acc}_{k+1}(1-\frac{1}{2}(1-\alpha)^2) + \underline{Acc}_{k+2}\frac{\alpha^2}{2}) \quad (2.81)$$

$$Num_{\text{sup}} = \frac{1}{2}(\overline{Acc}_{k-1}(1-\alpha)^2 + \overline{Acc}_k(1-\frac{\alpha^2}{2}) + \overline{Acc}_{k+1}(1-\frac{1}{2}(1-\alpha)^2) + \overline{Acc}_{k+2}\frac{\alpha^2}{2}) \quad (2.82)$$

$$Den_{\text{inf}} = Num_{\text{inf}} - Num_{\text{sup}} + \sum_{k=0}^p \overline{Acc}_k \quad (2.83)$$

$$Den_{\text{sup}} = Num_{\text{sup}} - Num_{\text{inf}} + \sum_{k=0}^p \underline{Acc}_k \quad (2.84)$$

Comme dans le cas des données précises, les valeurs de α maximisant localement la borne de $P(W; (D_i))$ sont déterminées en résolvant l'équation :

$$\frac{d}{d\alpha} (P(W; (D_i))) = 0 \quad (2.85)$$

avec $P(W; (D_i)) = f(\underline{P}(W; (D_i)), \overline{P}(W; (D_i)))$. Faut de mieux, nous avons choisi de prendre pour f la moyenne entre $\underline{P}(W; (D_i))$ et $\overline{P}(W; (D_i))$. L'équation (2.85) s'écrit donc :

$$\frac{d}{d\alpha} \left(\frac{Num_{\text{inf}}(\alpha)}{Den_{\text{inf}}(\alpha)} + \frac{Num_{\text{sup}}(\alpha)}{Den_{\text{sup}}(\alpha)} \right) = 0 \quad (2.86)$$

Contrairement au cas des données précises, la résolution analytique de l'équation (2.86) ne nous semble pas possible. On pourrait cependant trouver une approximation de $\hat{\alpha}$ par une méthode numérique. Celle-ci consisterait à parcourir l'axe de la partition de manière très fine et à rechercher tous les α vérifiant l'équation (2.86). Toutefois, nous cherchons à mettre au point une méthode de faible complexité, et le parcours de l'axe de la partition à la recherche de $\hat{\alpha}$ risque d'être pénalisant.

Nous nous sommes alors tourné vers l'approche utilisant le transfert des probabilités. L'expression analytique des bornes de $P(W; (D_i))$ est donnée dans la formule (2.61). La recherche de $\hat{\alpha}$ consiste en la résolution de l'équation (2.85), c'est à dire :

$$\frac{d}{d\alpha} \left(\begin{array}{c} [\underline{P}(H_{k-1}; (D_i)) + \overline{P}(H_{k-1}; (D_i))] \frac{1}{2} (1 - \alpha)^2 + \\ [\underline{P}(H_k; (D_i)) + \overline{P}(H_k; (D_i))] (1 - \frac{\alpha^2}{2}) + \\ [\underline{P}(H_{k+1}; (D_i)) + \overline{P}(H_{k+1}; (D_i))] (1 - \frac{1}{2} (1 - \alpha)^2) + \\ [\underline{P}(H_{k+2}; (D_i)) + \overline{P}(H_{k+2}; (D_i))] \frac{\alpha^2}{2} \end{array} \right) = 0 \quad (2.87)$$

La solution de cette équation est du même type que celle définie dans l'équation (2.79) :

$$\hat{\alpha} = \frac{Num}{Den} \quad (2.88)$$

avec :

$$Num = \underline{P}(H_{k-1}; (D_i)) + \overline{P}(H_{k-1}; (D_i)) - \underline{P}(H_{k+1}; (D_i)) - \overline{P}(H_{k+1}; (D_i)) \quad (2.89)$$

$$Den = \underline{P}(H_{k-1}; (D_i)) + \overline{P}(H_{k-1}; (D_i)) - \underline{P}(H_k; (D_i)) - \overline{P}(H_k; (D_i)) - \underline{P}(H_{k+1}; (D_i)) - \overline{P}(H_{k+1}; (D_i)) + \underline{P}(H_{k+2}; (D_i)) + \overline{P}(H_{k+2}; (D_i)) \quad (2.90)$$

Cette approche permet donc d'obtenir une solution littérale de $\hat{\alpha}$ sur l'intervalle $[m_k, m_{k+1}]$. L'estimation du mode global reprend la procédure définie dans le cas des données précises.

Pour illustrer la détermination du mode nous avons conservé les deux exemples du paragraphe précédent. Soit, dans un premier temps, à estimer le mode d'une distribution constituée de 2000 données réelles imprécises issues d'une distribution gaussienne centrée sur 3.4 et d'écart type 1.

De nombreux essais ont été réalisés. Nous présentons ici les résultats de l'un de ces essais très représentatif de l'ensemble des mesures. La figure 2.33 présente la partition floue de l'intervalle $[-10, 10]$, ainsi que l'accumulateur imprécis associé à celle-ci.

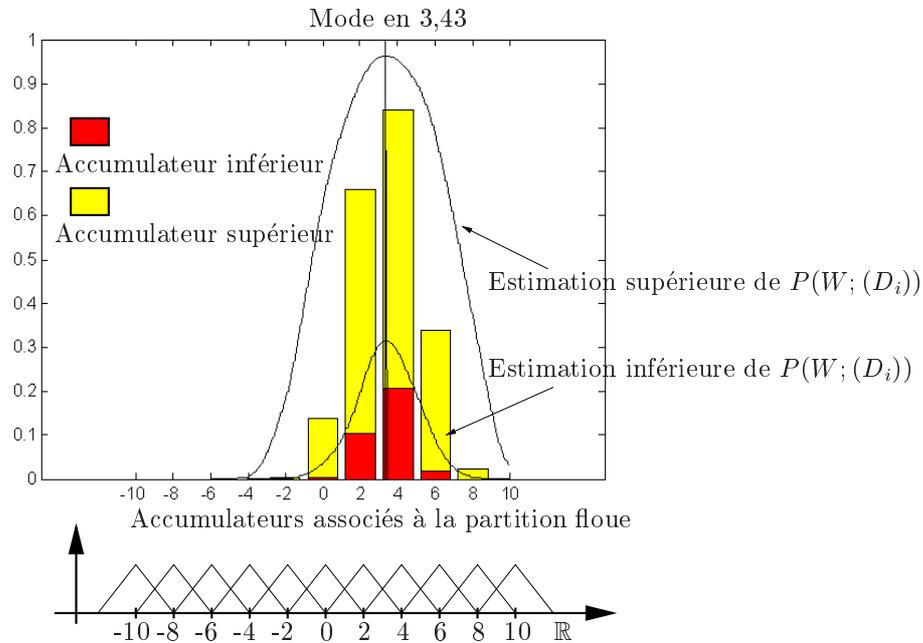


Figure 2.33 – Partition floue et accumulateur associé dans le cas de données imprécises.

Les bornes supérieures et inférieures de $P(W; (D_i))$ sont aussi représentées sur la figure 2.33 ainsi que l'estimation du mode. L'estimation du mode par une méthode classique aurait donné l'intervalle $[3, 5]$. Notre méthode d'estimation est plus précise et permet de fournir une estimation du mode à 3.43 (au lieu de 3.4).

Nous avons aussi réalisé des essais visant à estimer le mode avec un nombre de données variable. La figure 2.34 présente l'évolution de l'une de ces estimations du mode en fonction du nombre de données.

On peut voir que l'estimation du mode devient vite inférieure à la granularité de l'histogramme. Elle est d'ailleurs très rapidement proche de la valeur réelle du mode (beaucoup plus proche que quand les données sont précises).

Reprenons le deuxième exemple où la distribution précédemment étudiée est contaminée à 30% par une autre distribution gaussienne centrée sur -1,2 et d'écart type 1. Les résultats présentés sur la figure 2.35 sont représentatifs des nombreux essais réalisés. La figure 2.35 représente la partition floue de l'intervalle $[-10, 10]$, les deux bornes de $P(W; (D_i))$ et l'estimation du mode principal de $P(W; (D_i))$.

La figure 2.36 illustre quant à elle l'évolution de l'estimation du mode en fonction du nombre de données. On peut remarquer qu'une fois de plus, l'estimation du mode semble biaisée. En effet celle-ci se stabilise à 3.18 alors que le véritable mode de la distribution est en 3.4. Cependant, la précision de l'estimation du mode est supérieure à la granularité de l'histogramme.

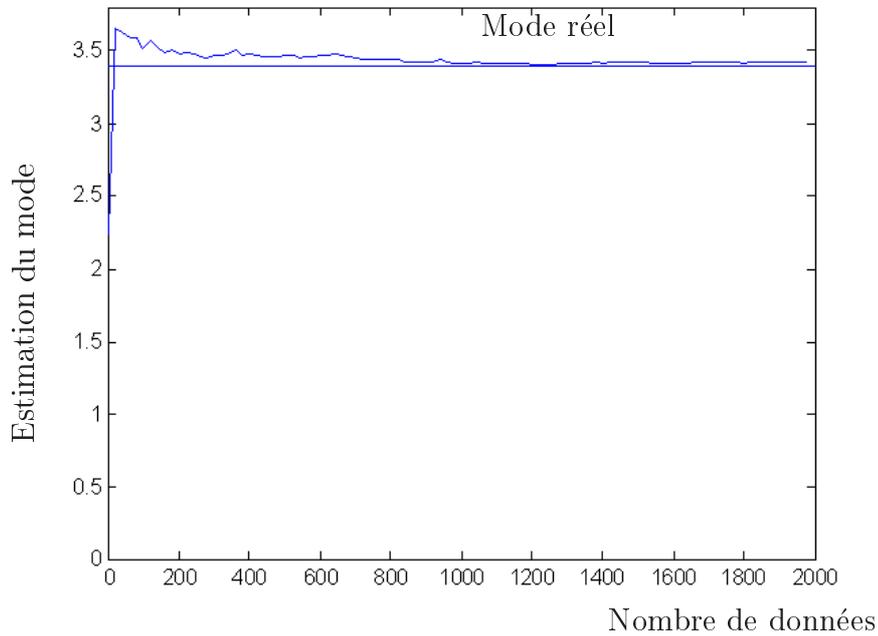


Figure 2.34 – Evolution de l'estimation du mode en fonction du nombre de données. Le véritable mode de la distribution se situe en 3.4.

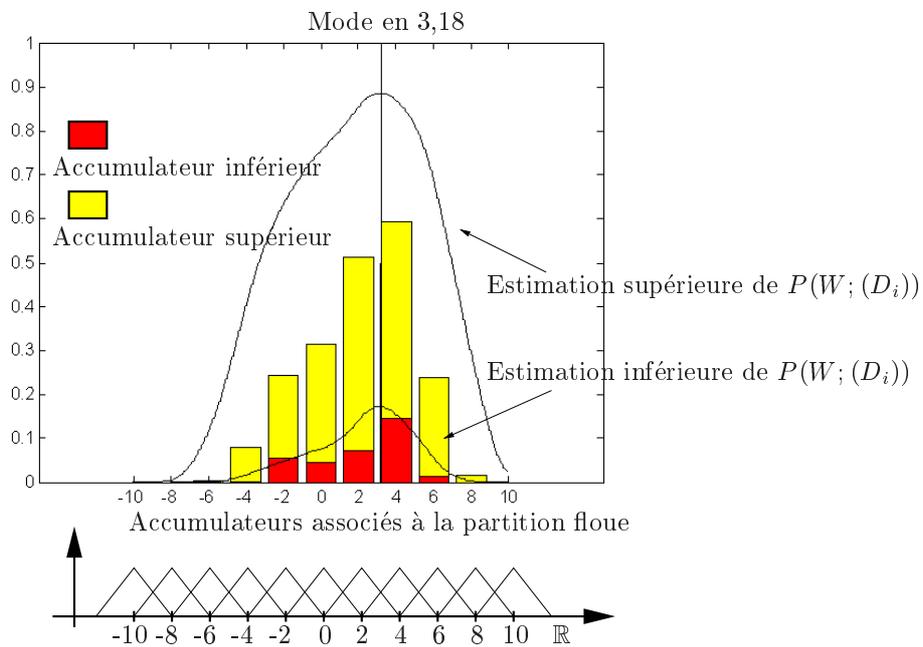


Figure 2.35 – Partition floue et accumulateurs associés pour une distribution multimodale dont le mode principal (70% des données) est en 3.4 et le mode secondaire (30% des données) est en -1.2.

Dans le dernier cas présenté, une méthode classique d'estimation du mode donne

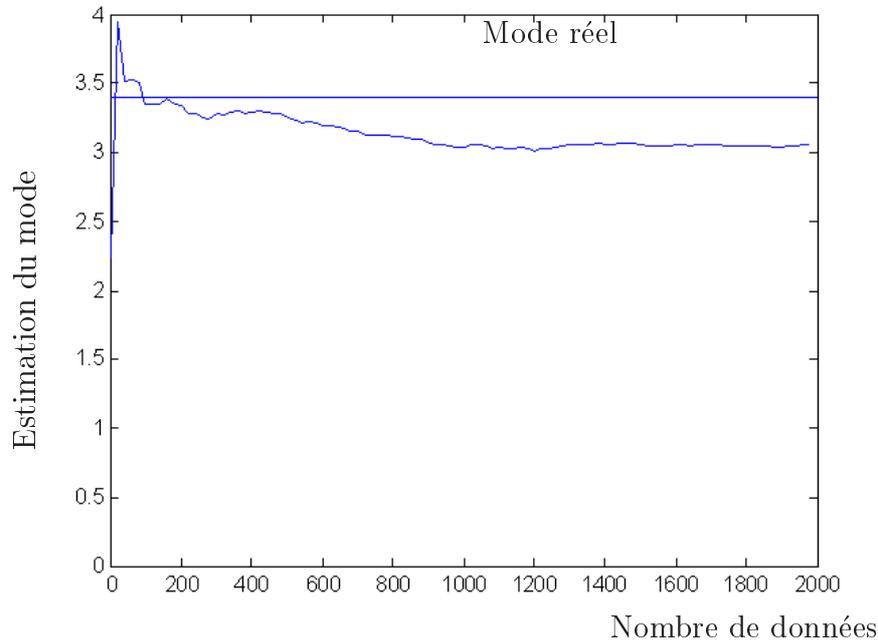


Figure 2.36 – Evolution de l'estimation du mode en fonction du nombre de données. Le véritable mode principal de la distribution se situe en 3.4.

l'intervalle $[3, 5]$, la moyenne de la distribution est estimée à 2.05, et la médiane à 2.89. On voit bien que ces trois estimateurs ne sont pas bien adaptés pour estimer le mode de la distribution. Notre estimateur est plus approprié pour estimer le mode d'une distribution de données contaminée.

2.5 Généralisation des histogrammes approchés en 2-D puis N-D

2.5.1 Histogrammes approchés en dimension 2

Jusqu'à maintenant, nous n'avons abordé que le cas de distributions mono dimensionnelles. Si la distribution est bi-dimensionnelle, l'histogramme approché sera lui aussi de dimension 2. Les cellules triangulaires de la partition floue 1-D se généralisent en cellules pyramidales en considérant la *t*-norme *min* [DP88]. L'allure de la partition 2-D construite sur un pavé $[e_{\min}, e_{\max}] \times [f_{\min}, f_{\max}]$ est représentée sur la figure 2.37.

Pour ce qui est de l'accumulation dans l'histogramme 2-D, les formules (2.18) et (2.19) se généralisent simplement en :

$$\overline{Acc_{k,l}} = \sum_{i=1}^N \sum_{j=1}^M \Pi(H_{k,l}; D_{i,j}) \quad (2.91)$$

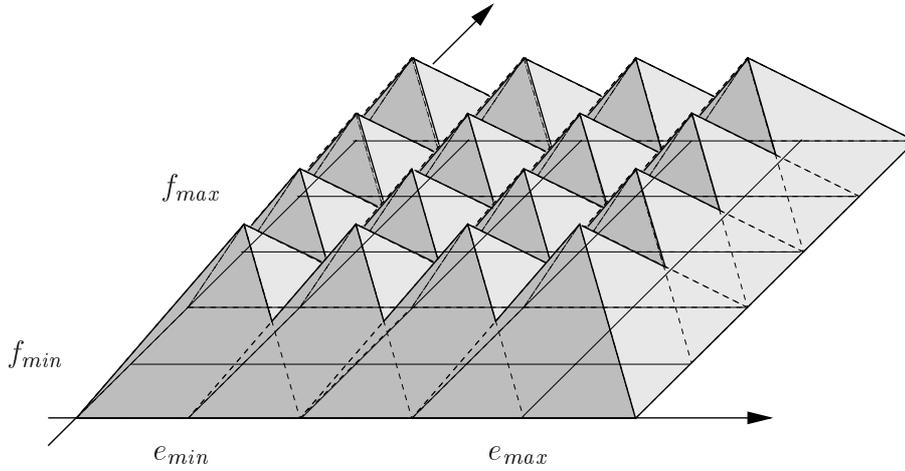


Figure 2.37 – Partition floue en deux dimensions sur l'espace $[e_{min}, e_{max}] \cdot [f_{min}, f_{max}]$.

$$\underline{Acc}_{k,l} = \sum_{i=1}^N \sum_{j=1}^M N(H_{k,l}; D_{i,j}) \quad (2.92)$$

où $H_{k,l}$ est la cellule située à la $k^{ième}$ ligne et à la $l^{ième}$ colonne de la partition 2-D. Avec la représentation des données que nous avons choisie (intervalles flous de fonction caractéristique triangulaire ou trapézoïdale), $D_{i,j} = D_i \times D_j$, (D_i, D_j) étant des intervalles flous représentant les projections de la donnée et \times le produit cartésien obtenu avec la t-norme min. L'expression de $\Pi(H_{k,l}; D_{i,j})$ et $N(H_{k,l}; D_{i,j})$ s'écrit :

$$\Pi(H_{k,l}; D_{i,j}) = \min(\Pi(H_k; D_i); \Pi(H_l; D_j)) \quad (2.93)$$

$$N(H_{k,l}; D_{i,j}) = \min(N(H_k; D_i); N(H_l; D_j)) \quad (2.94)$$

où H_k et H_l sont les projections de la cellule floue 2-D. Si les données sont précises, là encore, les deux accumulateurs sont confondus.

A priori, il est donc possible de généraliser toutes les formules proposées dans le cas 1-D. Cependant, dans le cas 2-D, les calculs de transferts pignistiques font intervenir les volumes d'intersection entre des pyramides et des cubes, et les formules des bornes de $P(W; (D_i))$ deviennent alors complexes.

Nous nous intéressons ici simplement à l'estimation du mode local de granularité $2\Delta \times 2\Delta$ dans un pavé de 4×4 cellules (généralisation de la recherche du mode local de granularité 2Δ dans un intervalle de quatre cellules adjacentes). A cause de la complexité de l'expression des bornes de $P(W; (D_i))$, nous n'avons pas réussi à déterminer la formule analytique de la position du mode (généralisation de l'équation (2.88)).

En vue de simplifier le problème, nous nous sommes tournés vers une approche projective. Cette approche suppose que le partitionnement de l'axe des réels est suffisamment fin pour qu'il n'y ait qu'un seul mode de la distribution dans chaque groupe de 4 cellules. Le calcul du mode 2-D sera alors ramené localement au calcul de deux modes 1-D. Ceci est illustré sur la figure 2.38.

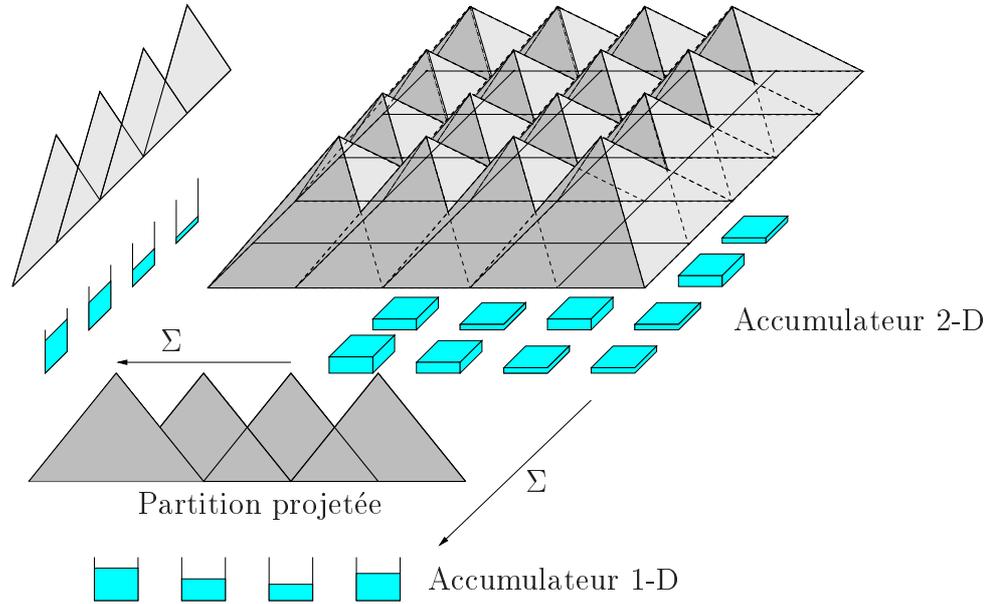


Figure 2.38 – Projection de la partition 2-D et de son accumulateur associé.

2.5.2 Evaluation de l'erreur de projection

L'hypothèse sous-jacente à la détermination du mode par les projections est que la distribution des données est localement séparable. C'est à dire que dans tout pavé $2\Delta \times 2\Delta$ de Ω on peut approcher la probabilité $p(e, f)$ par $p(e) \times p(f)$ avec $e \in [e_{\min}, e_{\max}]$ et $f \in [f_{\min}, f_{\max}]$. Si tel est le cas, rechercher localement le sous-ensemble W de $[m_k, m_{k+1}] \times [n_k, n_{k+1}]$ maximisant $P(W; (D_i))$ revient à chercher U (resp. V), sous-ensemble de $[m_k, m_{k+1}]$ (resp. $[n_k, n_{k+1}]$), maximisant $P(U/(D_i))$ (resp. $P(V/(D_i))$).

Dans le cas où l'hypothèse de séparabilité serait enfreinte, on peut s'attendre à une erreur d'estimation de $P(W; (D_i))$. Pour évaluer cette erreur, nous utilisons une approche inspirée des travaux de thèse de Christian Borgelt [Bor00], qui proposait de comparer la distribution reconstruite par projection avec la distribution originale. De nouveau, la distance de Kullback-Leibler nous permet de comparer les deux distributions.

La divergence de Kullback-Leibler est donnée dans la formule (2.58). Dans le cas présent, P_1 représente la distribution originale, c'est à dire celle issue de la distribution en 2-D. P_2 représente la distribution reconstituée à partir des projections. La formule (2.58) s'écrit alors :

$$I_{KLdiv}(P_1, P_2) = \sum_{e=e_1}^{e_4} \sum_{f=f_1}^{f_4} P_1(e, f) \times \log_2 \left(\frac{P_1(e, f)}{P_2(e) \times P_2(f)} \right) \quad (2.95)$$

avec :

$$P_2(e) = \sum_{i=1}^4 P_1(e, f_i) \quad (2.96)$$

$$P_2(f) = \sum_{i=1}^4 P_1(e_i, f)$$

Si l'on utilise les équations (2.95) et (2.96), on obtient alors :

$$I_{KLdiv}(P_1, P_2) = \sum_{i=1}^4 \sum_{j=1}^4 P_1(x_i, y_j) \left(\log_2(P_1(x_i, y_j)) - \log_2(\sum_{k=1}^4 P_1(x_i, y_k)) - \log_2(\sum_{k=1}^4 P_1(x_k, y_j)) \right) \quad (2.97)$$

Nous cherchons à savoir si la probabilité locale 2-D sous-jacente à l'accumulateur peut être assimilée à celle constituée du produit de ses deux projections marginales. Nous avons donc simulé un grand nombre d'accumulateurs 2-D de taille $M \times N$, puis évalué numériquement la divergence (formule (2.97)) entre la distribution originale et celle reconstituée à partir des projections, et ce pour chaque bloc 4×4 issu de l'accumulateur. L'exemple présenté illustre parfaitement les résultats obtenus pour nos nombreuses simulations. Il s'agit d'un accumulateur comportant 21×21 cellules. La distance de Kullback-Leibler a été calculée pour chaque bloc de 4×4 cellules et sa représentation graphique est illustrée sur la figure 2.40.

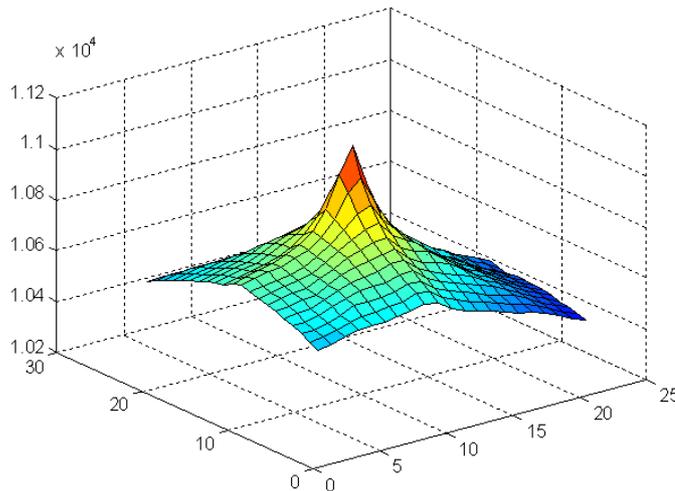


Figure 2.39 – Accumulateur construit sur une partition de 21 par 21 cellules.

La figure 2.40 contient $(21-4) \times (21-4)$ soit 289 calculs de distance. L'écart maximum est de 7.54×10^{-6} et l'écart moyen est de 5.3×10^{-7} . Donnons une interprétation plus pragmatique de cette distance. Si on ajoute aléatoirement des données dans les cellules

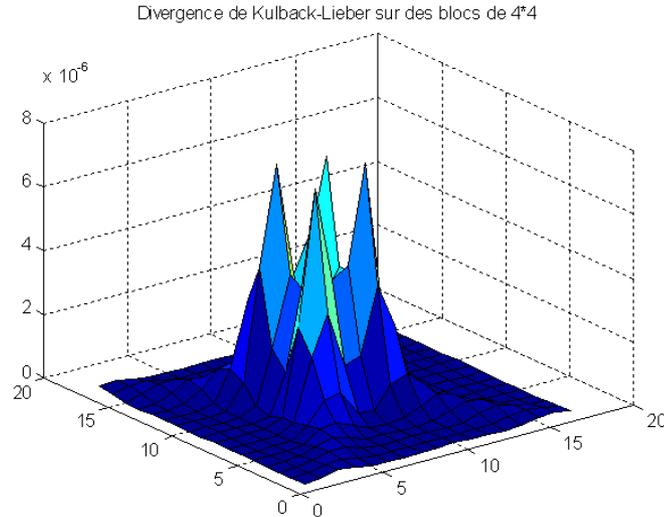


Figure 2.40 – Représentation graphique de la distance de Kullback-Lieber entre la distribution d'origine et la distribution projetée.

de l'histogramme, et que l'on calcule la distance entre les deux distributions obtenues, une distance de 10^{-5} correspond à un ajout de 16 votes sur 10^4 votes (soit une erreur inférieure à 0.15%).

On peut donc s'accorder sur le fait que l'approche projective introduit une très faible erreur sur l'estimation du mode. De plus elle aura pour effet de réduire énormément la complexité des calculs.

2.5.3 Histogrammes approchés en dimension N

Pour une variable mono-dimensionnelle, la partition floue fait intervenir des cellules triangulaires, donc de dimension 2. Pour une variable bi-dimensionnelle, la partition floue est constituée de pyramides (formes géométriques de dimension 3). En fait, la partition floue a une représentation graphique de dimension $N+1$ si la variable considérée est de dimension N . Au-delà de la dimension 2 la représentation graphique de la partition floue devient donc compliquée. Le calcul du transfert pignistique est alors impossible. Cependant, le raisonnement utilisant les projections reste valable. Toutes les formules (2.92)-(2.94) peuvent facilement s'étendre en dimension N . On peut alors les ré-écrire en :

$$\overline{Acc_H} = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} \Pi(H; D_{i_1, \dots, i_N}) \quad (2.98)$$

$$\underline{Acc_H} = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} N(H; D_{i_1, \dots, i_N}) \quad (2.99)$$

avec :

$$\Pi(H; D_{i_1, \dots, i_N}) = \min(\Pi(H_{i_1}; D_{i_1}), \dots, \Pi(H_{i_N}; D_{i_N})) \quad (2.100)$$

$$N(H; D_{i_1, \dots, i_N}) = \min(N(H_{i_1}; D_{i_1}), \dots, N(H_{i_N}; D_{i_N})) \quad (2.101)$$

où H_{i_k} est la projection de la cellule H sur la $k^{\text{ième}}$ dimension de l'espace considéré.

La détermination du mode ne peut alors s'effectuer qu'en raisonnant avec les projections.

2.6 Conclusion

Ce chapitre a présenté le nouvel outil de calcul statistique dont nous nous servirons pour l'estimation du mouvement majoritaire. Cet outil, appelé histogramme approché, s'appuie sur la théorie des sous-ensembles flous grossiers pour définir une extension des histogrammes classiques. Soient $(D_i)_{i \in [1, N]}$ N données. Les deux différences majeures entre les histogrammes classiques et les histogrammes approchés sont :

- Un histogramme approché est construit sur une partition floue de l'espace des données.
- L'accumulateur associé à chaque cellule de la partition est imprécis.

Une telle extension permet de sortir l'histogramme de son rôle classique de visualisation de la répartition des données pour en faire un véritable outil de calcul statistique. Nous avons vu dans le paragraphe 2.4 comment utiliser les histogrammes approchés pour estimer des fractiles et les modes d'une distribution de données.

Les spécificités des histogrammes approchés nous ont permis de remarquer :

- d'une part, que l'utilisation d'une partition floue de l'espace permet d'atténuer la dualité *précision/incertitude* généralement associée aux techniques de vote. Il a entre autre été montré que la précision d'estimation des modes et des fractiles était plus fine que la granularité de la partition de l'histogramme.
- d'autre part, que l'utilisation d'un accumulateur imprécis permet de prendre en compte l'imprécision des données, chose qui est rarement prévue avec les estimateurs classiques.

Cependant, tout n'a pas été étudié sur les histogrammes approchés. Beaucoup de travail théorique reste en suspens, notamment en ce qui concerne le principe d'estimation d'une densité de probabilité.

Le premier point théorique à approfondir concerne l'estimation de $P(W; (D_i))$ à partir des bornes inférieures et supérieures $\underline{P}(W; (D_i))$ et $\overline{P}(W; (D_i))$. Actuellement, nous nous contentons de prendre comme estimation de $P(W; (D_i))$ le centre de l'intervalle $[\underline{P}(W; (D_i)), \overline{P}(W; (D_i))]$. Une ouverture envisageable consiste en la minimisation de la distance entre $P(W; (D_i))$ et une fonction de $\underline{P}(W; (D_i))$ et $\overline{P}(W; (D_i))$. L'utilisation de la distance classique $d(a, b) = |a - b|$ conduit à définir $P(W; (D_i)) = \frac{\underline{P}(W; (D_i)) + \overline{P}(W; (D_i))}{2}$. Cependant, il serait plus judicieux d'utiliser des distances plus

spécifiques aux distributions de probabilité telles que les distances de Kullback-Leibler, de Bhattacharyya, ... [Bas88].

Une étude plus approfondie sur l'incidence de la granularité du partitionnement sur la reconstruction de $P(W; (D_i))$ serait aussi très utile. Elle permettrait certainement d'expliquer plus précisément les phénomènes d'étalement des distributions reconstruites. De plus, cette étude pourrait permettre d'établir un critère déterminant si la partition est en adéquation avec les données. En effet, nous avons pu constater que si la granularité de la partition est trop fine par rapport à l'imprécision des données, l'accumulateur inférieur est identiquement nul.

Un troisième point à étudier serait l'élaboration d'un critère permettant de qualifier la qualité de l'estimation de $P(W; (D_i))$. Pour l'instant nous avons comparé la densité estimée avec la vraie densité à l'aide de la divergence de Kullback-Leibler, mais comment faire si cette dernière n'est pas connue? Il faudrait envisager un critère qui ne soit fonction que des deux bornes de l'estimation de $P(W; (D_i))$. Nous avons pour l'instant essayé d'observer le rapport des deux bornes, mais sans pouvoir en tirer de conclusion.

Lors du passage en 2-D, nous avons été confronté à un problème pour calculer les formules de transfert pignistique de manière littérale. Une nouvelle méthode de calcul de ces transferts pignistiques pourrait permettre de travailler directement sur les distributions 2-D (voir N-D) sans avoir à calculer les projections. En l'occurrence, nous pensons utiliser une généralisation de la méthode géométrique de découpe de polygones de Southerland-Hodgman [SH74] pour calculer les volumes d'intersection entre W et (H_k) . Notre souci principal est d'obtenir une méthode de faible complexité. Il faudra donc comparer les performances de cette méthode avec celle utilisant les projections.

En ce qui concerne l'estimation des fractiles, seul le cas des données précises a été traité. Le cas des données imprécises a tout juste été abordé. Nous avons proposé quelques bases, et il faudrait réaliser des essais pour valider ces bases, notamment en ce qui concerne l'estimation d'un rang imprécis.

Chapitre 3

Estimation du mouvement apparent

Comme nous l'avons vu au chapitre 1, lorsque qu'une caméra se déplace dans un environnement, l'intensité des pixels de l'image qu'elle acquiert varie. En regardant les images on a une impression de mouvement plan, et c'est ce mouvement que l'on appelle mouvement apparent. Ce déplacement est une projection du mouvement 3-D dans le plan de l'image.

Si l'environnement dans lequel la caméra évolue est majoritairement statique, le mouvement apparent est directement relié au mouvement de la caméra. C'est dans ce cadre que s'inscrit le travail présenté dans ce chapitre. Il présente une méthode statistique d'estimation du mouvement apparent dominant. Le terme dominant signifie que ce mouvement est celui de la majorité des pixels de l'image. La méthode d'estimation de mouvement que nous proposons repose sur un processus d'accumulation dans un histogramme approché.

Ce chapitre est organisé de la manière suivante : le premier paragraphe traite de généralités sur le modèle de mouvement à estimer et de son influence sur l'estimation du mouvement dominant. Le deuxième paragraphe propose d'utiliser une modélisation floue du niveau de gris des pixels afin de rendre plus robuste le processus d'estimation de mouvement. Un processus multi-échelles issu de cette représentation floue du niveau de gris est aussi décrit. Le troisième paragraphe explicite comment les variations du niveau de gris d'un pixel sont reliées à son éventualité de déplacement. La méthode pour estimer l'éventualité de déplacement d'un pixel en fonction de la transformation considérée est présentée dans le paragraphe quatre. Le paragraphe cinq décrit le processus d'estimation du mouvement apparent dominant à l'aide des histogrammes approchés. Le sixième paragraphe propose une comparaison entre les différentes méthodes d'estimation de mouvement majoritaire au travers d'une série d'expérimentations. Le paragraphe sept présente l'apport en gain de temps de notre processus multi-échelles. Finalement, une conclusion dresse le bilan sur les apports et les inconvénients de notre approche et ouvrira de nouvelles perspectives à ce travail.

3.1 Principe de base

Nous avons déjà défini le mouvement apparent dominant dans le chapitre 1. Rappelons que si une caméra se déplace dans un environnement majoritairement statique, des variations cohérentes se produisent entre les niveaux de gris de la séquence d'images qu'elle acquiert. Cela donne une impression d'un mouvement, et c'est ce mouvement dans le plan image que nous souhaitons estimer.

3.1.1 Modèle de mouvement de la caméra

Généralement le repère associé à une caméra est défini en assimilant l'axe optique à l'axe Z et son plan sagittal au plan (X, Y) . Les axes X et Y sont choisis de façon à être parallèles aux côtés de la rétine. Une caméra libre dans l'espace peut effectuer un mouvement à six degrés de liberté décomposables en trois translations et trois rotations. Ceci est illustré sur la figure 3.1.

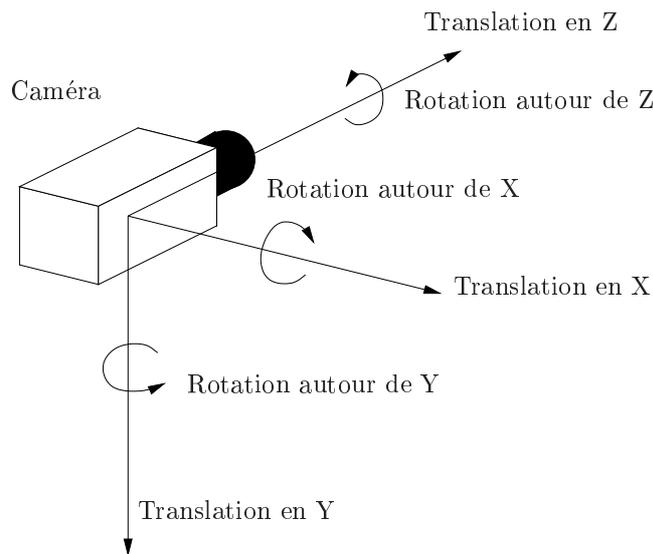


Figure 3.1 – Composante du mouvement d'une caméra dans l'espace.

Voyons tout d'abord les effets de chacun de ces déplacements tridimensionnels sur le mouvement des projections des objets fixes perçus par la caméra. Lorsque la caméra avance selon l'axe des Z , les projections des objets ont tendance à grossir. Réciproquement, si l'environnement de la caméra est fixe, un grossissement des objets est caractéristique d'une avancée suivant l'axe des Z , tandis qu'un rétrécissement est caractéristique d'un recul. Par analogie avec la photo, on appelle généralement cette transformation un effet de zoom. En utilisant le même raisonnement, un déplacement positif de la caméra suivant l'axe des X correspond à une translation de la projection des objets d'une quantité négative suivant l'axe des X (et vice-versa). Il en est de même

pour les translations suivant l'axe des Y . L'effet d'une rotation de la caméra autour de l'axe Z dans le sens trigonométrique (resp. anti-trigonométrique) se traduit par une rotation des objets projetés dans le sens anti-trigonométrique (resp. trigonométrique). En ce qui concerne les rotations autour des axes X et Y , les effets sont un peu plus complexes. Les déformations engendrées du fait des changements de perspective sont non-linéaires. La figure 3.2 illustre l'effet des rotations de la caméra autour des axes X et Y sur la projection des objets dans le plan image.

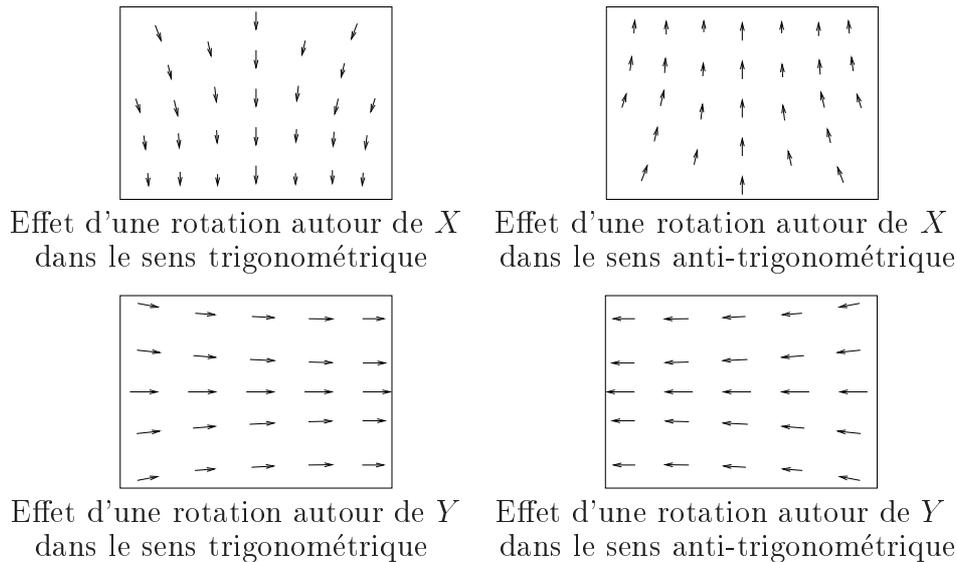


Figure 3.2 – Effet des rotations autour des axes X et Y sur la projection des objets.

En toute rigueur, on voit qu'un mouvement à 6 degrés de liberté de la caméra induit pour chaque point de l'image un mouvement apparent à six degrés de liberté. Un tel mouvement n'est pas identifiable en général. On ne peut l'estimer qu'en émettant des hypothèses de régularité du mouvement et de rigidité de l'environnement [SK99].

Cependant, si la scène filmée n'a pas trop de profondeur de champ, si les objets sont suffisamment éloignés de l'objectif de la caméra et si le mouvement de la caméra est de faible amplitude, alors la scène filmée peut être assimilée à un plan (entre deux images). Ainsi, tous les pixels de l'image subissent le même mouvement à six degrés de liberté. Toutes ces hypothèses semblent très contraignantes. Elles sont cependant facilement vérifiables dans la plupart des séquences d'images réelles. Tout du moins ces remarques sont vraies pour la majorité des pixels de l'image.

La difficulté de l'estimation des petits mouvements vient du fait qu'il est complexe de différencier tous les mouvements projetés [MB96]. En effet, comme cela est illustré dans la figure 3.3, une petite rotation autour de l'axe Y et une translation suivant l'axe X sont très similaires. Bien sûr, les déformations des objets ne sont pas rigoureusement les mêmes puisque les objets n'appartiennent pas à un plan. Cependant, dans les deux cas, l'observateur ne pourra identifier qu'un déplacement horizontal.

C'est pourquoi un modèle à quatre degrés de liberté est usuellement préféré à un modèle à six degrés de liberté. Tenter d'identifier un modèle à six degrés de liberté

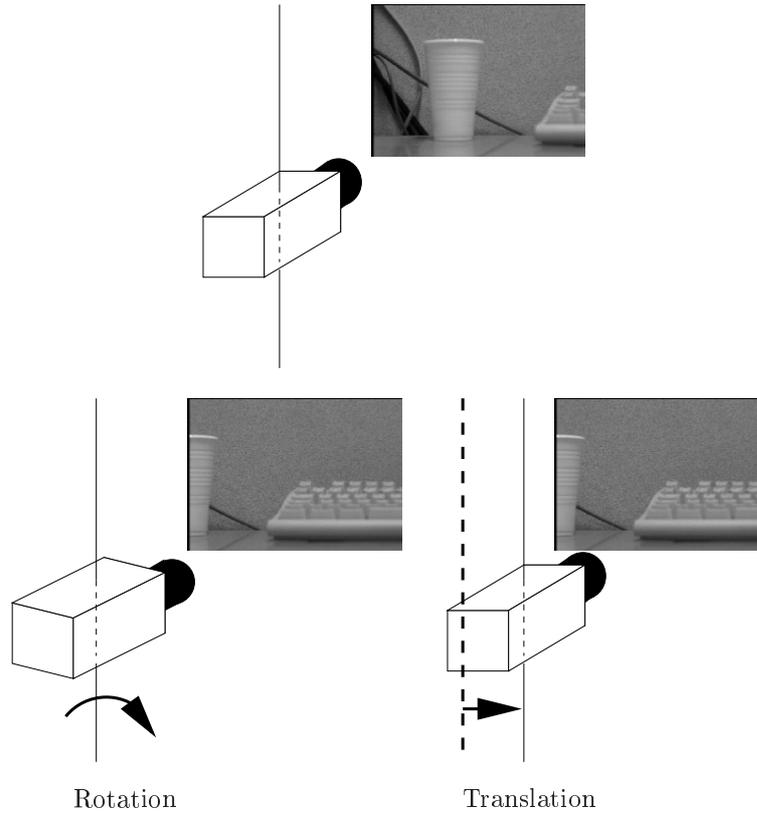


Figure 3.3 – Similitude entre une rotation autour de l’axe Y et une translation le long de l’axe X.

lorsque seulement quatre degrés sont perceptibles entraîne généralement une perte de la robustesse d’un identificateur [Lju99]. Nous proposons de modéliser le mouvement dominant apparent à l’aide de quatre paramètres (modèle affine [Fau93]) représentant deux translations dans le plan sagittal (pour les deux translations dans le plan sagittal et les deux rotations autour des axes de ce même plan), une rotation autour de l’axe Z normal au plan image (R_ϕ), et un facteur d’homothétie (α) que nous appellerons zoom par analogie avec le dispositif optique. Nous appellerons T_{hor} la translation horizontale et T_{ver} la translation verticale.

Comme les transformations de repères ne sont pas permutable, nous avons fixé un ordre arbitraire d’application des transformations, à savoir : translations suivant les axes horizontaux et verticaux, zoom et enfin rotation.

Considérons maintenant un point P de l’image repéré par le couple (lin_P, col_P) . Les notations sont précisées sur la figure 3.4 où (lin_C, col_C) représente les coordonnées du centre de l’image. Lors d’une translation de la caméra, le point P se retrouve décalé d’une quantité (T_{hor}, T_{vert}) en P' . Où T_{hor} (resp. T_{vert}) correspond à la translation horizontale (resp. verticale). Soit en notation matricielle :

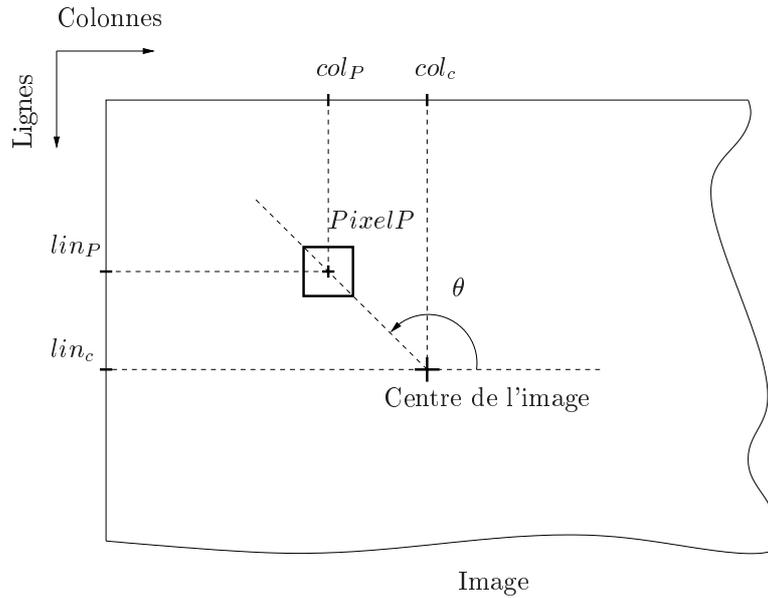


Figure 3.4 – Variables utilisées pour la modélisation du mouvement apparent.

$$P' = P - T \quad (3.1)$$

avec $P' = \begin{pmatrix} lin_{P'} \\ col_{P'} \end{pmatrix}$, $P = \begin{pmatrix} lin_P \\ col_P \end{pmatrix}$ et $T = \begin{pmatrix} T_{hor} \\ T_{ver} \end{pmatrix}$.

REMARQUE 6 *Les différentes transformations entre deux images ne sont généralement pas entières. Ainsi, les coordonnées du point P' ne sont, elles non plus, généralement pas entières. On a donc $(lin_{P'}, col_{P'}) \in \mathbb{R}^2$.*

Dans le cas où la scène est filmée de suffisamment loin pour être assimilée à un plan, la transformation induite par un zoom correspond en fait à une homothétie ayant pour centre F . En représentation de profil, l'effet du zoom peut être modélisé par la figure 3.5. Dans le plan image l'effet du zoom est illustré par la figure 3.6 :

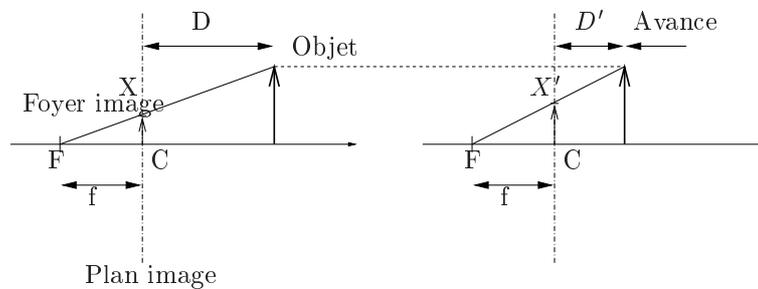


Figure 3.5 – Effet de l'avance de la caméra sur la projection d'un objet sur le plan image.

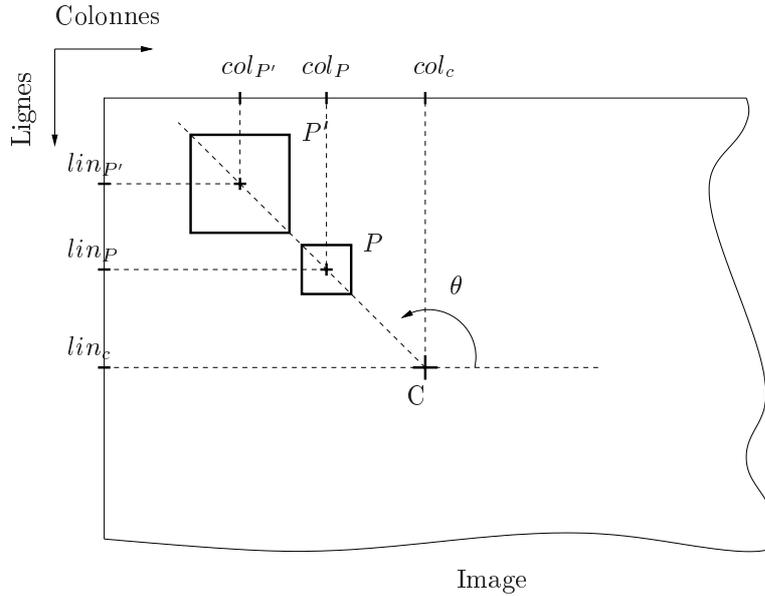


Figure 3.6 – Effet du zoom sur un pixel de l'image.

Si le centre de l'image n'est pas exactement sur l'axe focal de la caméra, le zoom se traduit par la combinaison : homothétie centrale plus translation. On peut donc supposer, sans perte de généralité, que le centre de l'image C est sur l'axe optique de la caméra. Ainsi, pour un point P transformé en P' par le zoom, le rapport d'homothétie est déterminé par :

$$\begin{aligned}\overline{CP} &= \frac{f \times T}{D + f} \\ \overline{CP'} &= \frac{f \times T}{D' + f}\end{aligned}\quad (3.2)$$

avec f la distance focale de la caméra. On en déduit :

$$\overline{CP'} = \overline{CP} \times \frac{D + f}{D' + f} = \alpha \times \overline{CP} \quad (3.3)$$

avec $\alpha \in]0, +\infty[$

En utilisant l'équation (3.3) et les notations définies sur la figure 3.4 on obtient :

$$\begin{aligned}\|CP\| \sin(\theta) &= \text{lin}_C - \text{lin}_P \\ \|CP\| \cos(\theta) &= \text{col}_C - \text{col}_P \\ \|CP'\| \sin(\theta) &= \text{lin}_C - \text{lin}_{P'} \\ \|CP'\| \cos(\theta) &= \text{col}_C - \text{col}_{P'}\end{aligned}$$

d'où :

$$\begin{aligned} \text{lin}_{P'} &= \text{lin}_C (1 - \alpha) + \alpha \text{lin}_P \\ \text{col}_{P'} &= \text{col}_C (1 - \alpha) + \alpha \text{col}_P \end{aligned}$$

soit en notation matricielle :

$$P' = (1 - \alpha) P_C + \alpha P \quad (3.4)$$

où P_C est la matrice correspondant au centre de l'image $P_C = \begin{pmatrix} \text{lin}_C \\ \text{col}_C \end{pmatrix}$. Il est à noter que les coordonnées du point P' ne sont pas nécessairement entières du fait du zoom.

Si l'on combine les deux équations (3.1) et (3.4), on obtient :

$$P_{\text{transformé}} = (1 - \alpha) P_C + \alpha P - T \quad (3.5)$$

Le dernier paramètre à prendre en compte dans le modèle de mouvement est la rotation autour de l'axe optique de la caméra. Encore une fois, si l'axe optique passe par le centre de l'image, la transformation due à la rotation de la caméra est illustrée sur la figure 3.7.

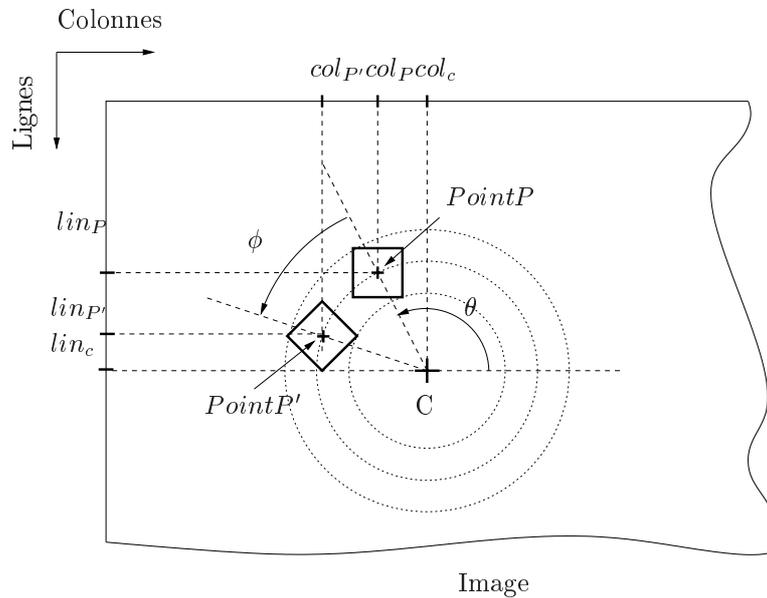


Figure 3.7 – Transformation d'un pixel suite à une rotation autour de l'axe optique de la caméra.

Pour une rotation d'un angle ϕ autour du centre de l'image, un point $P = \begin{pmatrix} \text{lin}_P \\ \text{col}_P \end{pmatrix}$ de l'image est transformé en $P' = \begin{pmatrix} \text{lin}_{P'} \\ \text{col}_{P'} \end{pmatrix}$ avec :

$$\begin{aligned} lin_{P'} &= lin_C (1 - \cos(\phi)) + (col_P - col_C) \sin(\phi) + lin_P \cos(\phi) \\ col_{P'} &= col_C (1 - \cos(\phi)) - (lin_P - lin_C) \sin(\phi) + col_P \cos(\phi) \end{aligned} \quad (3.6)$$

soit en notation matricielle :

$$P' = (\mathbb{I}_2 - R_\phi) P_C + R_\phi P \quad (3.7)$$

avec \mathbb{I}_2 la matrice identité d'ordre 2 et R_ϕ une matrice de rotation d'angle ϕ telle que :

$$R_\phi = \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix} \quad (3.8)$$

Si l'on combine les équations (3.5) et (3.7), on obtient la transformation d'un point par le modèle de mouvement à quatre paramètres :

$$P_{transformé} = (\mathbb{I}_2 - \alpha R_\phi) P_C + \alpha R_\phi P - \alpha R_\phi T \quad (3.9)$$

Dans le cadre de cette étude, le paramètre de zoom n'a pas été intégré dans la modélisation du mouvement apparent. Le terme α est donc pris égal à 1. La transformation d'un point de l'image par un mouvement $(T_{hor}, T_{ver}, R_\phi)$ s'exprime alors par :

$$P_{transformé} = (\mathbb{I}_2 - R_\phi) P_C + R_\phi P - R_\phi T \quad (3.10)$$

Nous verrons dans la conclusion, les raisons pour lesquelles nous n'avons pas, dans le cadre de ce travail, pris en compte le facteur de zoom dans le modèle de mouvement.

3.1.2 Modélisation floue du niveau de gris des pixels

Comme on l'a vu dans le début du paragraphe, un mouvement de la caméra se traduit par une variation de l'intensité des pixels. Cependant, des phénomènes, tels qu'une occultation partielle ou totale (objet se déplaçant devant la caméra), un changement de la luminosité globale (passage d'un objet devant la source d'éclairage) ou bien un changement des propriétés réfléchissantes d'une surface suite à un changement d'orientation, peuvent aussi induire des changements de niveaux de gris des pixels sans pour autant être reliés au déplacement de la caméra. Certaines techniques passent par l'identification de ces phénomènes (par exemple variation de l'illumination globale pour le flot optique). On cherche en général à minimiser l'impact de ces variations parasites sur l'estimation du mouvement. Plus précisément, on cherche à augmenter la robustesse de la méthode vis-à-vis des variations d'intensité des pixels qui ne seraient pas dues aux mouvements apparents. Nous supposons donc par la suite que la majorité des variations d'intensité des pixels est liée au mouvement apparent. Des essais, en fin de chapitre, illustreront la robustesse de notre méthode vis-à-vis de ces perturbations.

Chaque méthode d'estimation de mouvement nécessite des précautions d'emploi et des hypothèses pour être utilisée. La violation fréquente de ces hypothèses trop restrictives amène souvent à un échec du processus d'estimation de mouvement. Pour les méthodes basées sur l'appariement de points caractéristiques, on fait l'hypothèse que les attributs invariants associés à ces points (forte courbure du contour, surbrillance, etc.) sont conservés au long de la séquence vidéo. Ainsi, en cas d'occlusion de trop de points caractéristiques, la méthode a de fortes chances d'échouer. Il est alors indispensable de sélectionner un grand nombre de ces points pour s'assurer du succès de la méthode. Pour les techniques basées sur la mise en correspondance de zones (corrélacion) on considère que des motifs de l'image sont conservés au long de la séquence. Ici aussi la disparition partielle ou totale d'un motif due à une occlusion risque fort de perturber l'estimation de mouvement. De plus, ces méthodes sont souvent sensibles aux variations de niveau de gris des pixels du fait des métriques qu'elles utilisent lors de la mise en correspondance des motifs [JBC97]. Quant aux méthodes reposant sur le calcul du flot optique, elles émettent l'hypothèse que la distribution locale de l'illumination est conservée. Or, cette hypothèse est souvent transgressée dans le cas de déplacements trop grands. Bien sûr de nombreuses techniques ont été proposées pour s'affranchir partiellement de ces problèmes.

Nous proposons d'assouplir les hypothèses classiques d'appariement. Le principe de base de notre méthode d'estimation de mouvement repose sur une hypothèse très simple : dans une image à niveaux de gris, un pixel qui est blanc a peu de chance de devenir noir au cours du mouvement et vice-versa.

Pour simplifier le principe de représentation des pixels, nous avons défini deux classes duales : la classe des pixels blancs et la classe des pixels noirs [CSA01]. Un des inconvénients majeurs de cette approche est l'instabilité à la frontière entre les deux classes. Cet inconvénient peut être facilement contourné en utilisant une représentation floue de l'appartenance d'un pixel à une classe. La figure 3.8 illustre la division de l'espace des niveaux de gris en deux classes floues duales.

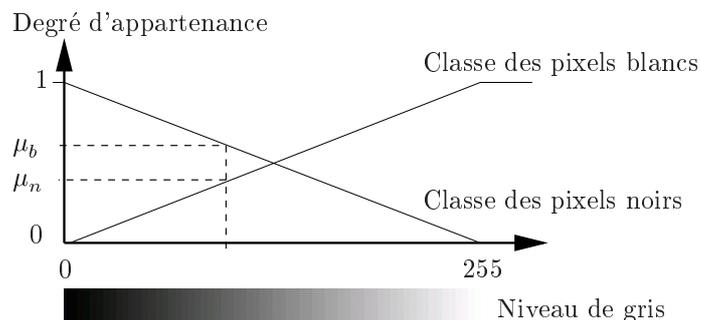


Figure 3.8 – Deux classes duales pour représenter le niveau de gris des pixels : pixels blancs et pixels noirs.

Chaque pixel d'une image est représenté par son degré d'appartenance à la classe des pixels blancs (μ_b) et son degré d'appartenance à la classe des pixels noirs (μ_n).

Nous verrons dans le paragraphe suivant que cette modélisation possède des avantages très intéressants. Entre autres, on est en droit d'attendre une certaine robustesse vis-à-vis des variations d'intensité globale du fait de l'utilisation d'une classification du niveau de gris des pixels. Les performances de cette modélisation sont illustrées dans le paragraphe 3.5. Les problèmes liés aux occlusions et autres phénomènes locaux, sont eux aussi évoqués dans paragraphe 3.5.

3.1.3 Représentation multi-échelles des images

Les représentations multi-échelles des images sont souvent employées dans les processus en vision assistée par ordinateur. Elles consistent à représenter les images à différents niveaux de précision. Soit une image de taille M_0 lignes par N_0 colonnes. Sa représentation au niveau 0 ou résolution initiale correspond à l'image elle-même. Elle est donc composée de $M_0 \times N_0$ pixels. Plus le niveau augmente, plus la précision de la représentation diminue. Aussi, cette même image à la résolution 1 aura pour taille $M_1 \times N_1$ pixels avec $M_1 < M_0$ et $N_1 < N_0$. Pour les méthodes multi-échelles classiques, on divise généralement une image en blocs de 2×2 pixels. On essaye ensuite de chercher de manière certaine quel serait le niveau de gris d'un pixel s'il devait représenter un bloc de quatre pixels. C'est à dire quel serait le niveau de gris des pixels de l'image si la caméra possédait une résolution deux fois moins importante ou si la focale était deux fois plus petite. Dans ces deux cas, la scène visualisée par une image de $2M \times 2N$ pixels est représentée par une image de $M \times N$ pixels. Plusieurs méthodes ont été employées. La plus classique consiste à minimiser l'erreur entre le niveau de gris du pixel à la résolution R_{n+1} (\widehat{Ng}) et celui des quatre pixels lui correspondant à la résolution R_n (Ng_i). Cette erreur est définie par :

$$Erreur = \sum_{i \in \text{bloc de } 2 \times 2 \text{ pixels}} \left(\widehat{Ng} - Ng_i \right)^2 \quad (3.11)$$

Cette approche conduit naturellement à prendre pour \widehat{Ng} la valeur moyenne de l'intensité des quatre pixels Ng_i . Ainsi on a :

$$Ng = \frac{1}{4} \sum_{i \in \text{bloc de } 2 \times 2 \text{ pixels}} Ng_i \quad (3.12)$$

Il existe d'autres méthodes ne considérant pas seulement des blocs de 2×2 pixels mais prenant en compte plus de pixels (filtrage gaussien) ou bien ne minimisant pas la même erreur, cependant elles ne sont pas présentées dans ce manuscrit. La méthode multi-échelles que nous avons défini ne répond pas aux mêmes critères. En effet, elle vise à conserver le maximum d'information sur les niveaux de gris des pixels et découle naturellement de la modélisation du niveau de gris à l'aide des deux classes floues duales. Le principe de notre approche multi-échelles est le suivant :

PROPOSITION 1

Soit $(P(2m, 2n); P(2m + 1, 2n); P(2m, 2n + 1); P(2m + 1, 2n + 1))$ avec $m \in [0, M]$, $n \in [0, N]$ un bloc de quatre pixels de l'image de taille $(2M, 2N)$ à la résolution R_n . A la résolution R_{n+1} ces quatre pixels sont remplacés par un seul pixel dont l'appartenance à la classe des pixels noirs est définie par :

$$\mu_n = \max \left(\begin{array}{l} \mu_n(P(2m, 2n)), \mu_n(P(2m + 1, 2n)), \\ \mu_n(P(2m, 2n + 1)), \mu_n(P(2m + 1, 2n + 1)) \end{array} \right)$$

et l'appartenance à la classe des pixels blancs est définie par :

$$\mu_b = \max \left(\begin{array}{l} \mu_b(P(2m, 2n)), \mu_b(P(2m + 1, 2n)), \\ \mu_b(P(2m, 2n + 1)), \mu_b(P(2m + 1, 2n + 1)) \end{array} \right).$$

Ainsi, on est sûr que la vraie valeur du niveau de gris \widehat{Ng} du pixel à la résolution R_{n+1} appartient à l'intervalle défini par $[255(1 - \mu_n), 255\mu_b]$. Pour illustrer ce procédé multi-échelles nous avons représenté sur la figure 3.9 la décomposition d'une image à différentes résolutions. Les deux images correspondent en fait aux valeurs prises par μ_n et μ_b , le blanc correspondant à $\mu = 1$ et le noir à $\mu = 0$. Ainsi, si l'image est sombre, l'image représentant les μ_b sera elle aussi sombre. En effet peu de pixels appartiennent à la classe "pixel blanc" ($\mu_b \simeq 0$). Par contre, l'image représentant les μ_n sera très claire car un grand nombre de pixels appartiennent à la classe "pixel noir". Ceci est illustré sur la figure 3.9, où nous avons représenté les μ_b et μ_n de l'image des arbres.

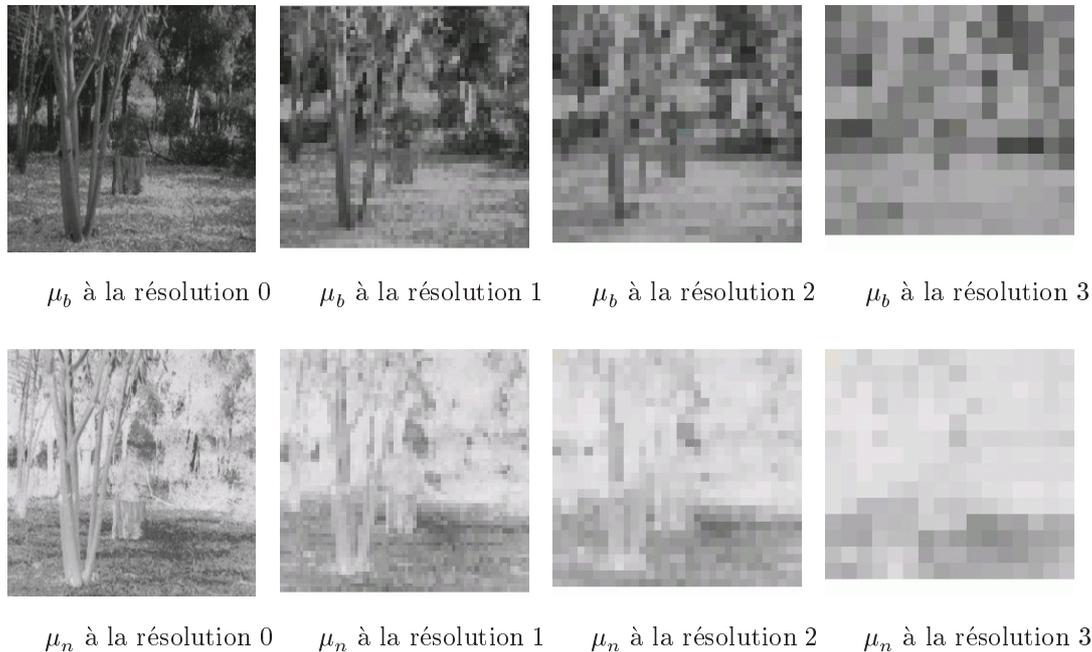


Figure 3.9 – Décomposition multi-échelle des images. Les images du haut représentent les valeurs de μ_b et celles du bas les valeurs de μ_n .

Comme on peut le voir sur la figure 3.9, plus la résolution de l'image se dégrade plus la somme des deux valeurs d'appartenance aux classes de niveaux de gris s'éloignent de un (au niveau 0 on a $\mu_b + \mu_n = 1$). Ceci traduit le fait que la précision avec laquelle

on connaît le niveau de gris d'un pixel décroît à mesure que le niveau de précision de l'image diminue. En effet, un manque d'informations sûres dû à une précision de représentation basse se traduit par un intervalle de définition large pour le niveau de gris \widehat{Ng} .

Les performances de cette représentation multi-échelles imprécise sont présentées dans le chapitre traitant des applications.

3.2 Evaluation de l'éventualité de déplacement d'un pixel en fonction de son niveau de gris

Soit un pixel P_1 situé à la $i^{\text{ème}}$ ligne et à la $j^{\text{ème}}$ colonne de l'image 1. Sa position est notée (i, j) . Considérons un déplacement de la caméra. Le pixel P_1 de l'image 1 peut-il se retrouver en P'_1 à la position (i', j') dans l'image 2? C'est à cette question que nous allons nous efforcer de répondre. Dans un premier temps, nous allons uniquement considérer l'éventualité de ce déplacement en fonction des niveaux de gris des pixels de l'image. Le déplacement d'un pixel est illustré sur la figure 3.10.

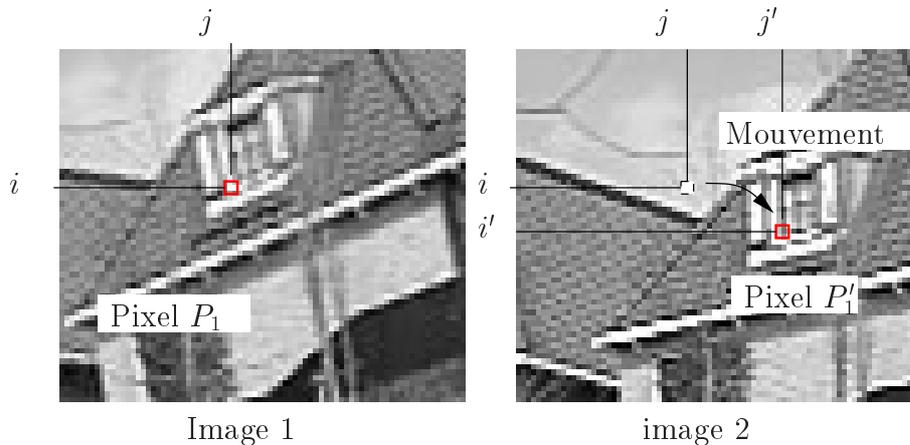


Figure 3.10 – Déplacement d'un pixel correspondant au mouvement engendré par (Tx, Ty, Rz) .

Le déplacement dans le plan image correspond à (u, v) avec :

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} i' \\ j' \end{pmatrix} - \begin{pmatrix} i \\ j \end{pmatrix} \quad (3.13)$$

Maintenant, comment peut on évaluer si un déplacement (u, v) est envisageable pour un pixel (i, j) ? Nous allons pour cela détailler le principe énoncé dans le paragraphe 3.1.2 à l'aide de la proposition suivante :

PROPOSITION 1 *Un pixel P_1 , de l'image 1, caractérisé par sa position (i, j) , a pu se déplacer en P'_1 de l'image 2 caractérisé par (i', j') si et seulement si P_1 et P'_1 appar-*

tiennent à la même classe de pixels. C'est-à-dire P_1 et P'_1 appartiennent à la classe des pixels blancs, ou P_1 et P'_1 appartiennent à la classe des pixels noirs.

Du fait de la modélisation floue des pixels, un pixel P_1 est caractérisé par ses deux degrés d'appartenance aux classes de pixels blancs et de pixels noirs respectivement notés $\mu_b(P_1)$ et $\mu_n(P_1)$. L'éventualité pour qu'un pixel P_1 se soit déplacé d'une quantité (u, v) en fonction du niveau de gris des pixels est alors évaluée de manière duale. Nous avons défini un processus de vote bi-modal, caractérisé par les bornes inférieure et supérieure de l'éventualité de déplacement des deux pixels. La borne supérieure est donnée par une fonction à valeur dans $[0, 1]$ appelée vote pour le déplacement (u, v) , définie par :

$$VotePour(P_1; P'_1) = \max[\min(\mu_b(P_1), \mu_b(P'_1)), \min(\mu_n(P_1), \mu_n(P'_1))] \quad (3.14)$$

Ce vote traduit le fait que les deux pixels appartiennent à la même classe de pixels. De la même façon, on peut établir une fonction duale à valeur dans $[0, 1]$ appelée vote contre le déplacement (u, v) et définie par :

$$VoteContre(P_1; P'_1) = \max[\min(\mu_b(P_1), \mu_n(P'_1)), \min(\mu_n(P_1), \mu_b(P'_1))] \quad (3.15)$$

Ce vote traduit le fait que les deux pixels n'appartiennent pas à la même classe. La borne inférieure de l'éventualité de déplacement est ensuite donnée par :

$$1 - VoteContre(P_1; P'_1) \quad (3.16)$$

Il existe une relation très forte entre les deux votes. En effet, on a généralement :

$$VotePour(P_1; P'_1) = 1 - VoteContre(P_1; P'_1) \quad (3.17)$$

REMARQUE 7 *Cette propriété n'est vraie qu'à la résolution 0. Elle devient fausse dès que l'on descend dans les niveaux de précision de représentation de l'image.*

Cependant, il existe une propriété très intéressante résultant de la modélisation floue du niveau de gris des pixels en conjonction avec le procédé de vote bi-modal. Cette propriété permet de s'affranchir des problèmes d'effets de bord dans le processus de mise en correspondance. Lorsque l'on cherche si un déplacement (u, v) a pu amener un pixel P_1 en dehors des limites de l'image 2, comme cela est illustré sur la figure 3.11, quel niveau de gris doit on prendre pour le pixel P'_1 ? Et donc, comment va-t-on pouvoir mesurer la compatibilité du niveau de gris de ces deux pixels?

Dans ce cas, puisque la valeur du niveau de gris du pixel est inconnue, elle peut tout aussi bien être blanche que noire, il vient alors naturellement :

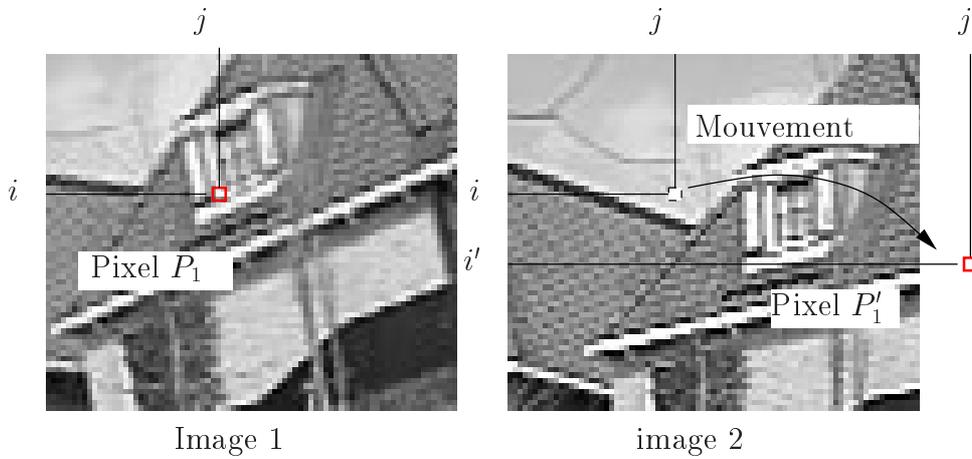


Figure 3.11 – Déplacement d'un pixel P_1 en dehors du cadre de l'image 2.

$$\mu_b(P'_1) = \mu_n(P'_1) = 1 \quad (3.18)$$

Ainsi les équations (3.14) et (3.15) s'écrivent :

$$VotePour(P_1; P'_1) = VoteContre(P_1; P'_1) = \max(\mu_b(P_1), \mu_n(P_1)) \quad (3.19)$$

L'absence d'information sur le niveau de gris du pixel P'_1 se traduit par un vote identique pour et contre le mouvement (u, v) . Ainsi, aucun des deux votes n'est privilégié par des considérations arbitraires sur le niveau de gris de ce pixel. Le problème lié aux bords de l'image est ainsi pris en compte implicitement par la technique de vote bi-modale que nous avons définie.

3.3 Estimation de l'éventualité de déplacement d'un pixel en fonction d'une transformation imprécise

Dans le cadre de notre étude, le modèle du mouvement apparent considéré est limité à trois paramètres $(T_{horizontale}, T_{verticale}, R_\phi)$. Pour estimer ce mouvement, nous avons discrétisé l'espace paramétrique. Ainsi nous avons formé un maillage en trois dimensions dont les noeuds sont les combinaisons des paramètres $T_{horizontale}$, $T_{verticale}$, et R_ϕ . La méthode que nous avons mise au point consiste à évaluer pour chaque pixel de l'image 1 la pertinence de l'ensemble des déplacements représentés par le domaine de l'espace paramétrique couvert par un pavé de la discrétisation.

Cette démarche s'apparente au calcul d'erreur classique. Toute discrétisation est considérée comme un pavage, et toutes les opérations sont vues sous un angle ensembliste. Dans un souci de minimiser l'influence d'une représentation granulaire, nous

adoptons une modélisation floue pour tout ensemble. Ainsi, la discrétisation spatiale de l'image est vue comme un pavage flou de \mathbb{R}^2 .

Pour que cette modélisation soit la plus neutre possible, nous avons choisi de représenter un pixel flou par un pavé mono-modal de support rectangulaire. Ainsi, sa représentation graphique est pyramidale comme représenté sur la figure 3.12.

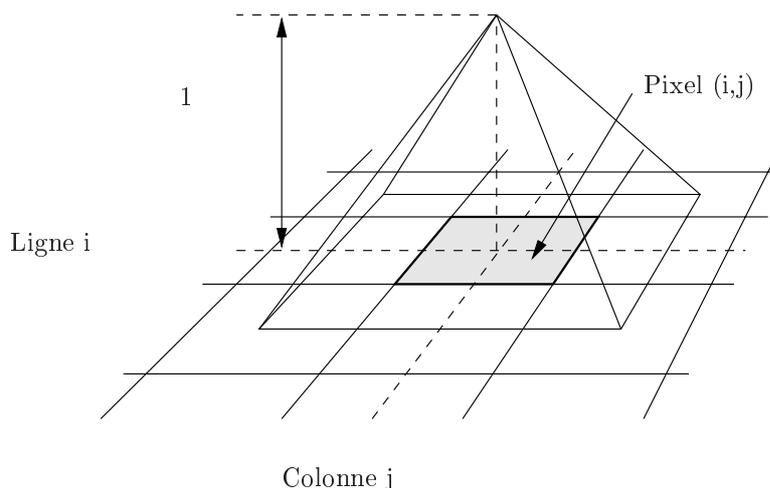


Figure 3.12 – Quantité floue représentant un pixel.

Dans les paragraphes suivants, nous proposons deux approches pour l'estimation de la pertinence du mouvement d'un pixel entre l'image 1 et l'image 2. Dans la première approche, nous abandonnons la représentation floue pour l'espace paramétrique en supposant (ce qui est classique) que la valeur des paramètres pour le mode du pavé est représentative de l'ensemble du pavé. Cette démarche est équivalente à supposer que le mouvement est précis. Dans une seconde approche, nous évaluons la pertinence de l'ensemble de mouvements compris à l'intérieur d'un pavé de la discrétisation de l'espace paramétrique. Cette approche revient à considérer que le mouvement est imprécis [Blo01]. On peut voir la première approche comme une simplification par rapport à la seconde. Son intérêt est avant tout d'être pédagogique pour la présentation du manuscrit. Elle permet aussi de réduire considérablement le temps d'exécution de l'algorithme en simplifiant les calculs.

3.3.1 Modèle de mouvement précis

Dans ce paragraphe, on considère que les paramètres du modèle de déplacement sont connus précisément. Pour illustrer notre propos, nous avons simulé le déplacement d'une caméra devant un paysage. Le mouvement est composé d'une translation vers la droite puis vers le haut et finalement d'une rotation dans le sens trigonométrique. Sur l'image, les points donnent l'impression de se déplacer dans le sens opposé. Ceci est illustré sur la figure 3.13.

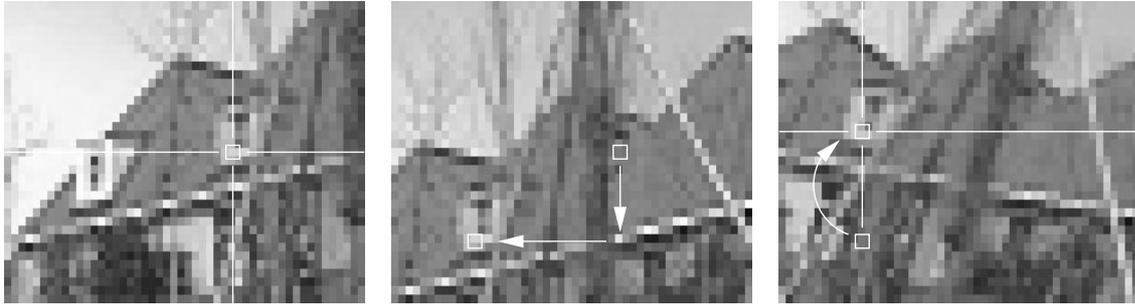


Figure 3.13 – Déplacement d'un point de l'environnement dans une séquence vidéo.

Puisque les pixels se "déplacent" dans l'image, les quantités floues les représentant subissent elles aussi les transformations dues au mouvement. Si les paramètres du modèle de déplacement (T_{hor} , T_{ver} , R) sont précis, alors le détail de ces transformations peut être représenté comme sur la figure 3.14.

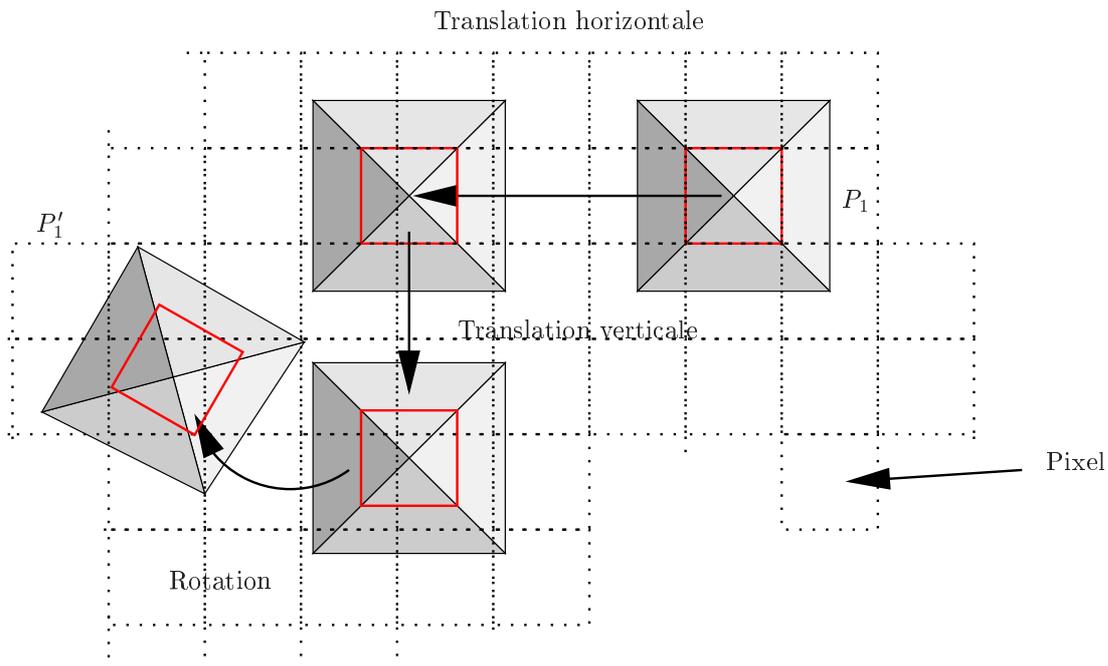


Figure 3.14 – Détail du déplacement d'un pixel dû au mouvement de la caméra.

Ce mouvement transforme la quantité floue représentant le pixel P_1 de l'image 1 (notée \mathcal{P}_1) en une quantité floue associée à P'_1 (notée \mathcal{P}'_1). Celle-ci peut être amenée à intersecter plusieurs quantités floues représentant des pixels P_2 de l'image 2 (notées \mathcal{P}_2). Dans ce cas, quels pixels de l'image 2 doivent être pris en compte ? Doivent-ils tous être considérés, ou seulement celui ayant la plus grande intersection avec la quantité floue associée à \mathcal{P}'_1 , ou bien encore doit on considérer seulement le plus proche ?

Nous devons donc déterminer si un pixel P_1 de l'image 1 a pu se déplacer en P_2 de l'image 2. Puisque les pixels sont représentés par des quantités floues, il existe un ensemble continu de valeurs représentant la compatibilité entre la quantité floue associée à P_1' et celles associées aux pixels de l'image 2. Nous proposons de représenter cet ensemble de compatibilités par ses bornes inférieure et supérieure. Ces deux bornes sont données par la nécessité et la possibilité de \mathcal{P}'_1 restreinte à \mathcal{P}_2 et s'écrivent respectivement :

$$N(\mathcal{P}'_1; \mathcal{P}_2) = \inf_{P_2 \in I_2} \left(\max \left(\mu_{\mathcal{P}'_1}, 1 - \mu_{\mathcal{P}_2} \right) \right) \quad (3.20)$$

$$\Pi(\mathcal{P}'_1; \mathcal{P}_2) = \sup_{P_2 \in I_2} \left(\min \left(\mu_{\mathcal{P}'_1}, \mu_{\mathcal{P}_2} \right) \right) \quad (3.21)$$

où I_2 est l'image 2 et μ_P est la fonction caractéristique linéaire de la quantité floue associée au pixel P . $N(\mathcal{P}'_1; \mathcal{P}_2)$ "mesure" à quel point la cellule associée à P_2 est incluse dans celle associée à P_1' , et $\Pi(\mathcal{P}'_1; \mathcal{P}_2)$ "mesure" à quel point les deux cellules s'intersectent.

Estimer ces deux quantités fait intervenir un calcul d'intersection de pyramides (donc plusieurs calculs d'intersection de plans). Par exemple, calculer la possibilité $\Pi(\mathcal{P}'_1; \mathcal{P}_2)$ revient à déterminer le point le plus élevé de l'intersection entre les deux fonctions caractéristiques pyramidales représentant les quantités floues \mathcal{P}'_1 et \mathcal{P}_2 . Ces opérations sont relativement coûteuses en temps de calcul, et dans un souci de gain de temps, nous avons opté pour une méthode légèrement dégradée.

La complexité du calcul vient du fait que l'orientation de la fonction caractéristique de P_1' est quelconque. Le calcul de $N(\mathcal{P}'_1; \mathcal{P}_2)$ et $\Pi(\mathcal{P}'_1; \mathcal{P}_2)$ se retrouve grandement simplifié si les deux pyramides ont la même orientation. Pour rendre ce calcul plus rapide, nous utilisons donc pour représenter \mathcal{P}'_1 , une fonction caractéristique englobante, ayant la même orientation que celles associées aux pixels P_2 de l'image 2, mais de précision moindre. Ceci est illustré sur les deux figures 3.16 et 3.17.

Cette technique revient à supposer que les variables de représentation de la position des pixels sont non-interactives. Ceci est classique en calcul d'erreur et correspond toujours au "pire cas". Dans ce cas, le calcul de $N(\mathcal{P}'_1; \mathcal{P}_2)$ et $\Pi(\mathcal{P}'_1; \mathcal{P}_2)$ peut s'effectuer sur les projections horizontales et verticales des deux quantités floues comme indiqué sur la figure 3.17 : où $N_{hor}(\mathcal{P}'_1; \mathcal{P}_2)$ (resp. $\Pi_{hor}(\mathcal{P}'_1; \mathcal{P}_2)$) est la nécessité (resp. possibilité) de \mathcal{P}'_1 restreinte à \mathcal{P}_2 en projection horizontale, et $N_{ver}(\mathcal{P}'_1; \mathcal{P}_2)$ (resp. $\Pi_{ver}(\mathcal{P}'_1; \mathcal{P}_2)$) est la nécessité (resp. possibilité) de \mathcal{P}'_1 restreinte à \mathcal{P}_2 en projection verticale. Les valeurs de $N(\mathcal{P}'_1; \mathcal{P}_2)$ et $\Pi(\mathcal{P}'_1; \mathcal{P}_2)$ sont données par l'utilisation de la t-norme min [DP88] :

$$N(\mathcal{P}'_1; \mathcal{P}_2) = \min(N_{hor}(\mathcal{P}'_1; \mathcal{P}_2), N_{ver}(\mathcal{P}'_1; \mathcal{P}_2)) \quad (3.22)$$

$$\Pi(\mathcal{P}'_1; \mathcal{P}_2) = \min(\Pi_{hor}(\mathcal{P}'_1; \mathcal{P}_2), \Pi_{ver}(\mathcal{P}'_1; \mathcal{P}_2)) \quad (3.23)$$

Ces deux bornes fixent un encadrement de la mesure de confiance pour que le pixel P_1 se soit déplacé en P_2 suite au mouvement paramétré par (T_{hor}, T_{ver}, R) .

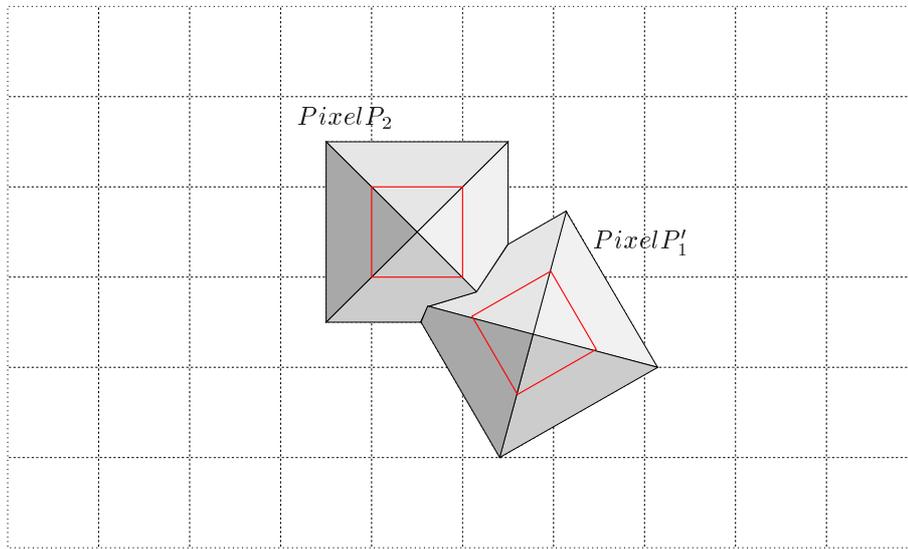


Figure 3.15 – Intersection entre les deux quantités floues associées à P_1' et P_2 .

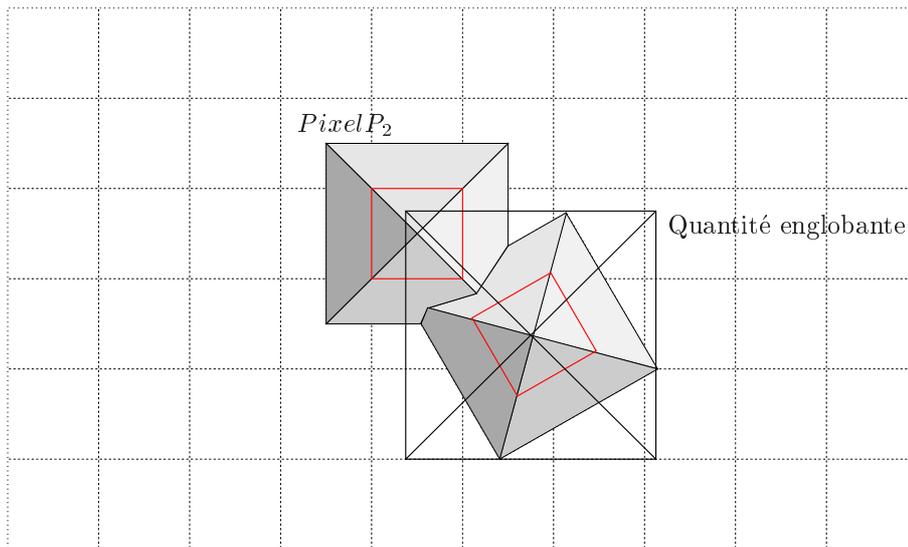


Figure 3.16 – Intersection entre la quantité floue englobante associée à P_1' et celle associée à P_2 .

3.3.2 Modèle de mouvement imprécis

Le modèle de mouvement précis n'est qu'une simplification de l'approche générale d'estimation de l'éventualité de déplacement d'un pixel. Nous présentons ici cette approche qui prend en compte la discrétisation de l'espace paramétrique.

D'un point de vue statistique (ce que nous avons vu au paragraphe précédent), les noeuds du maillage de l'espace paramétrique sont représentatifs de tout leur voisinage

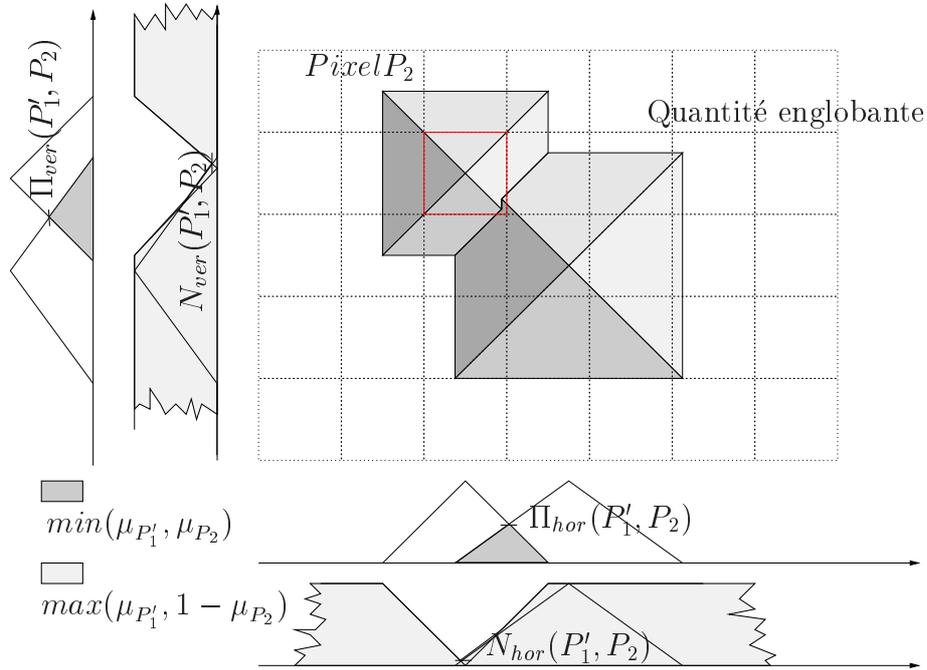


Figure 3.17 – Calcul des possibilité et nécessité de P_2 restreinte à P'_1 en projection horizontale et verticale.

(cf. figure 3.18).

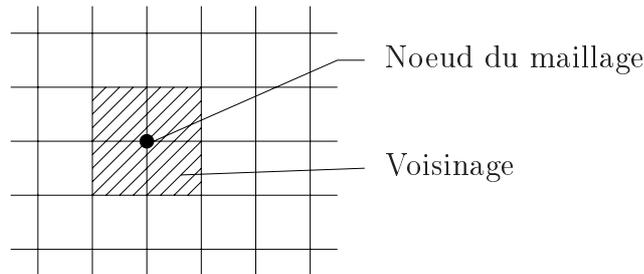


Figure 3.18 – Noeud du maillage et son voisinage.

D'un point de vue ensembliste, ce voisinage doit être pris en considération comme un intervalle, ou pavé, de valeurs possibles. En effet, la discrétisation de l'espace paramétrique introduit une imprécision sur les paramètres du mouvement. On est alors amené à adopter des techniques propres aux calculs ensemblistes [Wal91] pour le calcul de l'ensemble des valeurs possibles de la localisation du pixel P' sachant que T_{hor} , T_{ver} , R_ϕ et P sont des sous-ensembles flous définis par leurs modes : T_{hor0} , T_{ver0} , R_0 , et leurs supports : $T_{hor} \in [T_{hor0} - \Delta_{T_{hor}}, T_{hor0} + \Delta_{T_{hor}}]$, $T_{ver} \in [T_{ver0} - \Delta_{T_{ver}}, T_{ver0} + \Delta_{T_{ver}}]$, $R_\phi \in [R_0 - \Delta_R, R_0 + \Delta_R]$ et que P appartient au support de la quantité floue qui lui est associée. L'équation (3.10) doit alors être considérée comme une équation ensembliste.

La figure 3.19 illustre l'influence de l'imprécision des paramètres du modèle de déplacement sur l'ensemble des positions possibles de \mathcal{P}'_1 .

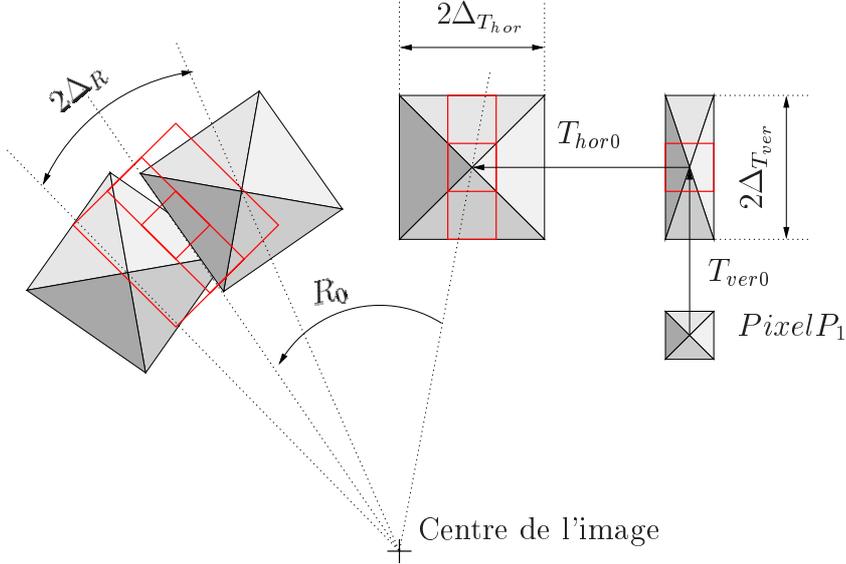


Figure 3.19 – Effet de l'imprécision des paramètres du modèle de mouvement.

Pour calculer \mathcal{P}'_1 , image de \mathcal{P}_1 par la transformation imprécise $(T_{hor}, T_{ver}, R_\phi)$, de nombreuses techniques peuvent être utilisées généralisant les méthodes développées dans [Wal91] et utilisant le principe d'extension [DP88]. Cependant, ces techniques sont très coûteuses en temps de calcul. Nous avons donc préféré utiliser un formalisme dérivé du calcul d'erreur classique [Moo79], utilisant le fait que les sous-ensembles flous que nous manipulons sont des “nombres flous non-interactifs” sur lesquels est appliqué un calcul de quantités floues “pessimiste”.

Dans ce cas calculer le noyau et le support de l'ensemble \mathcal{P}'_1 revient à calculer l'image du mode de \mathcal{P}_1 par le mode de $(T_{hor}, T_{ver}, R_\phi)$ noté $P'_{1Centre} = \begin{bmatrix} lin'_{1Centre} \\ col'_{1Centre} \end{bmatrix}$, et d'appliquer un calcul variationnel sur les supports [DP88]. Ainsi si $P'_1 = \begin{bmatrix} lin'_1 \\ col'_1 \end{bmatrix}$ est une fonction des ensembles flous $\mathcal{P}_1 = \begin{bmatrix} lin_1 \\ col_1 \end{bmatrix}$, T_{hor} , T_{ver} , R_ϕ et que l'on note : $lin'_1 = f_1(\mathcal{P}_1, T_{hor}, T_{ver}, R_\phi)$ et $col'_1 = f_2(\mathcal{P}_1, T_{hor}, T_{ver}, R_\phi)$, avec f_1 et f_2 définis par l'équation 3.6, alors le support de \mathcal{P}'_1 est de largeur $2\Delta_{col'}$ et de hauteur $2\Delta_{lin'}$ avec :

$$\Delta_{lin'_1} = \left| \frac{\partial f_1}{\partial lin_1} \right| \Delta_{lin_1} + \left| \frac{\partial f_1}{\partial col_1} \right| \Delta_{col_1} + \left| \frac{\partial f_1}{\partial R_\phi} \right| \Delta_{R_\phi} + \left| \frac{\partial f_1}{\partial T_{hor}} \right| \Delta_{T_{hor}} + \left| \frac{\partial f_1}{\partial T_{ver}} \right| \Delta_{T_{ver}}$$

$$\Delta col'_1 = \left| \frac{\partial f_2}{\partial lin_1} \right| \Delta lin_1 + \left| \frac{\partial f_2}{\partial col_1} \right| \Delta col_1 + \left| \frac{\partial f_2}{\partial R_\phi} \right| \Delta R_\phi + \left| \frac{\partial f_2}{\partial T_{hor}} \right| \Delta T_{hor} + \left| \frac{\partial f_2}{\partial T_{ver}} \right| \Delta T_{ver}$$

Soit en développant en fonction des expressions de f_1 et f_2 définies par l'équation (3.10) :

$$\Delta lin' = |\cos(\phi)| \Delta lin + |\sin(\phi)| \Delta col + \Delta T_{hor} + \left| (\sin(\phi)(lin_c - lin) + \cos(\phi)(col - col_c)) \right| \Delta \phi \quad (3.24)$$

$$\Delta col' = |\sin(\phi)| \Delta lin + |\cos(\phi)| \Delta col + \Delta T_{ver} + \left| (\sin(\phi)(col_c - col) + \cos(\phi)(lin_c - lin)) \right| \Delta \phi \quad (3.25)$$

La figure 3.20 illustre le calcul de l'ensemble fou englobant associé à P'_1 .

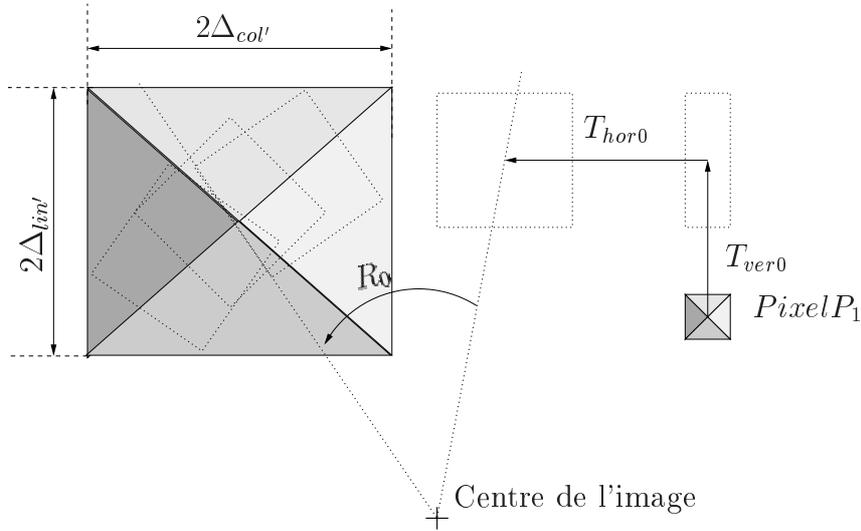


Figure 3.20 – Quantité floue englobante issue du calcul d'erreur.

Nous avons vu quelle était l'influence de l'imprécision des paramètres du modèle sur le déplacement d'un pixel. Il reste à estimer si un pixel P_1 de l'image 1 a pu se déplacer en P_2 de l'image 2. Pour cela, nous utilisons le même formalisme que dans le cas du modèle précis de déplacement. La compatibilité entre la quantité floue associée à P'_1 et celle associée à un pixel P_2 de l'image 2 est comprise dans un intervalle. Nous déterminons une borne supérieure et inférieure de cet intervalle à l'aide des formules de possibilité et nécessité restreintes projetées définies au paragraphe précédent (formules (3.22), (3.23)).

3.4 Estimation de la probabilité d'un mouvement

Les méthodes exposées dans les paragraphes 3.2 et 3.3 nous permettent maintenant d'estimer, pour tout pixel de l'image 1, son éventualité de déplacement sur l'image 2 pour un ensemble de mouvements $(T_{hor}, T_{ver}, R_\phi)$. Nous proposons d'utiliser le formalisme des histogrammes approchés présentés dans le chapitre 2 pour déduire de cette masse d'informations le déplacement majoritaire le plus probable. Il s'agit en fait du déplacement (ou de l'ensemble des déplacements) le plus fréquent en considérant l'éventualité de déplacement de tous les pixels de l'image 1 sur l'image 2. L'utilisation des histogrammes approchés doit permettre au final d'évaluer la probabilité d'occurrence de tout déplacement $(T_{hor}, T_{ver}, R_\phi)$.

La procédure d'accumulation (voir chapitre 2) peut être vue comme une transformée de Hough dite de m à m [CM91] où pour chaque ensemble de déplacement évalué défini par le pavé $(T_{hor}, T_{ver}, R_\phi)_i$ on "compte" le nombre de pixels acceptant cette éventualité de déplacement. Pour chaque $(T_{hor}, T_{ver}, R_\phi)_i$, et pour chaque pixel de l'image 1 on accumule alors cette valeur imprécise sur l'histogramme imprécis.

L'aspect pratique de la technique est décrit à présent. A cause de l'imprécision des différents éléments entrant en jeu dans le processus de vote, les éventualités de déplacement en fonction des transformations ou des niveaux de gris, sont représentées par des intervalles (borne inférieure et supérieure). Pour conserver l'homogénéité de notre méthode dans le traitement de l'imprécision des données, nous avons défini un processus de vote bi-modal. Celui-ci est basé sur le principe suivant :

PROPOSITION 2

Un pixel P_1 de l'image 1 vote possiblement pour le déplacement $(T_{hor}, T_{ver}, R_\phi)$ si, $\exists \mathcal{P}_2$ vérifiant $\mathcal{P}_2 \cap \mathcal{P}'_1 \neq \emptyset$ et tel que les niveaux de gris de P_1 et P_2 appartiennent à la même classe.

PROPOSITION 3

Un pixel P_1 de l'image 1 vote nécessairement pour le déplacement $(T_{hor}, T_{ver}, R_\phi)$ si, $\forall \mathcal{P}_2 \subset \mathcal{P}'_1$, les niveaux de gris de P_1 et P_2 appartiennent à la même classe.

Ces propositions expriment en fait une fusion des équations (3.14), (3.15), (3.22) et (3.23). Nous avons choisi de fusionner ces propositions avec une loi de composition de type sup-min [DP88].

Ainsi la borne supérieure du vote pour un déplacement $(T_{hor}, T_{ver}, R_\phi)$ est donnée par :

$$Vote_{SupP_1}(T_{hor}, T_{ver}, R_\phi) = Sup_{P_2 \in \text{Image 2}} (\min(\Pi(\mathcal{P}'_1; \mathcal{P}_2), VotePour(P_1, P_2))) \quad (3.26)$$

La deuxième proposition se traduit par : " P_1 vote nécessairement pour le déplacement (T_{hor}, T_{ver}, R) s'il n'existe aucun pixel P_2 vérifiant $\mathcal{P}_2 \subset \mathcal{P}'_1$ tel que les niveaux de gris de P_1 et P_2 appartiennent à des classes différentes".

Pour la borne inférieure du vote, cela s'exprime par la négation de " $\mathcal{P}_2 \subset \mathcal{P}_1$ " et "les niveaux de gris de P_1 et P_2 appartiennent à des classes différentes", soit :

$$Vote_{InfP_1}(T_{hor}, T_{ver}, R_\phi) = 1 - Sup_{P_2 \in \text{Image}_2}(\min(N(\mathcal{P}'_1; \mathcal{P}_2), VoteContre(P_1, P_2))) \quad (3.27)$$

Ces votes sont ensuite accumulés dans un histogramme approché construit sur une partition de l'espace paramétrique du mouvement. La borne supérieure du vote est accumulée dans l'accumulateur supérieur, et la borne inférieure du vote dans l'accumulateur inférieur. Cette procédure est répétée pour chaque pixel de l'image 1. Nous définissons ainsi un encadrement de la probabilité qu'un mouvement $(T_{hor}, T_{ver}, R_\phi)$ ait eu lieu. Il reste maintenant à estimer lequel de tous ces mouvements est le mouvement majoritaire.

3.5 Estimation du mouvement dominant

3.5.1 Introduction

Nous avons annoncé au début du chapitre que le mouvement apparent dominant correspondait au mouvement associé à la majorité des pixels. Dans le processus de vote, chaque pixel de l'image 1 vote pour l'ensemble des déplacements. La détermination du mouvement apparent majoritaire s'effectue donc en recherchant le mode de la distribution des mouvements (caractérisée par l'histogramme approché construit sur l'espace paramétrique du mouvement). La procédure complète de détermination de ce mode a été détaillée dans le chapitre 2.

Dans ce chapitre nous proposons d'illustrer les performances de l'algorithme que nous avons développé sur un certain nombre d'exemples de mouvements plans simulés. Ces exemples simulés nous permettent de comparer notre méthode avec des méthodes classiques telles que les méthodes différentielles ou les méthodes apparentées à la corrélation.

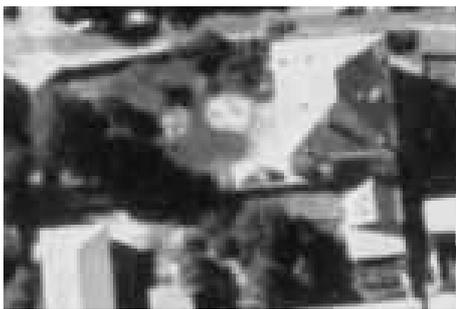
Les séquences d'images que nous avons générées sont issues d'une simulation de déplacement en translation et rotation d'une fenêtre sur une image en niveau de gris de grande dimension (1125×749 pixels). Cette procédure est illustrée sur la figure 3.21. L'image source ainsi que deux sous images y sont représentées. Dans le cas présenté, le déplacement est constitué d'une translation horizontale de 1.1 pixels, d'une translation verticale de 2.5 pixels et d'une rotation de 3° .

REMARQUE 8 Cette simulation correspond au cas où la scène filmée est loin de l'objectif de la caméra et comporte de très faibles variations de profondeur de champ, ce qui est souvent le cas dans les images aériennes ou satellites.

Dans un premier temps, nous présentons des exemples où seul le déplacement plan simulé est présent sur les images. Dans un second temps, nous testons la robustesse de



(a)



(b)



(c)

Figure 3.21 – Image source (a) (comportant un agrandissement de la zone utilisée), et les deux imagerie b (cadre foncé) et c (cadre clair) extraites de celle-ci.

l'algorithme en provoquant des variations parasites de l'illumination globale de l'image et en rajoutant, par animation, des mouvements parasites sur la séquence d'images.

Tous les tests proposés dans ce chapitre sont issus de séquences simulées pour

T_{hor}		
Erreur moyenne (en pixels)	Erreur maximum (en pixels)	Ecart type de l'erreur
0.11	0.45	0.09
T_{ver}		
Erreur moyenne (en pixels)	Erreur maximum (en pixels)	Ecart type de l'erreur
0.07	0.51	0.08
$Rotation$		
Erreur moyenne (en degrés)	Erreur maximum (en degrés)	Ecart type de l'erreur
0.27	0.94	0.25

Tableau 3.1 – Statistiques sur les erreurs commises.

estimer l'écart entre le véritable mouvement et le mouvement estimé. D'autres tests ont été réalisés avec des séquences d'images prises dans des conditions réelles. Comme pour de telles séquences on ne peut estimer l'écart entre les paramètres réels et les paramètres estimés, les résultats produits ne sont que qualificatifs. C'est pourquoi nous les avons inclus dans le dernier chapitre traitant des applications potentielles de notre technique. Dans ces tests, nous pourrions apprécier la robustesse de l'algorithme aux variations de profondeur de champ et aux occlusions.

Dans l'algorithme d'estimation de mouvement, la procédure de vote délivre les valeurs des deux bornes de l'accumulateur imprécis associé à la partition floue de l'espace paramétrique du mouvement. Ces deux bornes sont représentées sur la figure 3.22. Puisque l'accumulateur est un tableau à trois dimensions (les trois variables du modèle de déplacement), sa représentation graphique dans le plan est assez complexe. Nous l'avons donc représenté par coupes successives suivant le paramètre associé aux rotations. Un des accumulateurs "2-D" représenté sur la figure 3.22 correspond en fait à l'accumulateur associé à une rotation d'angle $R_k \in \{R_{-K}, \dots, R_K\}$ (ensemble des valeurs discrétisées de rotations), et à toutes les combinaisons de translations horizontales et verticales.

Le mouvement majoritaire est obtenu par estimation du mode de l'histogramme approché. Le résultat présenté dans ce paragraphe correspond à la simulation d'une centaine de déplacements identiques non-entiers à savoir $(T_{hor}, T_{ver}, R_\phi) = (1.1 \text{ pixels}, 2.5 \text{ pixels}, 3^\circ)$ pour différents couples d'images. L'espace de recherche du mouvement est délimité par $T_{hor} \in [-4, 4]$ pixels, $T_{ver} \in [-4, 4]$ pixels, $R_\phi \in [-10^\circ, 10^\circ]$. La partition floue est composée de 9 cellules pour chaque paramètre. La figure 3.23 illustre les résultats de l'estimation des trois paramètres du mouvement dominant. La figure 3.24 détaille l'erreur commise lors de l'estimation des trois paramètres.

On peut voir sur la figure 3.24 que l'erreur commise sur l'estimation des paramètres est plus petite que la granularité du partitionnement. En effet, la granularité de la partition pour les translations est de ± 1 pixel et celle pour les rotations est de ± 2.5 degrés. Les erreurs maximales sont bien inférieures à ces limites. Quelques statistiques sur les essais réalisés sont résumées dans le tableau 3.1 :

On peut voir dans ce tableau que l'erreur commise sur l'estimation des paramètres

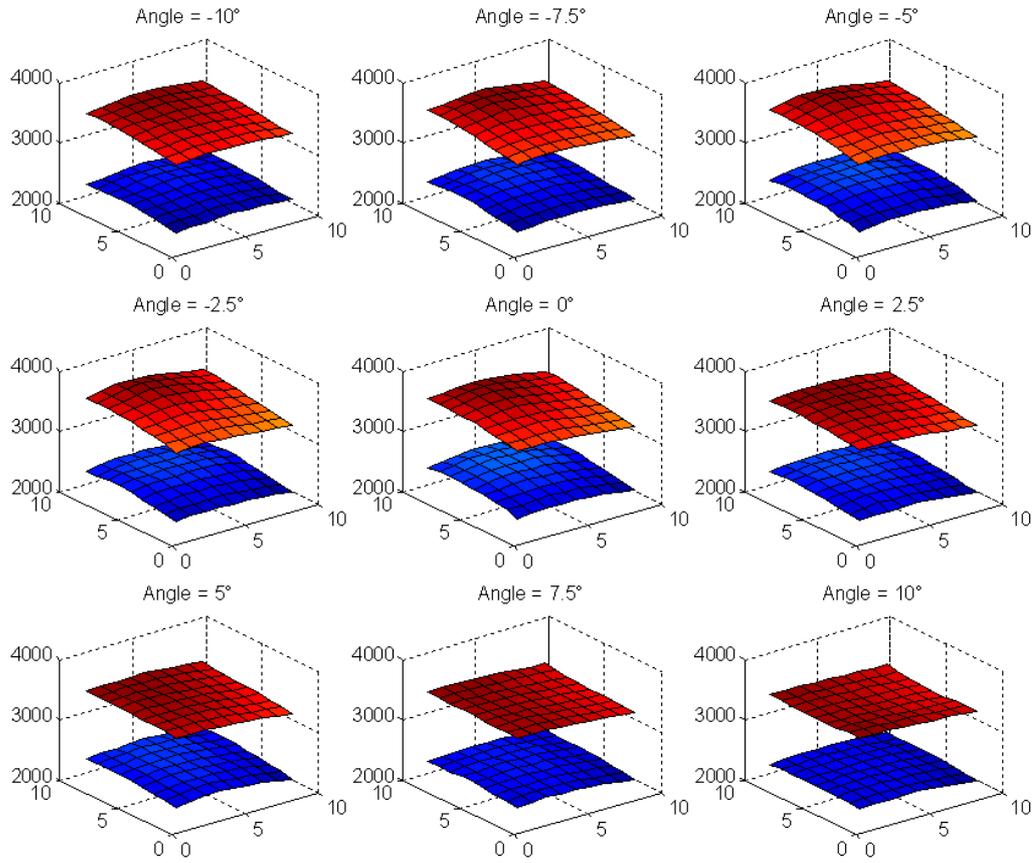


Figure 3.22 – Accumulateurs supérieurs et inférieurs associés aux paramètres du mouvement, représentés en coupes, pour différentes rotations R_ϕ .

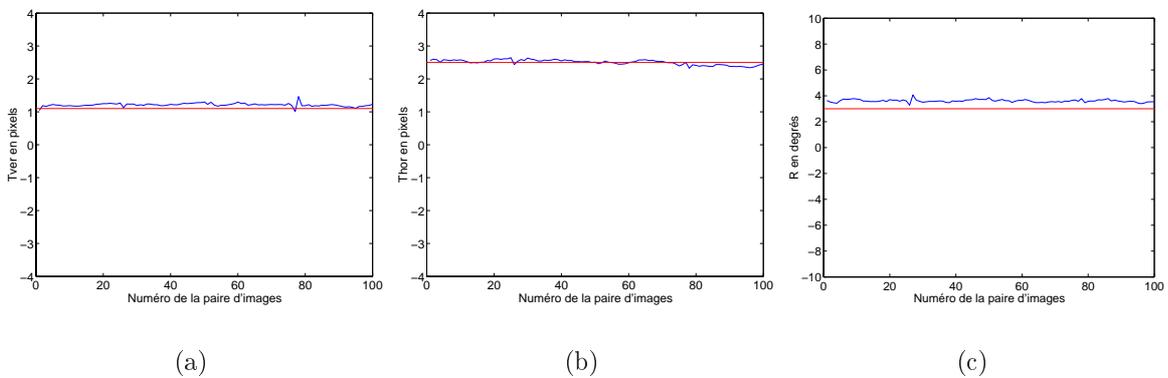


Figure 3.23 – Estimation de T_{hor} (a), T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images.

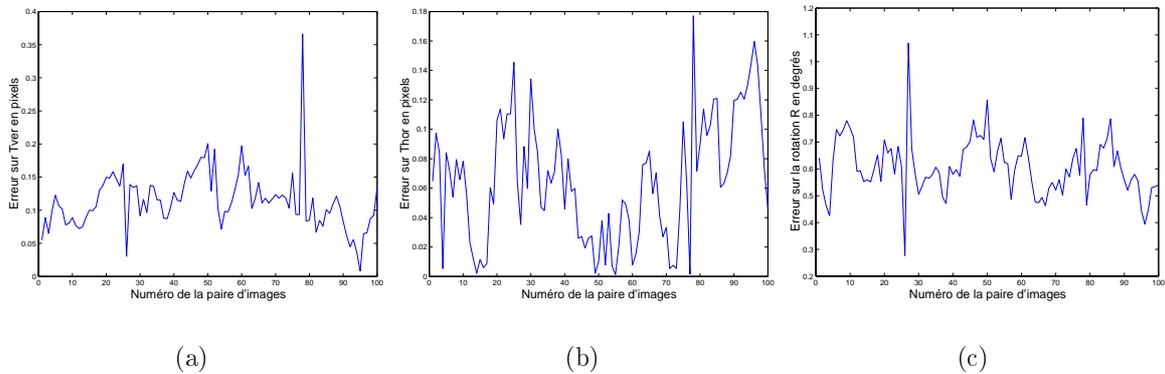


Figure 3.24 – Erreur commise pour l’estimation de T_{hor} (a), T_{ver} (b) et de la rotation R_ϕ (c) pour les 100 paires d’images.

est en moyenne de 10% de la granularité de la partition (pour chacun des paramètres). De plus, l’écart type étant très faible, la moyenne de la valeur absolue des erreurs est bien représentative de la majorité des erreurs.

3.5.2 Test de robustesse

Nous allons maintenant évaluer la robustesse de la méthode d’estimation de mouvement vis-à-vis des perturbations usuelles rencontrées dans une séquence vidéo.

Dans un premier temps, nous avons testé l’influence d’une variation de luminosité globale sur le processus d’estimation de mouvement. Les variations de luminosité sont fréquentes soit à cause du changement d’illumination de la scène filmée (ombres, changement du point de vue), soit à cause des systèmes d’adaptation automatique d’ouverture du diaphragme (WebCam). Pour simuler ce phénomène, nous avons repris la série de paires d’images utilisées au test précédent. Pour chaque paire, les images numéro 1 n’ont pas été modifiées. Par contre, nous avons artificiellement diminué la luminosité des images numéro 2 de 25%. Ainsi, les pixels blancs et gris devenant un peu plus foncés, quant aux pixels noirs, ils ne changent pas. En fait cela correspond à une diminution d’environ 60 du niveau de gris des pixels. Pour illustrer les changements que cela produit, une image de la séquence ainsi que cette même image avec la luminosité baissée de 25% sont représentées sur la figure 3.25.

Nous avons à nouveau estimé le déplacement apparent entre les deux images de chaque paire. Les résultats obtenus sont présentés sur les figures 3.26 et 3.27. Quelques statistiques sur les valeurs absolues des erreurs commises sont présentées dans le tableau 3.2. Ces résultats sont issus des nombreux essais réalisés et sont pleinement représentatifs du comportement de notre méthode.

Une variation de 25% de la luminosité globale est une perturbation assez importante vis-à-vis du niveau de gris des pixels. Cependant, on peut voir dans le tableau 3.2 que les erreurs sur l’estimation des trois paramètres restent faibles en moyenne (de l’ordre



Figure 3.25 – Image initiale (a) puis la même image avec une diminution de 25% de la luminosité (b).

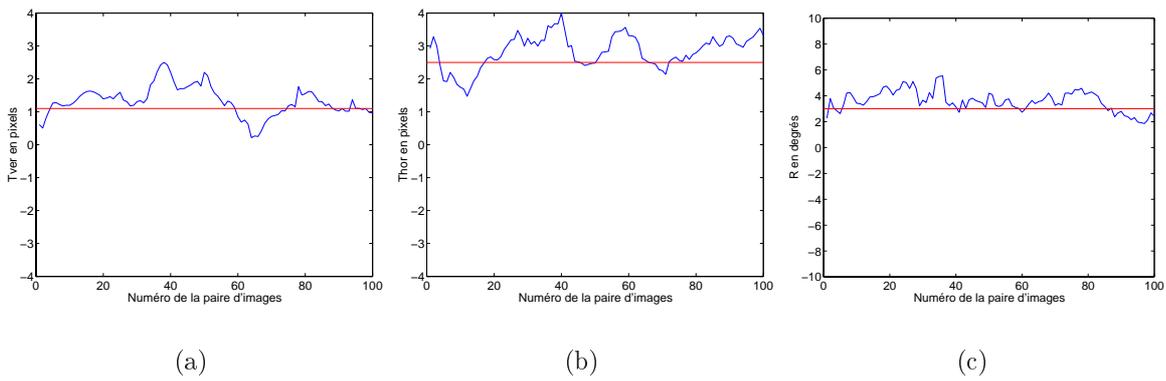


Figure 3.26 – Estimation de T_{hor} (a) T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images avec une variation de la luminosité de 25% entre les deux images.

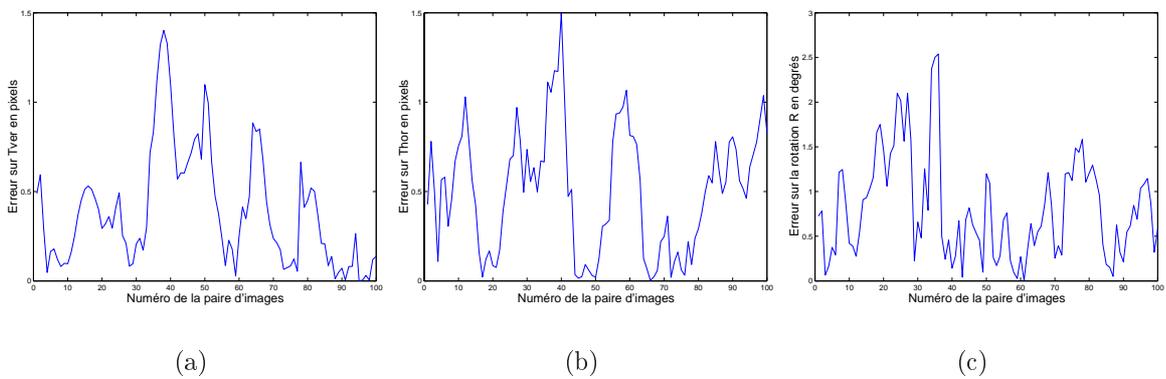


Figure 3.27 – Erreur sur l'estimation des paramètres du mouvement pour 100 paires d'images avec modification de la luminosité de 25% pour la seconde image de la paire.

T_{hor}		
Erreur moyenne (en pixels)	Erreur maximum (en pixels)	Ecart type de l'erreur
0.39	1.40	0.32
T_{ver}		
Erreur moyenne (en pixels)	Erreur maximum (en pixels)	Ecart type de l'erreur
0.50	1.50	0.34
$Rotation$		
Erreur moyenne (en degrés)	Erreur maximum (en degrés)	Ecart type de l'erreur
0.81	2.54	0.58

Tableau 3.2 – Statistiques sur les erreurs commises : influence d'une variation globale de luminosité de 25%

de la moitié du pas d'échantillonnage soit 0.5 pixels pour les translations et 0.9° pour la rotation). Les erreurs maximales ont par contre fortement augmenté par rapport aux précédents essais (1.5 pixel au lieu de 0.5 pixels pour les translations et 2.5° au lieu de 1° pour la rotation). Une explication réside dans le fait que la robustesse vis-à-vis des variations de luminosité dépend fortement de la nature des images traitées. Par exemple, prenons le cas de l'image 3.28 (a). Une variation de luminosité de 16%, seule, transforme 3.28 (a) en 3.28 (b).

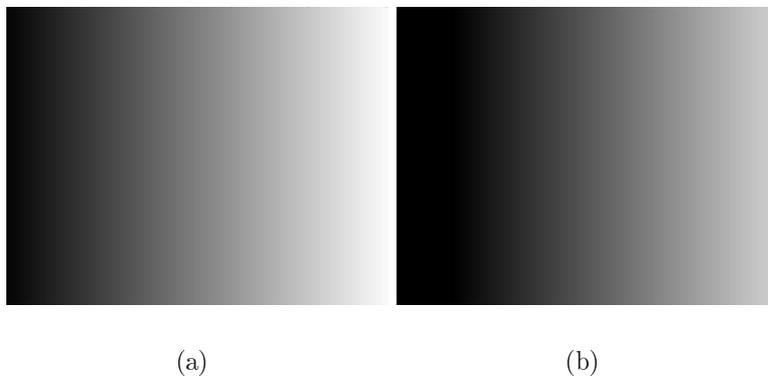


Figure 3.28 – Image d'un dégradé de niveaux de gris (a) et la même image avec une variation de l'intensité de 16% (b).

On est incapable de faire la différence entre cette variation de luminosité globale et un déplacement vers la gauche de la caméra. Dans ce cas, l'estimation du mouvement à de grandes chances d'être biaisée.

Dans un second temps, nous avons testé la robustesse de l'estimation de mouvement vis-à-vis d'un mouvement parasite. Pour cela, nous avons simulé le déplacement d'une caméra devant une scène fixe. Puis nous avons incrusté sur cette scène un objet se déplaçant. Celui-ci occulte une partie des pixels appartenant à l'image, et génère des pixels ayant un mouvement différent du mouvement de la caméra. L'objet, s'il est totalement présent dans l'image filmée occupe environ 15% de l'image. La figure 3.29

illustre l'estimation des paramètres du mouvement. La figure 3.30 représente les erreurs commises par le processus d'estimation. Ces erreurs sont synthétisées dans le tableau 3.3 à l'aide de la moyenne, de la valeur maximum et de l'écart type des erreurs.

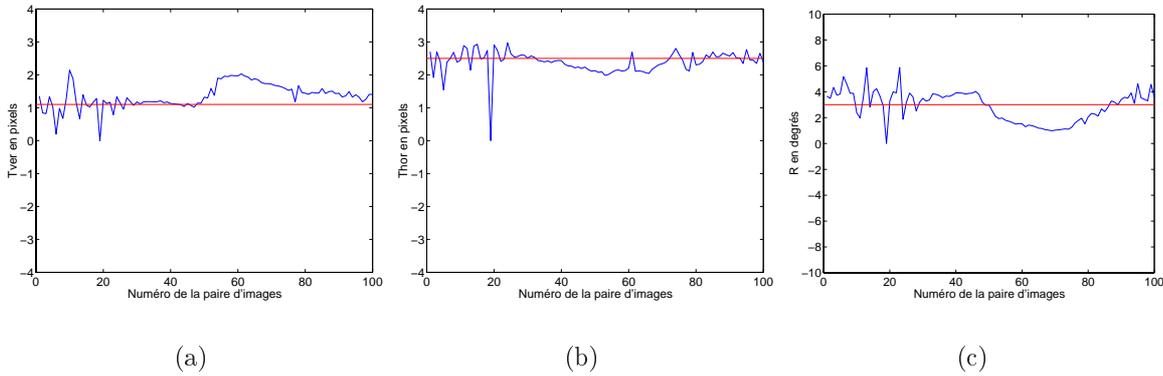


Figure 3.29 – Estimation de T_{hor} (a) T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images contenant un mouvement parasite.

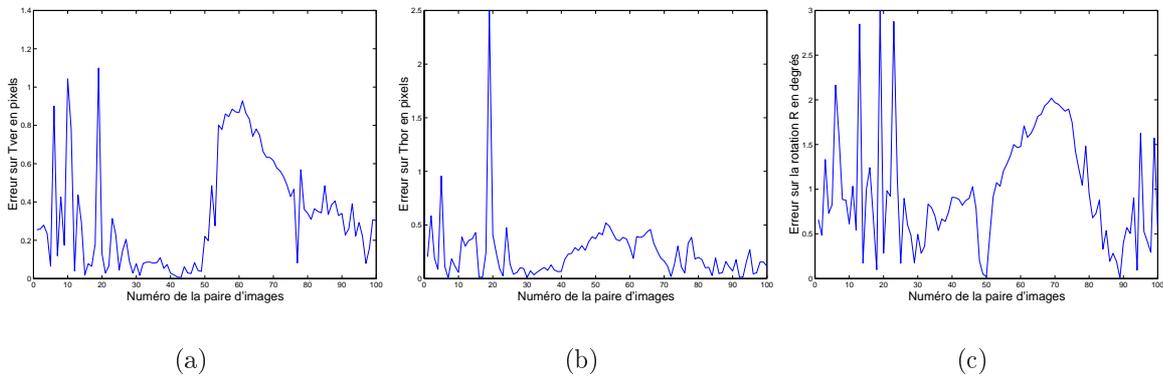


Figure 3.30 – Erreur sur l'estimation de T_{hor} (a) T_{ver} (b) et de la rotation R_ϕ (c) pour 100 paires d'images avec un mouvement parasite affectant 15% de la surface de l'image.

Les erreurs moyennes ont légèrement augmenté par rapport à l'expérience sans les perturbations mais restent cependant faibles face au pas d'échantillonnage des paramètres. Les erreurs moyennes passent de 0.1 pixels à 0.4 pixels pour les translations et de 0.3° à 1° pour les rotations. Seules les erreurs maximales ont subi une augmentation plus importante. Ces dernières sont environ multipliées par un facteur 3 (de 0.5 pixels à 1.5 pixels et de 0.9° à 3.5°). Ces essais montrent la robustesse de cette méthode vis-à-vis des mouvements parasites minoritaires. Cette robustesse peut s'expliquer par le fait que l'on estime le mouvement dominant à l'aide d'un vote majoritaire. En effet, si une minorité de pixels se retrouvent masqués ou ne votent pas pour le bon déplacement, ceci n'aura qu'une répercussion mineure sur l'ensemble des votes. Il en est de même pour les points qui changent de niveau de gris suite à un changement d'orientation de la caméra modifiant ainsi la quantité de lumière réfléchiée par une surface.

T_{hor}		
Erreur moyenne (en pixels)	Erreur maximum (en pixels)	Ecart type de l'erreur
0.34	1.1	0.29
T_{ver}		
Erreur moyenne (en pixels)	Erreur maximum (en pixels)	Ecart type de l'erreur
0.24	2.5	0.28
$Rotation$		
Erreur moyenne (en degrés)	Erreur maximum (en degrés)	Ecart type de l'erreur
1.01	3.50	0.87

Tableau 3.3 – Statistiques sur les erreurs commises : influence de mouvements parasites affectant 15% de la surface de l'image.

Des essais sur des séquences réelles seront présentés dans le chapitre 4. Nous y verrons entre autres la construction de mosaïques d'images à partir d'une séquence vidéo. Cependant, l'évaluation des performances dans ce cas reste plus subjective.

3.6 Comparaison avec les autres méthodes

Pour savoir si la formulation que nous avons employée pour la problématique "estimation de mouvement" est plus robuste que celle des méthodes classiques, nous avons souhaité comparer leurs performances respectives. Nous avons donc testé les différentes méthodes d'estimation de mouvement sur des séquences d'images identiques. L'exemple que nous avons utilisé pour illustrer notre propos est issu des nombreux essais que nous avons réalisés. Nous avons choisi d'utiliser une séquence présentant à la fois des mouvements de translation et de rotation, des variations globales d'intensité et des mouvements parasites. De façon à pouvoir tester toutes les méthodes dans de bonnes conditions d'utilisation (et particulièrement pour le flot optique où les mouvements doivent être de faible amplitude), nous avons généré une séquence de mouvements aléatoires compris entre plus ou moins 4 pixels pour les translations et plus ou moins 10° pour la rotation.

Pour les deux méthodes nécessitant une discrétisation de l'espace paramétrique (corrélation et histogramme approché), nous avons imposé le même échantillonnage des paramètres à savoir : $T \in [-4, 4]$ avec $\Delta T = 1$ pixel (soit 9 cellules) et $R_\phi \in [-10^\circ, 10^\circ]$ avec $\Delta R_\phi = 2.5^\circ$ (soit aussi 9 cellules).

Dans un premier temps, nous avons comparé les performances des méthodes dans le cas optimum, c'est-à-dire le cas d'une scène statique sous un éclairage constant. Les comparaisons des valeurs absolues des erreurs sont représentées dans le tableau 3.4. On peut voir sur ces résultats, que les algorithmes de corrélation ont des difficultés à estimer correctement un mouvement contenant des rotations. En effet, l'erreur commise sur ce paramètre est à peu près dix fois plus importante que lors de l'estimation avec notre méthode. Cependant, notre méthode met en oeuvre une interpolation, ce

que ne font pas les méthodes basées sur la corrélation. L'erreur commise lors de l'estimation du mouvement avec ces dernières méthodes est alors plus ou moins dictée par l'échantillonnage de l'espace paramétrique. En revanche, le temps de calcul des méthodes apparentées à la corrélation est considérablement moins important. On a en effet un facteur 100 entre les deux approches (0.9 secondes contre 90 secondes pour une paire d'image de 60×80 pixels sur un ordinateur type Pentium III cadencé à 866Mhz). Cependant, l'optimisation du temps de calcul de notre méthode n'est pas encore effectuée, et nous pouvons raisonnablement penser le diminuer d'un facteur 100. L'algorithme d'estimation du mouvement avec une approche basée sur le flot optique donne de bons résultats très similaires à notre méthode. Les erreurs moyennes commises sont globalement du même ordre de grandeur. Cependant, les erreurs maximales sont plus faibles.

Les résultats figurant dans le tableau 3.5 concernent l'essai où nous avons modifié artificiellement la luminosité de la seconde image de chaque paire. Là encore, on peut voir que les erreurs commises avec les méthodes de corrélation sont en moyenne plus importantes que celles obtenues avec notre méthode. On peut aussi noter que la mise en correspondance utilisant la transformation CENSUS est effectivement plus robuste que les autres méthodes de corrélation vis à vis des changements de luminosité. Ceci est dû au fait qu'elle utilise, pour estimer le mouvement, un ordonnancement des pixels en fonction de leurs niveaux de gris et pas directement leurs niveaux de gris. La robustesse de notre méthode vis-à-vis de ces perturbations s'explique par l'utilisation des classes de niveau de gris des pixels. Dans les deux cas, le passage par une classification (locale pour CENSUS et globale pour notre méthode) apporte une grande robustesse. A nouveau dans cet essai, l'erreur commise sur l'estimation des angles est beaucoup plus importante qu'avec l'estimation par histogramme approché. La méthode de flot optique utilisée par contre est très sensible aux variations de luminosité globale. Ceci est lié à l'hypothèse de conservation de l'illumination qui n'est pas respectée dans cette expérience. Dans ce cas, l'estimation du mouvement est totalement inexploitable.

Le dernier essai consiste à étudier la robustesse vis à vis des occlusions. Nous avons repris l'essai où nous avons inséré un objet se déplaçant avec son mouvement propre dans la séquence. La comparaison des résultats est présentée dans le tableau 3.6. On voit que notre méthode est plus robuste que les méthodes basées sur un calcul de corrélation. Tout particulièrement, la méthode CENSUS présente des difficultés à gérer les occlusions. L'ordonnancement des pixels se trouve fortement perturbé par un mobile se déplaçant librement dans la séquence vidéo. L'estimation du mouvement dominant est ainsi fortement biaisée. Pour l'estimation du mouvement apparent utilisant un calcul robuste du flot optique, les résultats obtenus sont très satisfaisants. Hormis une paire d'images où l'estimation n'a pas fonctionné, ce qui explique l'erreur maximale relativement importante, la présence d'occlusions générées par le mouvement parasite ne constitue pas une gêne pour ce type d'approche.

3.7 Processus multi-échelles d'estimation de mouvement

Nous présentons ici les résultats en termes de gain en temps de calcul que nous avons obtenus pour l'estimation de mouvement avec notre processus multi-résolution (PMR).

Nous avons constaté au cours de nos expérimentations que l'estimation de trois paramètres (deux translations et une rotation) est nettement plus longue que l'estimation de deux paramètres (2 translations seules). Nous avons limité l'estimation multi-résolution à la recherche des translations dans un souci de gain de temps. La rotation est uniquement estimée à la résolution 0 (niveau le plus fin de représentation de l'image). Ceci ce justifie lorsque l'on considère une rotation faible entre les deux images. Par ailleurs, lorsque l'image est dégradée les faibles rotations disparaissent. **Le PMR est donc présenté pour un modèle de mouvement ne contenant que deux translations.** Le raisonnement que nous effectuons se généralise aisément dans le cadre d'un PMR sur les trois paramètres.

3.7.1 Apport de la représentation multi-échelles

Le processus multi-résolution consiste à rechercher le mouvement apparent à l'aide d'une approche pyramidale allant de la résolution la plus grossière vers la plus fine. Le mouvement est donc estimé sur des images représentées à différents niveaux de précision. Un déplacement de grande amplitude à l'échelle la plus précise se traduit par un déplacement d'amplitude de plus en plus faible à mesure que le niveau de résolution de l'image diminue. La taille de la zone de recherche du mouvement est définie à la résolution la plus grossière et peut alors être conservée tout au long du PMR. L'utilisation du PMR permet d'estimer des déplacements de grande amplitude tout en gardant un espace paramétrique de taille réduite.

Nous utilisons les histogrammes approchés pour trouver le mode principal d'une distribution statistique. Pour ce faire, nous accumulons des informations imprécises probabilisées pour construire une approximation de la densité de répartition de cette distribution statistique. Or, pour pouvoir garantir un minimum de pertinence à une estimation statistique, celle-ci doit être réalisée sur au moins trente échantillons pour un moment d'ordre un ou cent échantillons pour un moment d'ordre deux [WW91]. Dans le cas un peu particulier qui nous intéresse, ce sont les pixels d'une image qui votent pour les déplacements et qui constituent les "échantillons". Pour que le vote d'un pixel de l'image 1, pour une transformation T , soit informatif, il faut que son image par cette transformation se situe à l'intérieur des limites de l'image 2 (sinon les votes pour et contre le déplacement sont identiques *cf.* paragraphe 3.2). Il faut donc qu'à n'importe quelle échelle de représentation, on ait suffisamment de pixels votant pour les déplacements envisagés. Cette contrainte nous amène à définir une règle reliant la résolution la plus grossière de représentation des images et l'amplitude maximale du mouvement à estimer.

Les pixels de l'image 1 ayant le plus de chance de se retrouver, après application d'une transformation, à l'intérieur des limites de l'image 2 sont les pixels du centre de l'image 1 (dans le cas d'un mouvement composé de translations). Ces pixels sont compris dans une zone rectangulaire de α lignes et β colonnes (figure 3.31). Il faut donc que le nombre de pixels compris dans cette zone ($N_{\alpha\beta} = \alpha \times \beta$) soit suffisamment grand pour pouvoir effectuer l'estimation de mode.

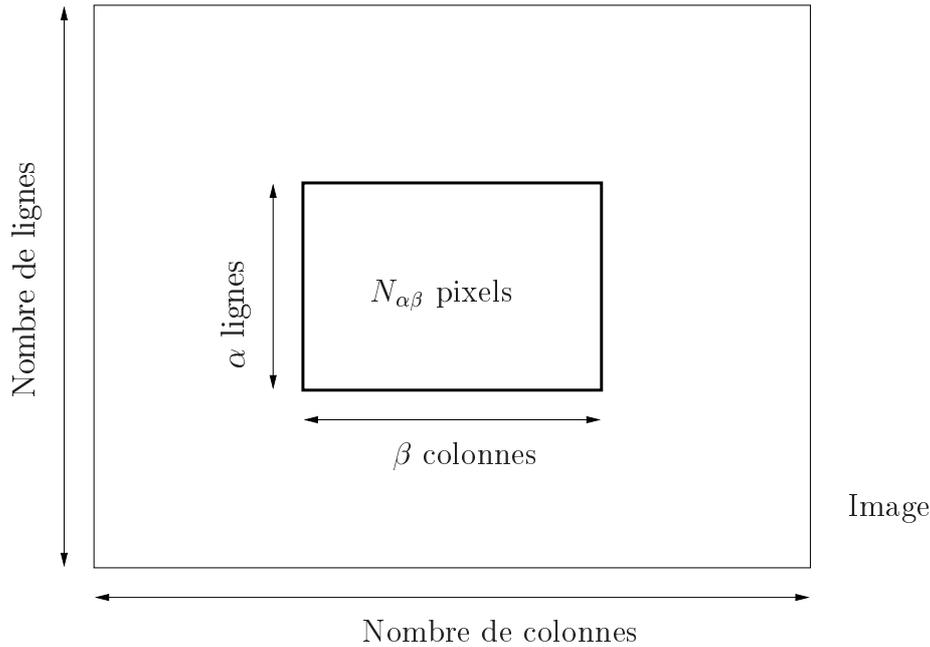


Figure 3.31 – Zone centrale de l'image 1 contenant les pixels ayant le plus de chance de se retrouver dans l'image 2 suite à une transformation.

La condition pour que les pixels se retrouvent, après transformation, dans les limites de l'image 2 est illustrée sur la figure 3.32. Il faut impérativement que le nombre de lignes (resp. de colonnes) de l'image soit supérieur à la taille de la zone de recherche du pixel, c'est à dire, α (resp. β) auquel on ajoute deux fois la translation verticale (resp. horizontale) maximum. On ajoute deux fois la translation maximale pour que la zone de recherche couvre un déplacement allant de moins la translation maximale à plus la translation maximale.

Cette propriété doit être vérifiée dans le cas le moins favorable, c'est à dire quand l'image est la plus petite, autrement dit à la résolution la plus grossière. Ainsi, la contrainte reliant la résolution maximale et le déplacement d'amplitude maximum est définie par :

$$N_{LinMin} = \alpha + \text{Taille Zone Recherche Lin} \quad (3.28)$$

$$N_{ColMin} = \beta + \text{Taille Zone Recherche Col} \quad (3.29)$$

Les grandeurs N_{LinMin} et N_{ColMin} sont définies par :

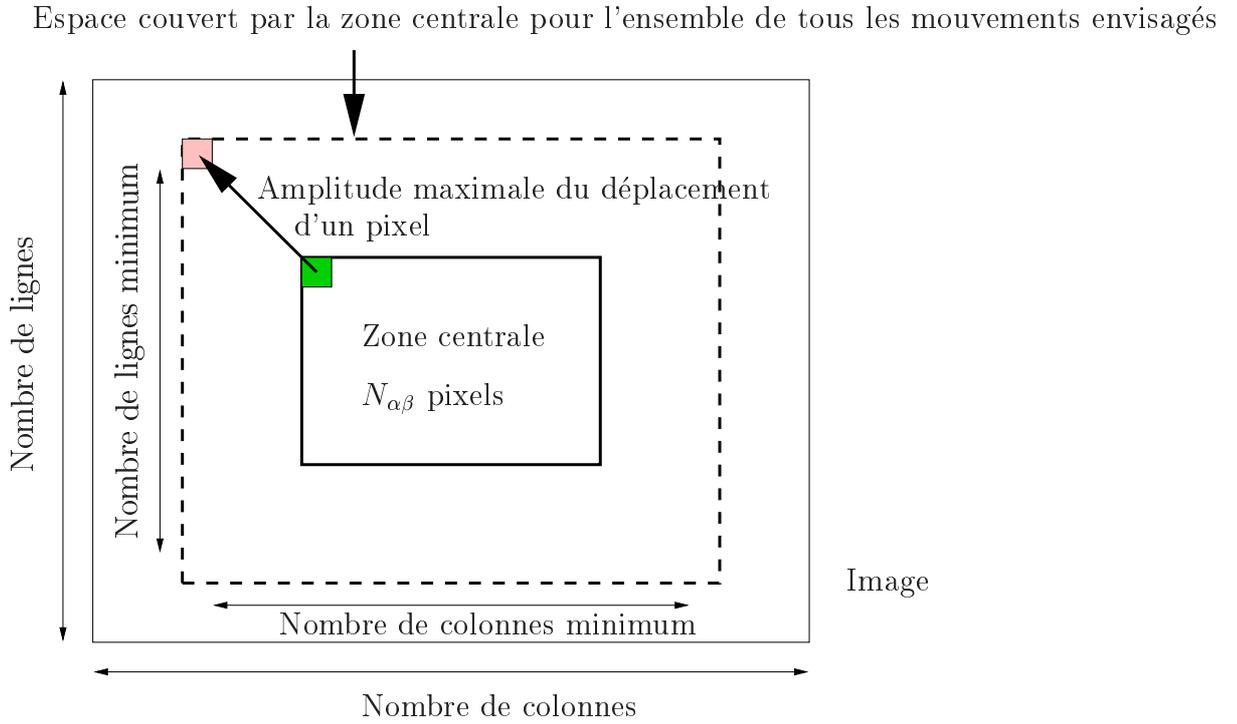


Figure 3.32 – Taille minimale de l'image à une résolution donnée fonction de la zone de recherche du mouvement.

$$N_{LinMin} = \frac{N_{lin}}{2^{résolution_{max}}} \quad (3.30)$$

$$N_{ColMin} = \frac{N_{col}}{2^{résolution_{max}}} \quad (3.31)$$

où N_{lin} et N_{col} sont les dimensions de l'image à la résolution la plus fine. La taille de la zone de recherche est fonction de l'étendue du mouvement que l'on veut estimer. En effet, pour estimer un mouvement compris entre $\pm DepMax$ pixels à la résolution la plus fine, la zone de recherche doit être au minimum de $\frac{2DepMax}{2^{résolution_{max}}}$ à la résolution la plus grossière. Le 2 au numérateur vient du fait que l'on s'autorise un déplacement allant de $-DepMax$ à $+DepMax$. Le dénominateur est lié au fait que le processus multi-échelles divise le nombre de lignes et le nombre de colonnes par 2 à chaque changement de résolution. Ainsi, la taille de la zone de recherche pour les lignes et les colonnes est donnée par :

$$N_{LinMin} = \alpha + \frac{2 \times DepMax_{lin}}{2^{Résolution_{max}}} \quad (3.32)$$

$$N_{ColMin} = \beta + \frac{2 \times DepMax_{col}}{2^{Résolution_{max}}} \quad (3.33)$$

En utilisant les équations (3.30), (3.31), (3.32) et (3.33) on obtient :

$$\text{Résolution}_{\max 1} = \log_2 \left(\frac{N_{lin} - 2 \times \text{DepMax}_{lin}}{\alpha} \right) \quad (3.34)$$

$$\text{Résolution}_{\max 2} = \log_2 \left(\frac{N_{col} - 2 \times \text{DepMax}_{col}}{\beta} \right) \quad (3.35)$$

De ces deux équations, on déduit :

$$\text{Résolution}_{\max} = \min(\text{Résolution}_{\max 1}, \text{Résolution}_{\max 2}) \quad (3.36)$$

Nous ne pouvons pas donner une formulation mathématique exacte du gain en temps de calcul. En effet, chaque transformation imprécise d'un pixel de l'image 1 (cf. chapitre 3) est amenée à recouvrir plusieurs pixels de l'image 2. Le nombre de pixels recouverts dépend de la position du pixel, de la transformation considérée, etc. Nous avons donc défini un temps élémentaire, correspondant au temps maximum nécessaire au calcul de l'éventualité de déplacement d'un pixel. Le gain en temps de calcul sera donné en fonction de ce temps élémentaire, et ne sera qu'une approximation du vrai gain.

Soit une image de taille (N_{lin}, N_{col}) à la résolution 0. Si on souhaite estimer un déplacement de $\pm \text{DepMax}_{lin}$ et $\pm \text{DepMax}_{col}$, sans le processus multi-résolution, il faudra :

$$\text{Nombre de calculs} = 2 \times \text{DepMax}_{lin} \times 2 \times \text{DepMax}_{col} \times N_{lin} \times N_{col} \quad (3.37)$$

Dans le cas multi-résolution, il faudra :

$$\begin{aligned} \text{Nombre de calculs} &= \sum_{\text{Résolution}=0}^{\text{Résolution}_{\max}} 2 \times \frac{\text{DepMax}_{lin}}{2^{\text{Résolution}_{\max}}} \times 2 \times \frac{\text{DepMax}_{col}}{2^{\text{Résolution}_{\max}}} \\ &\times \frac{N_{lin}}{2^{\text{Résolution}}} \times \frac{N_{col}}{2^{\text{Résolution}}} \end{aligned} \quad (3.38)$$

$$\text{Nombre de calculs} = \sum_{\text{Résolution}=0}^{\text{Résolution}_{\max}} \frac{\text{DepMax}_{lin} \times \text{DepMax}_{col} \times N_{lin} \times N_{col}}{2^{2(\text{Résolution}_{\max} + \text{Résolution} - 1)}} \quad (3.39)$$

3.7.2 Influence de l'estimation du mode de la distribution des mouvements sur le PMR

Il y a un aspect que nous n'avons pas encore abordé : l'influence du temps de détermination du mode de la distribution des mouvements. Elle nécessite elle aussi un

certain nombre de calculs. Comme nous l'avons vu dans le chapitre 2, l'estimation du mode se fait par blocs de 4^τ cellules de la zone de recherche, où τ est le nombre de paramètres du modèle de déplacement. La taille de la zone de recherche doit au moins comprendre 4 cellules pour chaque paramètre à la résolution la plus grossière. De cette constatation on peut déduire une nouvelle contrainte sur la résolution maximum :

$$\left(2 \times \frac{DepMax_{col}}{2^{Résolution_{max}}} > 4\right) \& \left(2 \times \frac{DepMax_{lin}}{2^{Résolution_{max}}} > 4\right) \quad (3.40)$$

On en déduit :

$$Résolution_{max} < \min \left(\log_2 \left(\frac{DepMax_{col}}{2} \right), \log_2 \left(\frac{DepMax_{col}}{2} \right) \right)$$

Soit encore :

$$Résolution_{max} < \min (\log_2 (DepMax_{col}), \log_2 (DepMax_{col})) - 1 \quad (3.41)$$

Finalement, la résolution maximum sera le minimum des résolutions calculées aux équations (3.36) et (3.41).

$$Résolution_{max} = \min \left(\log_2 \left(\frac{N_{lin} - 2 \times DepMax_{lin}}{\alpha} \right), \log_2 \left(\frac{N_{col} - 2 \times DepMax_{col}}{\beta} \right), \left(\min (\log_2 (DepMax_{col}), \log_2 (DepMax_{col})) - 1 \right) \right) \quad (3.42)$$

Si le PMR est uniquement appliqué aux translations, la recherche du mode se fera sur des pavés de 4×4 cellules de l'histogramme des mouvements (*cf.* chapitre 2). Notons Tm le temps nécessaire à l'estimation du mode local sur un pavé de 4×4 cellules. Sans le PMR il y a $(2 \times DepMax_{lin} - 3) \times (2 \times DepMax_{col} - 3)$ blocs de 4×4 cellules dans l'histogramme. L'estimation du mode prendra donc :

$$Temps = (2 \times DepMax_{lin} - 3) \times (2 \times DepMax_{col} - 3) \times Tm \quad (3.43)$$

Avec le PMR, la taille de la zone de recherche est diminuée, par contre un histogramme approché est calculé pour chaque résolution. Ainsi, le temps pour l'estimation du mode sera de :

$$Temps = \left(2 \times \frac{DepMax_{lin}}{2^{Résolution_{max}}} + 1 - 3\right) \times \left(2 \times \frac{DepMax_{col}}{2^{Résolution_{max}}} + 1 - 3\right) \times (Résolution_{max} + 1) \times Tm \quad (3.44)$$

Prenons un exemple concret : soit une image de taille $N_{lin} = 480$, $N_{col} = 640$. le déplacement maximum à estimer est de : $DepMax_{lin} = DepMax_{col} = 25$. Sans le

processus multi-résolution il faut 768.10^6 calculs élémentaires. Supposons qu'il faille un minimum de 30 pixels pour le processus d'estimation ($N_{\alpha\beta} > 30$). On prendra pour simplifier le calcul $\alpha = \beta = 6$. Le calcul de la résolution maximale à l'aide de l'équation 3.42 donne : $Résolution_{\max} = \min(3, 6, 6) = 3$. On peut alors estimer que le nombre de calculs élémentaires avec le PMR vaut environ 20.10^6 (soit 2.6% du nombre calculé sans le PMR). Pour ce qui est de l'estimation du mode, sans le PMR on a $2209 \times Tm$ contre $42.25 \times Tm$ avec le PMR (soit environ 1.9% du temps précédent). Il est donc évident que le processus multi-résolution nous permet d'améliorer le temps de calcul de notre procédé d'estimation de mouvement. De plus, l'importance du PMR est croissante à mesure que la taille de l'image augmente (vis-à-vis de la réduction du nombre de calculs élémentaires).

En effectuant l'estimation de la rotation à la résolution la plus fine, et à titre purement indicatif, nous avons mesuré le temps de calcul de l'estimation de mouvement pour deux images de taille 176×120 pour une zone de déplacement de ± 10 pixels en translation horizontale et verticale (par pas de 1 pixel) et de $\pm 10^\circ$ pour la rotation (par pas de 1°). L'essai a été réalisé sur un ordinateur de type Pentium III cadencé à 866MHz. Avec le PMR le temps d'exécution est de 494 secondes (8 min 14 sec) alors que sans le PMR le temps d'exécution est de 4264 secondes (1h11 min 4 sec), soit un rapport de 11.6%. L'intérêt du PMR est largement mis en évidence par ces résultats.

Comme on peut le voir sur les résultats expérimentaux, l'amélioration du temps de calcul lié au processus multi-résolution ne correspond pas aux calculs théoriques donnés. Le rapport théorique entre les deux temps de calcul aurait dû être de 6.3%, et nous avons pratiquement le double. Cependant, les temps donnés ne concernent que les parties de calcul de l'algorithme. Or, la programmation multi-résolution nécessite l'allocation récursive d'espace mémoire pour stocker les images et des appels de fonction eux aussi récursifs, qui sont très coûteux en temps de calcul. Le gain en temps de calcul peut être établi en réalisant une étude minutieuse de la complexité de notre algorithme.

3.8 Vers le temps réel?

Nous nous sommes bien sûr intéressé à la possibilité de réaliser un algorithme en temps réel, c'est-à-dire un temps pour l'estimation de mouvement entre deux images inférieur ou proche du délai d'acquisition soit $1/25^{\text{ème}}$ de seconde.

Tel qu'il vient d'être présenté, notre algorithme ne peut être exécuté dans un laps de temps aussi court. Pour en accélérer le calcul, nous avons opté pour un certain nombre de techniques calculatoires.

Dans un premier temps nous réduisons la dimension de l'espace paramétrique à deux, ce qui revient à n'estimer que les translations.

Dans un second temps, au lieu d'utiliser tous les pixels des deux images pour calculer l'histogramme approché de la répartition du mouvement, nous n'utilisons qu'un échantillon réduit de l'image tiré pseudo-aléatoirement pour éviter de biaiser l'estimation.

Finalement nous avons remarqué que lorsque la subdivision de l'espace des translations était entière, un grand nombre de facteurs pouvaient être précalculés.

Avec ces approximations, nous avons effectué des estimations de mouvement entre deux images de taille 640 par 480 pixels pour une zone de recherche des translations de 6×6 pixels. Pour atteindre quasiment le temps réel sur une machine type Pentium IV cadencé à 1.2GHz, nous sommes obligés de prendre moins d'un pourcent des points de l'image soit environ 2500 pixels (au lieu des 307200 pixels de l'image au total). L'estimation du mouvement avec seulement 1% des pixels est malgré tout satisfaisante.

3.9 Conclusion et perspectives

Ce chapitre a présenté une nouvelle approche pour l'estimation du mouvement apparent dans une séquence vidéo. Elle se différencie des approches classiques (flot optique, corrélation, etc.) par le fait qu'elle se base sur des hypothèses beaucoup moins contraignantes vis-à-vis de la modélisation des variations d'intensité des pixels. Cette approche est basée sur une représentation du niveau de gris des pixels à l'aide de deux classes duales floues. L'espace des mouvements de la caméra est discrétisé, et pour chacune des configurations de cet espace, un vote de l'ensemble des pixels est effectué. Ces votes sont accumulés dans un histogramme approché construit sur ce partitionnement de l'espace paramétrique du mouvement. Le mouvement apparent majoritaire est ensuite recherché comme étant le mode de la distribution des mouvements représentée sur l'histogramme approché.

Les essais réalisés pour tester les performances de notre méthode nous ont permis de remarquer que :

- Les performances de ce type d'estimation de mouvement sont très satisfaisantes. En effet les erreurs commises lors du processus d'estimation de mouvement sont largement inférieures à la granularité de l'échantillonnage de l'espace paramétrique.
- Notre méthode possède un bon comportement vis-à-vis des perturbations. Que ces perturbations proviennent de variations de l'éclairage ambiant ou d'un mouvement parasite, les erreurs restent inférieures en moyenne à la granularité du partitionnement. La modélisation des niveaux de gris des pixels à l'aide de deux classes duales floues permet de s'affranchir partiellement des problèmes de variation d'illumination. Pour ce qui est des occlusions ou des mouvements parasites, le fait de rechercher le mode principal de la distribution permet de s'affranchir en partie de ces problèmes. Le mouvement parasite sera représenté comme un mode secondaire de la distribution de mouvements et sera donc rejeté dans le processus d'estimation.

Cette méthode possède cependant quelques travers et nécessite des améliorations. Actuellement, le temps de calcul est beaucoup trop long. Il est en effet proportionnel à la taille de l'image multipliée par la taille de la zone de recherche du mouvement. Le traitement d'une petite image (60 par 80 pixels), pour un espace paramétrique du mouvement de $9 \times 9 \times 9$ cellules, prend environ 90 secondes par rapport à 0.9 secondes

pour une corrélation sur un Pentium III cadencé à 866 MHz. L'estimation d'un mouvement de plus grande amplitude est donc compromis par la lenteur de l'algorithme. Le problème du temps de calcul prohibitif est donc un point crucial à améliorer. Pour le réduire, nous avons utilisé une estimation du mouvement à l'aide de la représentation multi-échelles des images définie au paragraphe 3.1.3. Celle-ci permet de rechercher le mouvement à l'aide d'une approche pyramidale de la résolution la plus grossière vers la plus fine. Cependant, ce processus n'a pas encore été comparé en détail avec les procédures multi-échelles classiques, nous ne pouvons donc pas encore pleinement mesurer son efficacité. D'autre part, les gains en terme de temps de calcul sont présentés dans ce chapitre. Nous avons pu conclure que plus la taille des images est grande, plus l'intérêt d'un processus multi-résolution est important. Une étude plus poussée sur la complexité de l'algorithme permettrait sans doute de quantifier plus exactement le gain de temps apporté par notre procédure multi-résolution.

Dans le cas où le modèle de mouvement ne contient que des translations, nous avons déjà réussi à diminuer considérablement le temps de calcul (proche du temps réel). Pour cela, nous avons limité l'échantillonnage de l'espace des translations à des valeurs entières. Ceci permet d'effectuer des pré-traitements et de réduire ainsi la masse de calculs à réaliser à chaque étape de notre algorithme. Ces résultats, encourageants, nous laissent croire qu'il est aussi possible de réduire le temps de calcul dans le cas d'un modèle de déplacement contenant une rotation.

Le deuxième point à améliorer concerne la précision de la méthode. Actuellement, nous effectuons une approximation de la quantité floue associée à la transformation d'un pixel par un mouvement imprécis $(T_{hor}, T_{ver}, R_\phi)$, dans un but d'accélérer le calcul de $\Pi(P'_1; P_2)$ et $N(P'_1; P_2)$. Le fait de prendre une quantité floue englobante nous fait perdre en précision. Il serait donc judicieux de trouver une méthode pour définir plus finement la quantité associée à P'_1 . Nous espérons que l'utilisation des formules de découpe de polygones de Southerland-Hodgman [SH74] nous permettra de calculer les valeurs de $\Pi(P'_1; P_2)$ et $N(P'_1; P_2)$ avec une procédure relativement simple (même dans le cas où les deux quantités floues n'ont pas la même orientation).

Il reste par ailleurs des points que nous n'avons pas exploré. En effet, nous n'avons pas mesuré l'influence du nombre de classes de pixels sur le processus d'estimation de mouvement. Faut-il rajouter une classe intermédiaire de pixels gris ou bien d'autres classes intermédiaires? Dans ce cas, la complexité des formules (3.14) et (3.15) irait croissant à mesure de l'augmentation du nombre de classes. De plus, il faudra s'assurer que l'ajout de classes ne se fait pas au détriment de la robustesse de la méthode. Il s'agit donc de trouver le meilleur compromis s'il existe.

Finalement, une autre extension envisageable pour cette méthode d'estimation de mouvement consisterait à traiter les images en couleur et non plus seulement les images en niveaux de gris. Cette extension impliquerait une nouvelle modélisation des pixels. En effet les deux classes pixel noir et pixel blanc ne suffisent plus, il faut prévoir une classe pour chacune des trois composantes chromatiques de l'image : rouge, vert, bleu. De plus, les relations simples du type $\mu_b = 1 - \mu_n$ ne sont plus de mise avec les trois couleurs. La formalisation de ce problème risque donc d'être complexe. Cependant, le

fait d'introduire trois composantes au lieu d'une seule (la luminance) permettrait peut être de rendre l'estimation de mouvement encore plus robuste.

Erreur	T_{hor} en pixels			T_{ver} en pixels			Rotation en degrés		
	Moyenne	Maximum	Ecart type	Moyenne	Maximum	Ecart type	Moyenne	Maximum	Ecart type
Histo	0.12	1.31	0.15	0.11	2.92	0.29	0.31	3.28	0.48
SAD	0.96	1.99	0.54	0.88	1.96	0.59	2.93	9.86	2.30
ZSAD	0.49	0.99	0.27	0.54	0.99	0.30	4.90	9.99	2.84
SSD	1.59	2.95	0.73	0.54	0.99	0.30	2.99	7.50	2.04
ZSSD	1.59	2.95	0.73	0.54	0.99	0.30	2.99	7.50	2.04
NCC	1.56	2.95	0.75	0.55	1.48	0.33	2.81	7.50	2.0
ZNCC	1.56	2.95	0.75	0.55	1.48	0.33	2.81	7.50	2.0
Census	0.57	2.0	0.41	0.89	1.99	0.59	2.93	9.86	2.30
Flot optique	0.13	0.24	0.04	0.14	0.34	0.06	0.08	0.44	0.07

Tableau 3.4 – Statistiques sur les erreurs commises : absence de bruit.

Erreur	T_{hor} en pixels			T_{ver} en pixels			Rotation en degrés		
	Moyenne	Maximum	Ecart type	Moyenne	Maximum	Ecart type	Moyenne	Maximum	Ecart type
Histo	0.39	1.40	0.32	0.5	1.50	0.34	0.81	2.54	0.58
SAD	0.94	3.95	0.97	1.75	2.96	0.74	2.43	6.39	1.75
ZSAD	0.49	0.99	0.27	0.54	0.99	0.30	4.90	10	2.84
SSD	1.59	2.95	0.73	0.54	0.99	0.30	2.99	7.50	2.04
ZSSD	1.59	2.95	0.73	0.54	0.99	0.30	2.99	7.50	2.04
NCC	1.56	2.95	0.75	0.55	1.48	0.33	2.81	7.50	1.99
ZNCC	1.56	2.95	0.75	0.55	1.48	0.33	2.81	7.50	1.99
Census	0.57	2.0	0.41	0.90	1.99	0.59	2.93	9.86	2.30
Flot optique	25.88	63.20	19.06	17.22	68.85	19.00	3.46	4.01	0.21

Tableau 3.5 – Statistiques sur les erreurs commises : variation globale de luminosité de 25%.

Erreur	T_{hor} en pixels			T_{ver} en pixels			Rotation en degrés		
	Moyenne	Maximum	Ecart type	Moyenne	Maximum	Ecart type	Moyenne	Maximum	Ecart type
Histo	0.34	1.1	0.29	0.24	2.5	0.28	1.01	3.5	0.87
SAD	0.88	1.1	0.42	1.22	2.5	0.96	1.9	3.5	1.21
ZSAD	1.04	1.1	0.23	1.26	2.5	0.97	1.95	3.5	1.22
SSD	1.21	2.1	0.31	1.26	2.5	0.97	1.95	3.5	1.22
ZSSD	1.22	2.1	0.33	1.26	2.5	0.98	1.95	3.5	1.22
NCC	1.17	2.1	0.26	1.28	2.5	0.98	1.97	3.5	1.22
ZNCC	1.2	2.1	0.36	1.22	2.5	0.96	1.9	3.5	1.21
Census	1.94	4.1	1.86	0.9	6.5	1.17	3.91	6.49	2.7
Flot optique	0.49	22.6	2.28	0.32	5.27	0.76	4.49	4.55	0.05

Tableau 3.6 – Statistiques sur les erreurs commises : présence d'un mouvement parasite.

Chapitre 4

Applications

Dans le chapitre 3, nous avons réalisé une estimation de mouvement sur des séquences simulées. Nous pouvions alors comparer les performances de notre algorithme à celles des méthodes usuelles (corrélation, flot optique, etc.). Dans ce chapitre, nous présentons des résultats obtenus sur des séquences réelles acquises par une caméra en mouvement devant une scène majoritairement statique. Ces séquences sont plus complexes que les séquences simulées. Elles contiennent une multitude de phénomènes parasites, comme des variations de l'illumination globale, des occlusions et surtout des variations de profondeur dans la scène filmée. Tous ces paramètres rendent l'estimation de mouvement plus difficile et sont souvent des sources d'erreurs pour les méthodes classiques d'estimation de mouvement. Nous tentons dans ce chapitre d'illustrer le fait que même dans des conditions s'éloignant considérablement des "hypothèses de fonctionnement", notre algorithme fournit des résultats satisfaisants. Pour visualiser les performances de notre algorithme, nous l'avons employé dans deux problématiques qui sont la création de mosaïque d'images et la stabilisation d'images.

Le premier paragraphe de ce chapitre propose d'évaluer de manière qualitative et visuelle les performances de notre algorithme d'estimation de mouvement apparent majoritaire. Nous traitons le cas de séquences d'images réelles avec et sans perturbations (variations de luminosité, occlusions ...). Le deuxième paragraphe présente une des applications principales de notre méthode : la création de mosaïques d'images. Après un bref état de l'art nous donnons des résultats obtenus dans différentes conditions d'utilisation. Dans le troisième paragraphe nous élargissons le champ des applications de notre algorithme à la stabilisation d'images. Finalement nous concluons sur l'ensemble de ces essais et sur l'apport de notre algorithme par rapport aux techniques classiques.

4.1 Estimation d'un mouvement réel

4.1.1 Effet des variations de profondeur de champ

Dans le chapitre 3, les expérimentations réalisées étaient uniquement des simulations de mouvements plans. Rappelons que ces expérimentations avaient pour but l'évaluation des performances de notre algorithme et de les comparer avec quelques méthodes usuelles d'estimation de mouvement. Dans ce paragraphe, nous présentons des résultats obtenus pour l'estimation de mouvements réels. L'hypothèse que nous avons employée pour l'estimation de mouvement à savoir : la scène filmée est vue de suffisamment loin pour être assimilée à un plan parallèle au plan image (*i.e.* les variations de profondeur de la scène sont quasiment nulles) n'est plus rigoureusement respectée. Quelle sera la répercussion de ce non respect des hypothèses sur l'estimation de mouvement ? C'est ce que nous illustrons dans la suite de ce paragraphe.

En fait, en supposant que l'on puisse représenter la scène par une succession de plans situés à différentes profondeurs par rapport à la caméra, le mouvement que nous allons estimer correspond au mouvement du plan le plus visible dans l'image. En effet, tous les points appartenant à un plan parallèle au plan sagittal se déplacent de la même façon sous l'hypothèse que le mouvement est plan. L'estimation du mouvement apparent dominant donnera le mouvement des points appartenant au plan comprenant un maximum de points de l'image.

Lorsque l'on ne connaît pas le mouvement réel effectué par la caméra, la seule façon de juger si l'estimation du mouvement est correcte est purement qualitative. La méthode de visualisation de la qualité de l'estimation que nous avons choisie consiste en la superposition des images dont on a estimé le mouvement relatif en associant un effet de transparence à l'image du dessus. Les images sont superposées en utilisant le mouvement estimé. Si l'estimation de la transformation est correcte, les deux images sembleront bien ajustées. Sinon, l'image résultante présentera des défauts d'alignement. La précision de l'estimation est alors uniquement évaluée visuellement.

Prenons par exemple les deux images présentées sur la figure 4.1. Ces deux images sont issues d'une séquence où une caméra se déplace en filmant un bloc de maisons avec un parterre de fleurs et un arbre en premier plan. Le mouvement est principalement composé d'une translation horizontale.

Puisque tous les éléments de l'image n'appartiennent pas à un même plan, ils possèdent tous des mouvements différents. En l'occurrence, l'arbre au premier plan semble se déplacer plus vite que les maisons qui sont en arrière plan. L'estimation du mouvement apparent dominant à l'aide de notre algorithme donne : $T_{hor} = 10.3$ pixels, $T_{ver} = 1.7$ pixels et la rotation $R_\phi = 0.5^\circ$.

Quelques explications sont alors nécessaires. Il est évident que l'arbre (en surface occupée sur l'image) est minoritaire, ainsi l'estimation du mouvement apparent majoritaire ne correspondra pas au mouvement de l'arbre. Ceci se traduit par un mauvais recalage de l'arbre entre les deux images. Nous avons représenté sur la figure 4.2 la



Figure 4.1 – Images de la séquence maison : (a) Image 1, (b) Image 2.

superposition des deux images décalées du mouvement estimé. Par transparence, on voit bien que l'arbre des deux images n'est pas bien positionné. Au contraire, la partie arrière de l'image comprenant les maisons et une portion des fleurs (pouvant être assimilée à un plan car elle présente très peu de variations de profondeur), correspond à la zone la plus grande dans l'image. L'algorithme désigne donc le mouvement de cette partie comme le mouvement apparent majoritaire. Comme on peut le voir sur la figure 4.2 les maisons sont bien recalées.



Figure 4.2 – Superposition des deux images décalées du mouvement estimé ($T_{hor} = 10.7$; $T_{ver} = 1.7$; $R_{\phi} = 0.5^{\circ}$)

Le deuxième exemple que nous présentons contient une rotation plus importante. La figure 4.3 représente les deux images entre lesquelles nous avons essayé d'identifier la transformation. Il s'agit d'une vue de l'intérieur du laboratoire. L'intérêt de ces images repose sur le fait qu'entre les objets au premier plan et ceux au fond, la différence de

profondeur est grande par rapport à la distance entre la scène et la caméra. Tous les objets n'auront donc pas le même déplacement.



Figure 4.3 – Images du laboratoire présentant de fortes variations de profondeur : (a) Image 1, (b) Image 2.

La figure 4.4 illustre le recalage effectué après estimation du mouvement à l'aide de notre algorithme.



Figure 4.4 – Superposition des deux images décalées.

Comme on peut le voir, malgré les variations de profondeur, le recalage des deux images est majoritairement correct. L'écran d'ordinateur au premier plan n'est pas correctement mis en correspondance, mais cela est dû au fait que son mouvement n'est pas le mouvement dominant.

Les deux exemples présentés illustrent bien le comportement de notre algorithme

vis-à-vis de la non-planéité de la scène. Il assure une mise en correspondance robuste des zones de la scène correspondant au mouvement majoritaire identifié.

4.1.2 Effet des variations d'intensité globale

La deuxième expérience consiste à réaliser un déplacement entre deux prises de vue accompagné d'une variation de luminosité sur la deuxième image. Nous avons pris plusieurs images de scènes en intérieur. Au cours de l'acquisition de la séquence d'images, nous avons éteint certaines sources d'éclairage de la pièce. Ceci se traduit par une fluctuation parasite des niveaux de gris des pixels. La figure 4.5 présente deux images issues de cette séquence pour lesquelles nous donnons les résultats d'estimation du mouvement dominant.



Figure 4.5 – Images du laboratoire présentant de fortes variations de profondeur ainsi qu'une variation de luminosité ambiante : (a) Image 1, (b) Image 2.

L'intensité lumineuse moyenne de la seconde image est nettement inférieure à celle de la première. Contrairement aux expérimentations réalisées en théorie, la baisse de luminosité n'est pas identique en tout point de l'image. Tous les pixels ne subissent en effet pas la même variation de niveau de gris. On peut voir par exemple dans l'encadrement de la porte en haut à gauche de l'image qu'au lieu de diminuer, l'intensité lumineuse semble augmenter. Ce phénomène est très certainement dû à une adaptation automatique de l'ouverture du diaphragme de la caméra. De plus, la scène filmée n'est pas rigoureusement plane. Ainsi la méthode doit être robuste aux deux types de perturbations en même temps à savoir : scène non plane et variation de l'intensité des pixels entre deux images de la séquence.

La figure 4.6 montre la superposition des deux images positionnées conformément au mouvement estimé par notre algorithme. Malgré la forte variation de luminosité, le recalage est globalement bien effectué.

L'exemple suivant est issu d'une séquence d'images acquise à l'extérieur du laboratoire. Il s'agit d'un mur recouvert de pierres de taille. Là encore, les conditions



Figure 4.6 – Superposition des deux images décallées avec la variation de luminosité.

d'éclairage en extérieur, induisent des changements de luminosité ambiante au cours de la séquence. La figure 4.7 présente deux images extraites de cette séquence pour lesquelles la profondeur varie peu. Par contre, on peut voir entre ces deux images une variation d'intensité provoquée par le passage d'un nuage.

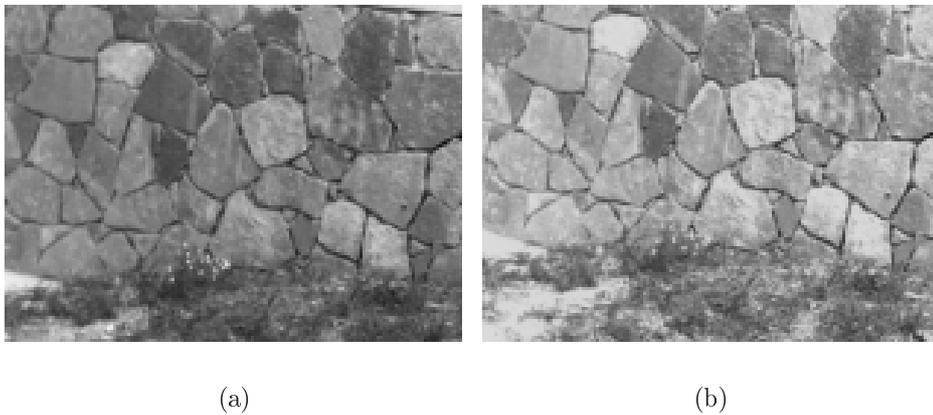


Figure 4.7 – Images issues d'une séquence prise en extérieur présentant des variations de luminosité ambiante : (a) Image 1, (b) Image 2.

L'estimation du mouvement apparent majoritaire à l'aide de notre algorithme nous permet de recalibrer les deux images l'une par rapport à l'autre comme présenté sur la figure 4.8.

On peut voir que, malgré la variation de luminosité, les deux images s'alignent correctement. Le fait que le mur soit une surface quasi-plane élimine les problèmes liés aux fluctuations de profondeur de champ. Ainsi, tous les points de l'image se déplacent quasi-identiquement et les deux images sont donc bien recalées.



Figure 4.8 – Superposition des deux images du mur avec variation de luminosité.

De plus la caméra étant manipulée à la main, le mouvement effectué n'est pas exactement composé de deux translations dans le plan sagittal et d'une rotation autour de l'axe optique. Le bon comportement de la méthode d'estimation de mouvement dans ces conditions illustre sa robustesse vis-à-vis du modèle de mouvement employé.

4.1.3 Effet des occultations et mouvements parasites

Lorsque des objets n'appartenant pas à la scène statique filmée passent devant l'objectif de la caméra, cela a pour effet de générer des occultations, c'est-à-dire la disparition temporaire d'une partie de la scène. Nous allons illustrer, à l'aide d'un exemple, la robustesse de notre algorithme vis-à-vis de telles perturbations (dans la limite du raisonnable).

La séquence présentée sur la figure 4.9 a été réalisée en intérieur. Le mouvement de la caméra est majoritairement composé d'une translation vers la droite. Une personne passe devant l'objectif de la caméra durant la séquence provoquant ainsi la disparition temporaire d'éléments de la scène.

Puisque le déplacement de la personne n'est pas majoritaire dans l'image, le mouvement estimé est celui de la caméra (ou plutôt de l'environnement statique). On peut voir sur la figure 4.10 le recalage des quatre images. Grâce à un effet de transparence, on peut voir le "fantôme" de la personne. Son image est présente quatre fois dans la séquence qui correspond aux quatre positions qu'elle occupait lors de l'acquisition des images. L'image de l'environnement reste stable et se superpose bien tout au long de la séquence.

Si l'objet se déplaçant devant la caméra devient prédominant en termes de surface projetée dans l'image, alors l'algorithme estime le déplacement de l'objet. Ceci est illustré sur la figure 4.11. La caméra est statique et seul le livre se déplace. Comme

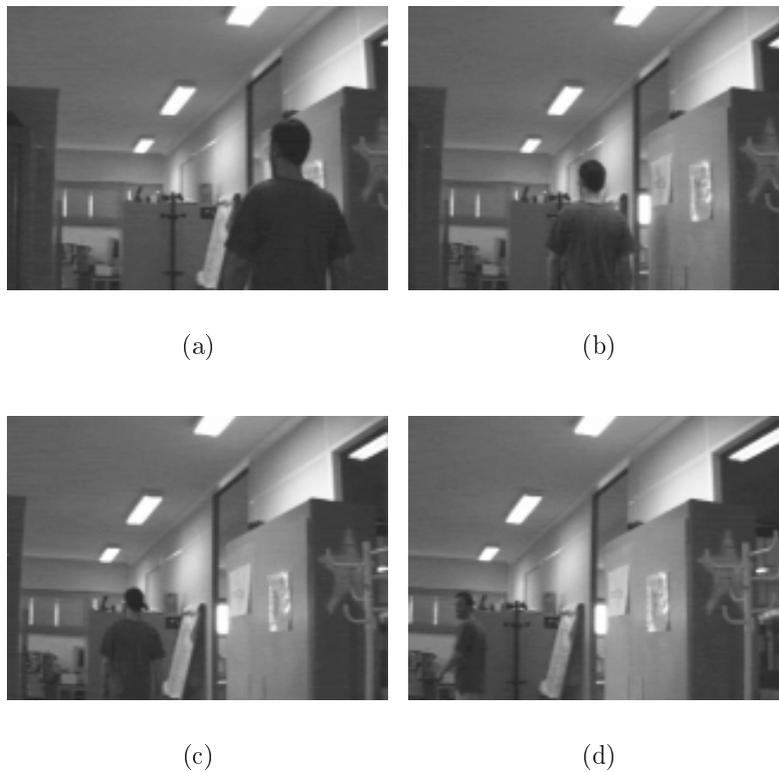


Figure 4.9 – 4 Images issues d’une séquence prise en intérieur avec présence d’un mouvement parasite.



Figure 4.10 – Superposition des quatre images avec effet de transparence pour indiquer la position de la personne tout au long de la séquence.

celui-ci est prédominant sur l’image, le mouvement estimé correspond à celui du livre et non à celui de la scène.

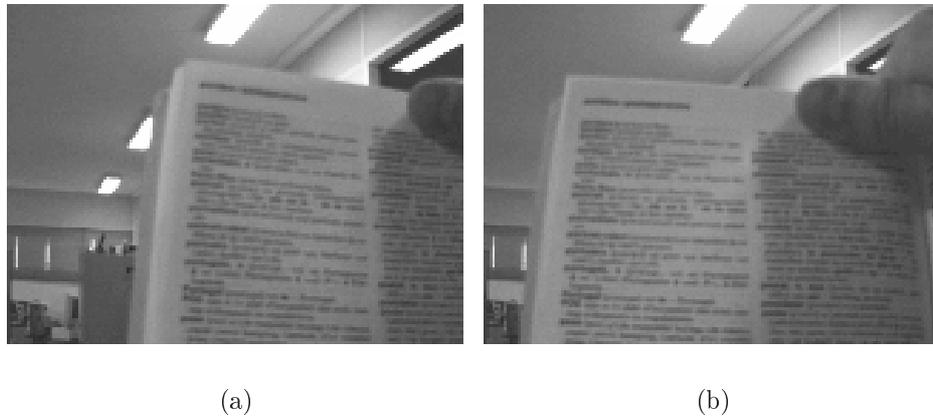


Figure 4.11 – Séquence présentant un mouvement parasite majoritaire (mouvement du livre).

Comme l'illustre la figure 4.12, le recalage du livre est bien effectué ; cependant le reste de la scène ne l'est pas. En effet, on peut voir deux lampes au plafond et deux coins de fenêtre en haut à droite de la figure 4.12 alors qu'il n'y en a qu'une sur les images 4.11 a) et b).

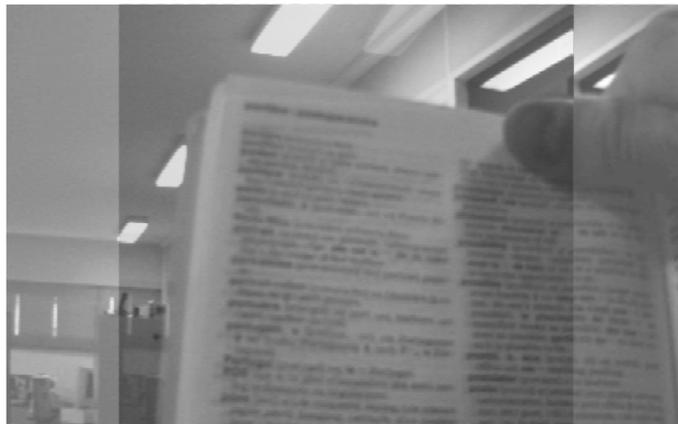


Figure 4.12 – Mise en correspondance des deux images avec le mouvement du livre comme mouvement dominant.

Ce résultat n'est pas réellement une limite de l'algorithme. Le mouvement estimé correspond bien au mouvement majoritaire dans l'image. Cependant, ce mouvement majoritaire n'est pas celui de la caméra. Si on veut que le mouvement estimé soit en rapport avec le déplacement de la caméra, il faut que les objets se déplaçant devant l'objectif restent minoritaires (en nombre de pixels) sur l'image.

4.2 Mosaïques d'images

Lorsque l'on prend une photo d'une scène, la taille de la zone de l'environnement représentée sur l'image est limitée par le champ de vue de l'appareil (fonction des caractéristiques de l'objectif, du diaphragme et de la surface photosensible). Pour capturer une image d'une scène plus large, il suffit de prendre une optique avec une distance focale plus petite. Cependant on a une perte de résolution qui peut être pénalisante pour certaines applications. Plusieurs approches existent pour palier cette réduction du champ d'observation. La première consiste à utiliser des caméras omnidirectionnelles, c'est-à-dire possédant un dispositif optique utilisant des miroirs ou des lentilles déformantes. Elles permettent d'obtenir des vues cylindriques ou sphériques de l'environnement. Cependant l'augmentation du champ visuel se fait encore au détriment de la précision du rendu de l'image. La deuxième approche consiste à prendre plusieurs clichés et à les réajuster les uns par rapport aux autres pour obtenir une image de plus grande taille. C'est le principe de la mosaïque d'images [IA98]. L'avantage d'un tel procédé est qu'il n'est pas nécessaire de posséder de matériel particulier, un simple appareil photo ou une caméra sont suffisants. D'autres techniques combinent les deux méthodes précédentes afin de créer une représentation de la scène filmée sous un angle solide d'environ 4π stéradians [NK00].

4.2.1 Etat de l'art

La création de mosaïques d'image est très en vogue depuis ces dix dernières années. Les champs d'applications sont variés et couvrent des domaines allant de l'imagerie médicale [SJW⁺97], à l'imagerie satellitaire en passant par l'exploration des fonds marins [GBCA01].

Généralement le processus de mosaïque d'images se décompose en deux étapes : l'estimation du mouvement et la fusion des images dans un référentiel unique. L'étape d'estimation de mouvement vise à estimer la transformation dans un repère "local" entre les deux images. Les méthodes d'estimation du mouvement exploitables ont déjà été traitées dans le chapitre 1. L'étape de fusion des images consiste à positionner et à afficher toutes les images dans un même repère général. Pour l'affichage, plusieurs techniques existent [GSV98]. Si l'on considère la position d'un pixel (i, j) dans la mosaïque finale, il se peut que plusieurs pixels des images de la séquence soient superposés en (i, j) au cours du recalage de toutes les images. Dans ce cas quelle valeur prendre ? On peut voir ce choix comme celui d'un opérateur temporel. En effet les pixels se superposant en (i, j) dans le repère global (celui de la mosaïque) sont issus d'images prises à différents instants. L'opérateur prend donc en entrée tous les pixels des images de la séquence se positionnant en (i, j) aux différents instants d'acquisition. Les quatre techniques les plus courantes utilisent : soit la première valeur du pixel se positionnant en (i, j) ("use first"), soit la dernière valeur du pixel se positionnant en (i, j) ("use last"), soit une moyenne de l'intensité de tous les pixels se chevauchant en (i, j) , ou enfin la médiane de ces mêmes intensités. La méthode utilisant la moyenne permet

d'atténuer les variations temporelles parasites du niveau de gris des pixel, alors que la méthode basée sur la médiane permet de supprimer l'influence d'objets se déplaçant vite devant l'objectif (ces derniers n'étant présents qu'un bref instant sur l'image). Nous avons retenu la méthode "use last" car elle est la plus simple à implanter et la moins coûteuse en temps de calcul.

Suivant le type de mouvements de la caméra considéré, les techniques utilisées pour construire les mosaïques d'images diffèrent. Par exemple lorsque le mouvement de la caméra est plan (images satellite) le support recueillant les images est un plan [ZFD97], lorsque le mouvement est circulaire (mouvement panoramique), les images sont généralement assemblées ou plaquées sur un "collecteur" cylindrique. Dans notre cas, nous utiliserons un support plan pour projeter toutes les images.

Dans [PH97], les auteurs remarquent que les déformations dues à l'alignement des images sont moins importantes au centre des images que sur leurs bords. Ils proposent donc de n'utiliser qu'une "bande" rectangulaire du centre de l'image (strip) pour reconstruire la mosaïque. Une extension de ce principe est proposée dans [RPF97] et [PRAZ00] où les bandes sont toujours choisies au centre de l'image mais dont les limites ne sont plus rectangulaires mais perpendiculaires à la direction du flot optique. Ceci a pour but de minimiser les déformations provoquées par des mouvements comprenant des zooms et des effets de parallaxe.

4.2.2 Résultats obtenus

Nous présentons dans ce paragraphe quelques résultats que nous avons obtenus en matière de création de mosaïques d'images. Pour obtenir une mosaïque, il faut être capable de replacer toutes les images d'une séquence dans un même repère. Nous allons donc, dans un premier temps, voir comment retrouver, à partir d'une succession de transformations de repères, la transformation globale permettant de replacer l'image (i) dans l'image globale (la mosaïque). Il s'agit en fait d'une relation de récurrence. Considérons l'équation (3.10) :

$$P_2 = (\mathbb{I}_2 - R_{\phi_1}^2) P_C + R_{\phi_1}^2 P_1 - R_{\phi_1}^2 T_1^2$$

Celle-ci permet de passer du repère attaché à l'image 1 au repère de l'image 2. Si l'on répète cette opération pour passer du repère 2 au repère 3 on a :

$$P_3 = (\mathbb{I}_2 - R_{\phi_2}^3) P_C + R_{\phi_2}^3 P_2 - R_{\phi_2}^3 T_2^3$$

Si l'on remplace P_2 par son expression et que l'on regroupe les termes on obtient :

$$P_3 = (\mathbb{I}_2 - R_{\phi_2}^3 R_{\phi_1}^2) P_C + R_{\phi_2}^3 R_{\phi_1}^2 P_1 - R_{\phi_2}^3 (R_{\phi_1}^2 T_1^2 + T_2^3) \quad (4.1)$$

Ce raisonnement se généralise aisément en posant :

$$P_i = (\mathbb{I}_2 - A_i) P_C + A_i P_2 - B_i \quad (4.2)$$

avec la relation de récurrence :

$$A_{i+1} = R_{\phi_i}^{i+1} A_i \quad (4.3)$$

$$B_{i+1} = R_{\phi_i}^{i+1} (B_i + T_i^{i+1}) \quad (4.4)$$

Une fois les transformations globales calculées, on est capable de placer toutes les images dans un même repère. C'est ce que nous avons illustré à l'aide de plusieurs exemples.

La première mosaïque présente une séquence fréquemment employée dans la communauté du traitement de l'image. Elle est constituée de 150 images. La scène est prise d'un véhicule se déplaçant vers la gauche. Au premier plan on distingue un arbre et en arrière plan un champ de fleurs et quelques maisons. La figure 4.13 présente les images 1, 10, 50, 90 et 130 de la séquence.

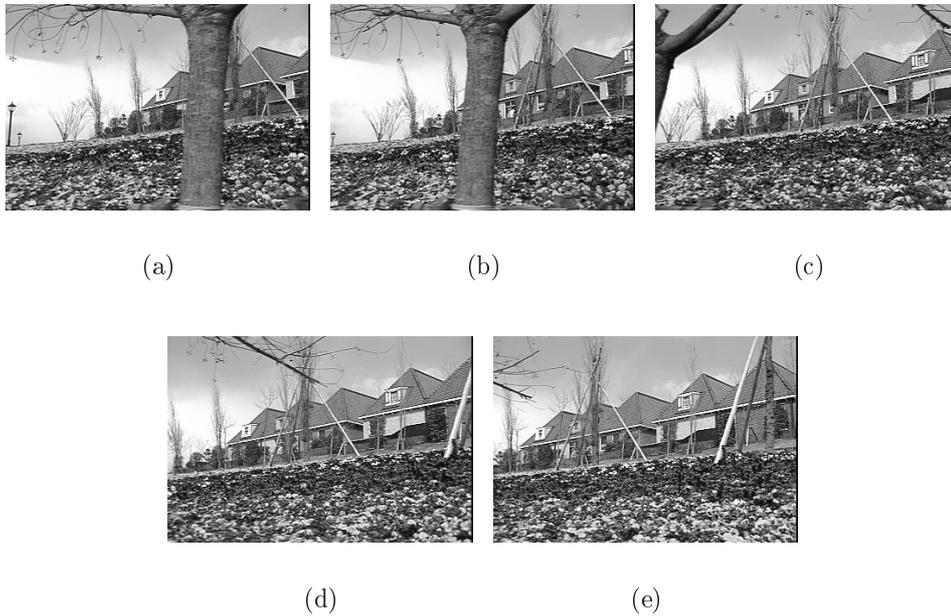


Figure 4.13 – Images de la séquence des maisons et du champs de fleurs.

L'image 4.14 illustre la mosaïque construite à partir de cette séquence. On peut voir, sur la mosaïque finale, que l'arbre au premier plan est déformé. Ceci est dû au fait que son mouvement ne correspond pas au mouvement majoritaire. Ainsi son image sera mal recalée tout au long de la séquence. Le fait que son image soit écrasée à gauche de la mosaïque vient du fait que nous utilisons une technique "use last" pour assembler les images.



Figure 4.14 – Mosaïque des 150 images de la séquence.

Le deuxième exemple que nous présentons est une séquence prise lors d'une des expéditions d'IFREMER sur l'épave du Titanic. Elle a été obtenue en réalisant un travelling horizontal devant la coque du bateau. La figure 4.15 représente quelques images de la séquence. Comme on peut le voir les conditions d'éclairage sont très mauvaises (4023m de profondeur !!). De plus la source de lumière, une lampe, est artificielle et crée un halo lumineux autour du centre de l'image. La position de ce halo ne se déplace pas dans l'image : il se situe toujours approximativement au centre de l'image. Les méthodes basées sur une extraction de points caractéristiques peuvent se tromper et prendre les points de la frontière de la zone éclairée pour estimer le mouvement ce qui reviendrait à estimer un mouvement biaisé tout au long de la séquence.

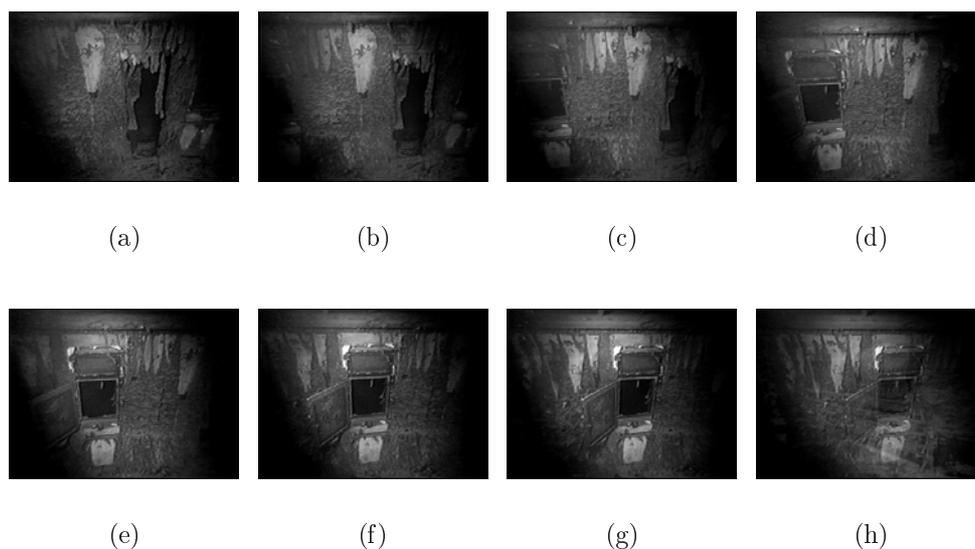


Figure 4.15 – Images d'une séquence visualisant la coque du Titanic sous de mauvaises conditions d'éclairage [IFREMER].

Nous avons donc estimé le mouvement entre les différentes images, et nous les avons

recalées pour former une mosaïque. Celle-ci est présentée sur la figure 4.16. Pour que la mosaïque soit représentative, nous avons tronqué une partie à gauche et à droite de chaque image afin de supprimer la zone sombre. Sans quoi l'utilisation de la méthode "use last" pour créer la mosaïque aurait produit une image quasiment entièrement noire.



Figure 4.16 – Mosaïque des 326 images de la séquence Titanic Wall.

Cette mosaïque illustre encore une fois le bon comportement de notre algorithme dans de mauvaises conditions d'éclairage et en présence de variations d'intensité locales.

La dernière mosaïque présentée est issue d'une séquence d'un fond marin comprenant un champ d'amphores. La présence d'algues ondulant au gré des courants provoque des mouvements parasites par rapport au déplacement de la caméra. Le passage d'un poisson sur quelques images de la séquence provoque lui aussi un mouvement parasite. Toutes ces difficultés s'ajoutent aux piètres conditions d'éclairage propres aux séquences sous-marines. De plus, le mouvement du plongeur tenant la caméra n'est pas bien modélisé par un mouvement plan (deux translations et une rotation autour de l'axe optique). Le mouvement effectué relève plutôt d'un mouvement panoramique avec la caméra inclinée vers le bas et des effets de perspective. Nous pourrions donc visualiser par le biais de la mosaïque la robustesse de l'estimation par rapport aux hypothèses sur le modèle de mouvement. La figure 4.18 illustre la mosaïque construite à partir de cette séquence.

On peut voir sur la mosaïque, que les mouvements parasites ne gênent pas l'estimation de mouvement apparent dominant. La mosaïque reste cohérente malgré ces perturbations. Cependant on voit apparaître sur la mosaïque des déformations sur certains objets. Ces déformations sont liées aux effets de parallaxe. En effet, les variations de profondeur ne sont pas négligeables par rapport à la distance séparant la scène de la caméra. Le fait que le mouvement de la caméra ne soit pas exactement un mouvement plan amplifie d'autant plus ces effets. Puisque le modèle de mouvement pour l'estimation, et a fortiori la construction de la mosaïque, ne tiennent pas compte de

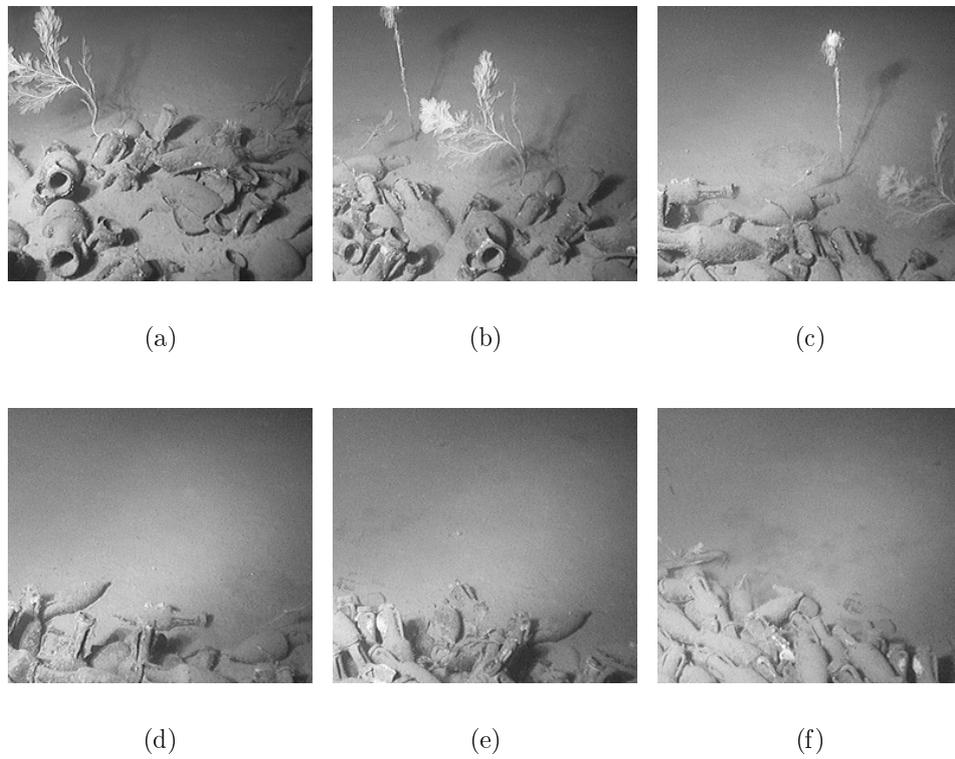


Figure 4.17 – Images de la séquence du champ d'amphores.

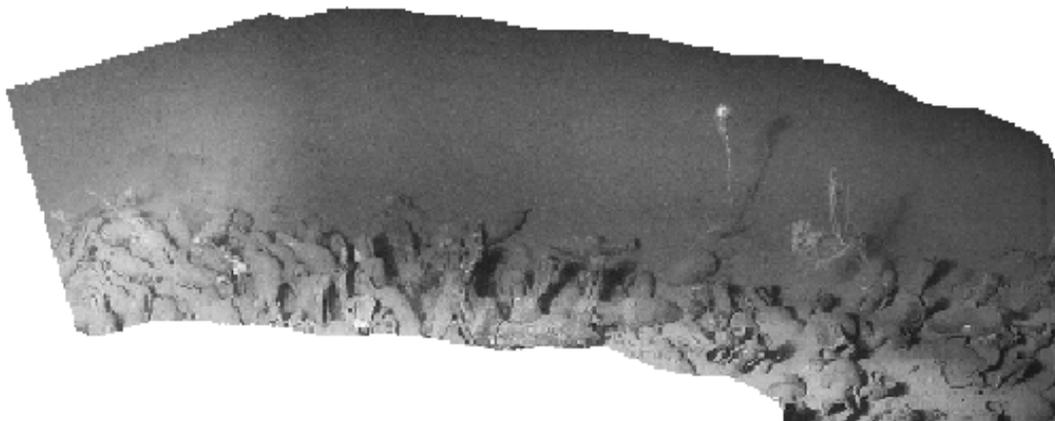


Figure 4.18 – Mosaïque des 300 images de la séquence du champ d'amphores.

ces perturbations, il est normal que l'image apparaisse légèrement déformée. On notera cependant que lorsque ces effets restent relativement faibles (début de la séquence), l'aspect de la mosaïque est tout à fait cohérent.

4.3 Stabilisation d'images

4.3.1 Quel est le principe ?

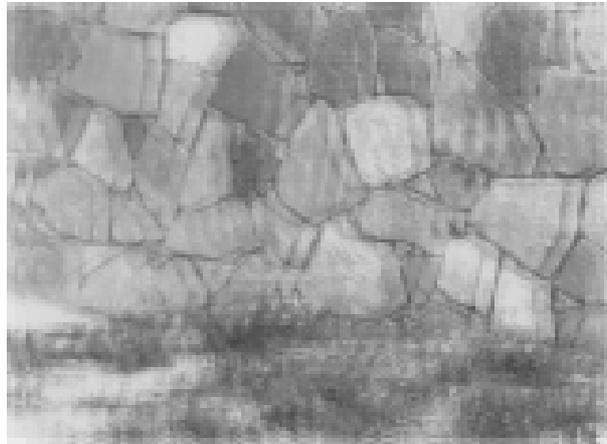
Il est fréquent lors la visualisation de séquences vidéo, que les images ne soient pas stables, c'est-à-dire qu'elles "vibrent" en quelque sorte. Ce manque de stabilité peut avoir différentes origines comme par exemple : le manque de stabilité ou les mouvements parasites du porteur de la caméra (opérateur humain ou véhicule).

Il résulte de ce manque de stabilité de l'image une gêne pour la visualisation du film. Le but de la stabilisation d'images est de supprimer ces vibrations parasites et ainsi de rendre plus "confortable" la visualisation du film. Afin de compenser les mouvements parasites, une estimation des différentes composantes du mouvement apparent est nécessaire. Considérons une séquence vidéo où l'on filme une scène statique en tremblant. On peut considérer qu'une image de la séquence n'est qu'une fenêtre au travers de laquelle le spectateur regarde la scène filmée. Sans stabilisation, le tremblement dans la séquence vidéo se traduira par une fenêtre de visualisation fixe et l'image qui vibre à l'intérieur. Avec stabilisation, l'effet sera inversé. En effet, l'image restera fixe et c'est le contour de la fenêtre qui vibrera. Comme l'attention du spectateur est généralement portée sur le contenu de l'image et non sur son contour, la visualisation sera plus confortable. Ceci est illustré sur la figure 4.19.

Différentes approches sont envisageables pour stabiliser les images d'une séquence vidéo. On les regroupe en deux catégories : les approches matérielles et les approches logicielles.

Les solutions matérielles utilisent soit des techniques d'asservissement à partir de mesures de rotation effectuées par des capteurs de type gyroscope, soit des systèmes mécaniques de type amortisseurs / ressorts. Il est possible de stabiliser soit le capteur CCD de la caméra, soit une plate-forme sur laquelle est fixée rigidement la caméra. La stabilisation du capteur CCD est généralement intégrée par le constructeur. Il existe de nombreux caméscopes grand public qui proposent une stabilisation de l'image. L'autre solution, externe, est de placer la caméra sur une plate-forme stabilisée à l'aide de capteurs d'accélération et d'actionneurs. L'étude d'un tel système a été menée par des étudiants néerlandais de l'université de Twente [BBV⁺95]. Les problèmes d'une plate-forme stabilisée sont dans un premier temps son coût prohibitif et en second son manque de portabilité. Il existe cependant des solutions commercialisées (steady cam), notamment utilisées par les media pour les reportages télévisés où le cameraman doit se déplacer en courant.

Ces inconvénients font naturellement pencher la balance vers la deuxième approche : la solution logicielle. Le fait de pouvoir stabiliser des séquences déjà tournées impose aussi l'utilisation d'une approche logicielle. Les algorithmes de stabilisation d'images comportent généralement trois phases : la première traite la séquence d'images et estime le mouvement de la caméra, la deuxième vise à compenser les mouvements gênants (vibrations), et la troisième gère l'affichage des images [MC98].



(a)



(b)

Figure 4.19 – a) Séquence instable et b) séquence stabilisée.

Les algorithmes d'estimation de mouvement ont été présentés dans le chapitre 1. On trouve majoritairement dans la littérature des algorithmes utilisant une mise en correspondance de points d'intérêts [CFR99] [MC95]. [MC98] propose une étude comparative de différents modèles de déplacement et de leurs performances respectives.

Certaines techniques se contentent alors de recalculer les images dans un repère de référence [MC95] [CFR99], alors que d'autres, dans le cas de mouvements hautement bruités séparent le mouvement apparent en deux composantes : le mouvement utile et le mouvement parasite [ZZyJ98]. Généralement le mouvement parasite est assimilé aux hautes fréquences (vibrations) alors que le mouvement utile correspond aux basses fréquences. Un filtrage passe-haut est donc utilisé pour extraire les hautes fréquences du signal du mouvement. Ces mouvements sont ensuite compensés en recalculant les images

dans un repère de référence en fonction des mouvements de hautes fréquences. La vidéo résultante laissera l'impression de mouvement tout en supprimant les instabilités.

4.3.2 Résultats obtenus

Dans un premier temps nous présentons une expérimentation illustrant le premier type de stabilisation d'images. Nous avons appelé cette stabilisation la stabilisation globale car tout le mouvement apparent est utilisé pour recalibrer les images dans un référentiel commun. Dans cette expérience, la première phase consiste à estimer le mouvement entre chaque image de la séquence. La transformation est ensuite exprimée dans le référentiel attaché à la première image. Sur la figure 4.20 on peut voir quelques images de la séquence des amphores. Sur la figure 4.21 le cadre blanc représente la position de la première image. L'image courante a été recalée dans le référentiel de la première image puis affichée.

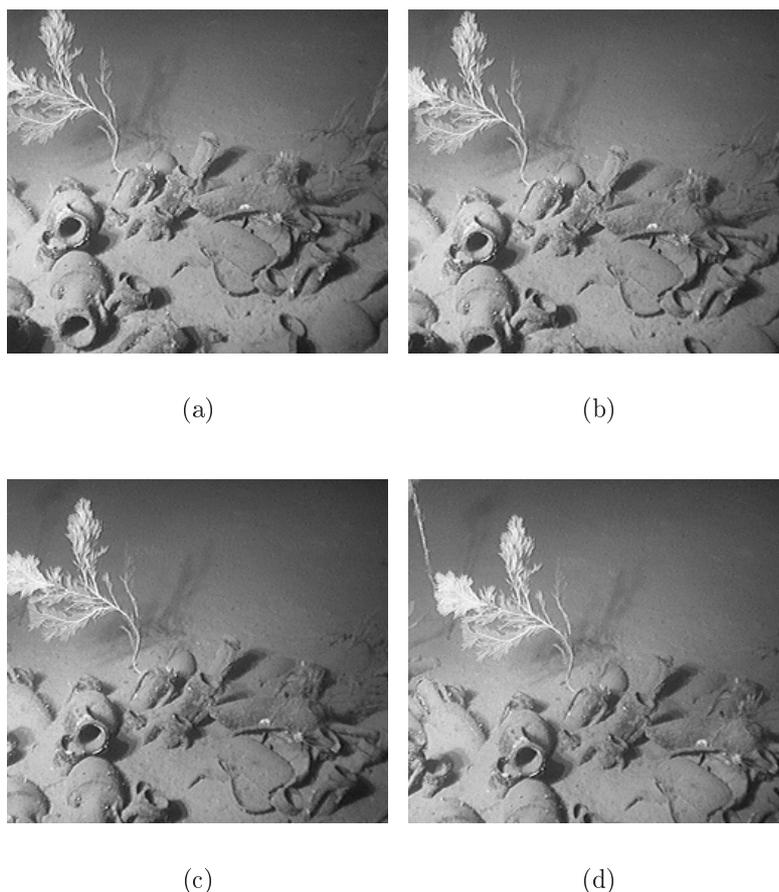


Figure 4.20 – Images de la séquence des amphores : n°1 (a), n°10 (b), n°20 (c), n°30 (d).

On peut voir sur la figure 4.21 que les différents motifs de l'image ne semblent pas changer de position. Cette expérience s'est confirmée sur d'autres séquences, on

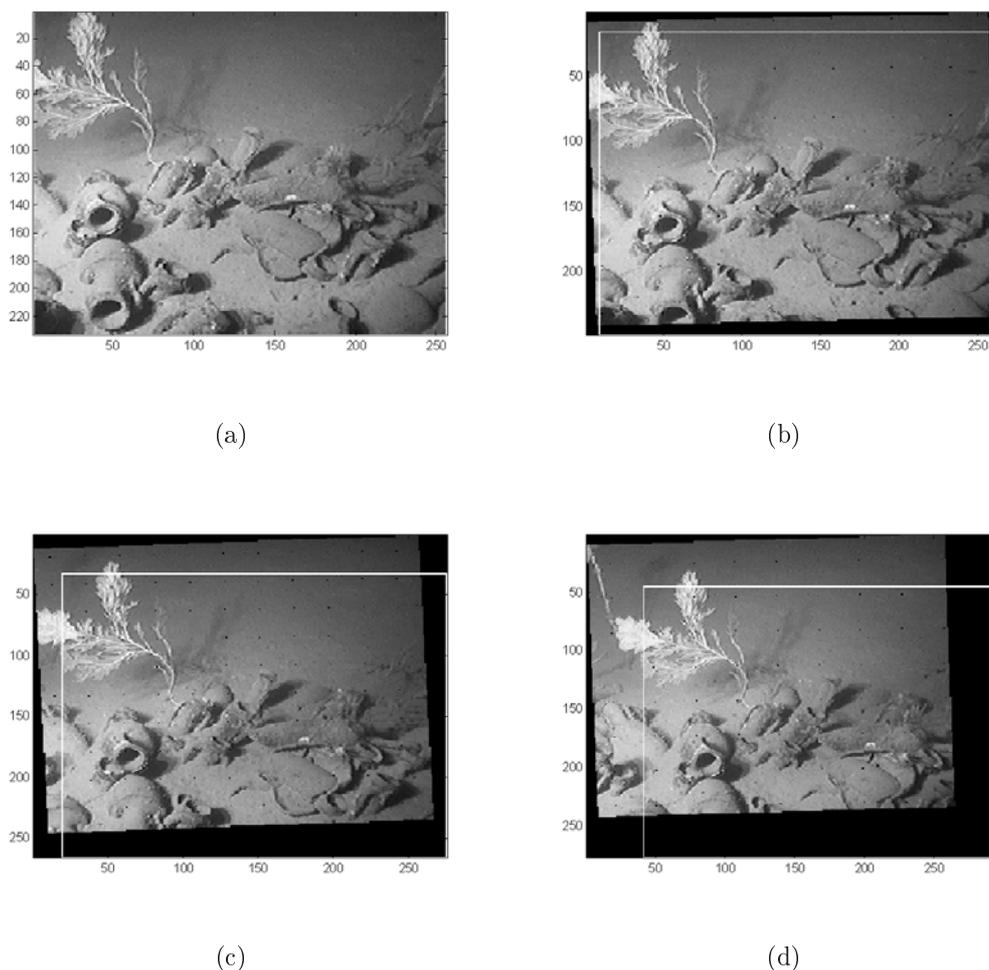


Figure 4.21 – Images stabilisées de la séquence des amphores : n°1 (a), n°10 (b), n°20 (c), n°30 (d).

peut donc affirmer que notre algorithme se prête bien à des problèmes de stabilisation d'images.

En ce qui concerne le deuxième cas de stabilisation d'images, que nous avons appelée stabilisation partielle, notre méthode d'estimation de mouvement a été employée pour améliorer la visualisation, en temps différé, de séquences d'images acquises par une caméra embarquée sur un hélicoptère. Cette étude a été réalisée dans le cadre du stage de DEA de Thomas Blum au LIRMM [Blu01]. La caméra est soumise à de fortes vibrations. Le but du travail était de proposer une méthode de filtrage permettant de compenser visuellement les mouvements de hautes fréquences dus à l'hélicoptère en utilisant une version simplifiée (temps réel) de notre estimateur de mouvement dominant. Il s'agissait donc d'estimer le mouvement, puis de séparer les composantes de hautes fréquences et de basses fréquences du mouvement. Plusieurs solutions ont été envisagées et comparées pour filtrer le mouvement. La première étude s'est portée sur

un modèle masse-ressort pour amortir les oscillations. Cependant la difficulté rencontrée pour gérer le coefficient de raideur des amortisseurs a fait abandonner cette solution. La deuxième solution retenue a été d'utiliser un filtre prédictif le "Guess Filter" [SA96]. Celui-ci présente l'avantage d'introduire un déphasage moindre par rapport au filtrage de Kalman, et offre de très bonnes performances en termes de robustesse, de précision et de fiabilité. La dernière approche a consisté à utiliser un filtre exponentiel. Son intérêt repose sur le fait qu'il est très rapide en temps d'exécution. Il permet ainsi de ne pas contraindre l'algorithme de stabilisation au niveau du temps de calcul. Tous les essais comparant ces trois approches sont présentés dans [Blu01].

Nous n'avons pas illustré ces expérimentations car il est difficile de rendre compte sur le papier des effets de la stabilisation vis-à-vis des oscillations. Néanmoins les expérimentations sont disponibles sous forme de séquences sur un CD-ROM.

4.4 Conclusion

Nous avons présenté quelques résultats d'expérimentations afin d'illustrer les performances de notre algorithme d'estimation de mouvement dans le cas d'images réelles. Ces expérimentations mettent en avant les performances de l'algorithme dans des conditions non-optimales d'utilisation. Il en est ressorti que cet algorithme se comportait bien dans le cas de fluctuations de l'illumination de la scène, d'occlusions ou de mouvements parasites présents sur la séquence vidéo. De même, sa robustesse vis-à-vis de l'inadéquation à l'hypothèse de mouvement plan a été illustrée. Cependant, afin de pouvoir utiliser de manière plus performante cette méthode d'estimation pour identifier des mouvement plus complexes, il faut augmenter le nombre de paramètres du modèle de mouvement.

Dans un deuxième temps nous avons présenté deux des applications principales de cet algorithme : les mosaïques d'images et la stabilisation d'images. Nous avons illustré au travers de quelques exemples les qualités de notre méthode pour la reconstruction de mosaïques dans le cas d'un mouvement assimilable à un mouvement plan. Si le mouvement s'éloigne trop de cette hypothèse, on voit alors apparaître des déformations sur les objets (cf. mosaïque sur la séquence des amphores). Pour prendre en considération des mouvements plus complexes, on est obligé, dans ce cas aussi, d'augmenter le nombre de paramètres du modèle de mouvement. Ceci se traduira inévitablement par un temps d'exécution de l'algorithme plus important.

Pour la stabilisation d'images, nous avons montré, par l'intermédiaire de quelques expérimentations, que notre algorithme donnait de bons résultats. Il reste cependant un point crucial dans le cas de la stabilisation avec filtrage du mouvement. En effet, comment doit on fixer la fréquence de coupure du filtre pour rejeter uniquement les mouvements parasites ? Une fréquence mal choisie se traduira inévitablement par un mauvais recalage et donc un mauvais affichage des images. Or le problème est que cette fréquence est difficile à déterminer. Elle dépend souvent de l'application et est généralement fixée suite à une série d'expérimentations.

Conclusion générale et perspectives

Avec les progrès réalisés durant les vingt dernières années en matière de miniaturisation, et surtout avec la baisse des prix des appareils vidéo, un grand nombre d'applications utilisant les caméras comme capteurs ont vu le jour. Celles-ci soulèvent sans cesse de nouvelles problématiques. Parmi les problèmes récurrents liés au traitement d'images, l'estimation du mouvement reste, encore de nos jours, un défi majeur.

Dans ce manuscrit, nous nous sommes plus particulièrement intéressés à l'estimation du mouvement de la projection des objets sur la rétine de la caméra, appelé couramment mouvement apparent. Ce mouvement apparent peut être induit soit par le mouvement propre des objets perçus, soit par le mouvement de la caméra. Dans ce dernier cas, sous certaines hypothèses, le mouvement projeté semble être uniforme donnant l'impression que l'on peut superposer les images issues de la séquence. Ce mouvement correspond en fait au mouvement apparent de la majorité des pixels de l'image. Nous l'appelons mouvement apparent majoritaire. Nous avons été amenés à étudier les différentes méthodes existantes permettant d'estimer le mouvement apparent (corrélation, flot optique, mise en correspondance de primitives). Nous avons tenté d'extraire les points forts, les points faibles et les limitations des approches classiques. Nous avons remarqué que la plupart des méthodes basaient leurs approches sur des hypothèses très contraignantes sur la relation entre les variations des niveaux de gris de pixels au mouvement apparent. Par exemple, pour les méthodes de type flot optique, on considère que l'illumination globale est conservée d'image en image. Les méthodes apparentées à la corrélation reposent sur l'hypothèse qu'un motif (ensemble de pixels) est conservé tout au long de la séquence. La violation de ces hypothèses compromet le bon fonctionnement de ces méthodes et en réduit donc la robustesse. Nous avons basé notre raisonnement sur une approche faisant appel aux théories des possibilités, des probabilités imprécises et des transferts de croyance. Cette approche nous a permis d'atténuer fortement l'influence des hypothèses sur le fonctionnement de notre estimateur. Ce type d'estimateur est à reclasser dans les démarches dites avec "a priori minimal"

Nous avons pour objectif de réaliser une méthode d'estimation de mouvement qui soit à la fois précise et robuste vis-à-vis des perturbations. Nous avons donc orienté nos travaux sur l'estimation du mouvement apparent majoritaire, tandis que la plupart des travaux précédemment cités s'articulaient autour d'une estimation du mouvement

moyen. L'estimation d'une quantité majoritaire nous a poussé à nous tourner vers des techniques permettant l'estimation de densités de probabilité. L'utilisation des histogrammes nous a alors semblée naturelle.

Les histogrammes classiques sont généralement employés comme outils de visualisation des densités de distributions statistiques plutôt que comme outils de calcul sur ces distributions. Ils ne permettent pas de prendre en compte les défauts liés à l'imprécision ou à l'incertitude des données. De plus, leur granularité et la façon dont sont fixées les bornes du partitionnement de l'espace influe considérablement sur les résultats de l'estimation que l'on peut en extraire.

C'est pourquoi, nous avons été amené à définir un nouvel outil statistique : les histogrammes approchés. Cet outil s'appuie sur la théorie des sous-ensembles flous grossiers et constitue une extension des histogrammes classiques. On peut noter deux différences notoires entre les histogrammes classiques et les histogrammes approchés. D'une part, un histogramme approché est construit sur une partition floue de l'espace des données. D'autre part, l'accumulateur associé à chaque cellule de la partition de l'espace est imprécis.

Grâce à l'utilisation d'un accumulateur imprécis, il est possible de prendre en compte l'imprécision des données d'entrée. Celle-ci se répercute sur l'imprécision de l'accumulation. On estime alors non pas la densité de probabilité mais un intervalle auquel appartient cette densité. On rejoint par cette approche la notion de probabilité imprécise.

La partition floue permet de réduire également la dualité *précision / incertitude* généralement associée aux techniques de vote. L'estimation de la densité de probabilité fait intervenir une interpolation "naturelle" améliorant sa précision sans pour autant augmenter son incertitude. Nous avons montré que cette estimation est moins dépendante du partitionnement que dans le cas des histogrammes classiques.

Nous avons testé les histogrammes approchés sur de nombreuses expérimentations. Nous avons reconstruit, à partir de données précises ou imprécises, l'allure de leurs densités de probabilité. Nous nous sommes attachés à tester le cas de variables issues de distributions mono-modales et multi-modales. Dans tous les cas, la reconstruction de la distribution est satisfaisante (au sens de la divergence de Kullback-Leibler).

Nous avons montré qu'il était possible de créer un estimateur de fractiles à partir des histogrammes approchés. Cet estimateur est particulièrement intéressant car plus rapide que les méthodes utilisant le tri des données, surtout dans le cas où l'estimateur de fractile doit être utilisé de façon récursive. Cet estimateur s'est avéré très précis.

Ce qui nous a intéressé principalement dans cet outil est la possibilité d'estimer le mode d'une distribution à partir de données imprécises. Nous avons aussi montré, sur des distributions statistiques simulées et réelles, que la précision de l'estimation du mode est plus fine que la granularité du partitionnement de l'histogramme. De plus cette estimation est moins biaisée, lorsque la distribution est multi-modale, que dans le cas des outils généralement employés pour l'estimation du mode (moyenne, médiane).

Malgré ces résultats encourageants, de nombreux travaux restent en suspens concer-

nant les histogrammes approchés. Dans un premier temps, une étude théorique plus approfondie de l'estimation de la probabilité d'un événement à partir des bornes supérieures et inférieures de cette probabilité s'impose. Nous utilisons actuellement une transformation permettant de passer de l'espace des accumulations dans l'espace des probabilités imprécises. Le mode est recherché comme étant l'élément maximisant la moyenne des deux bornes de la distribution de probabilité. Nous pensons que cette approche est insuffisante et mérite un travail théorique plus poussé.

D'autre part, de plus amples expérimentations sont nécessaires afin de mieux évaluer l'influence de la granularité du partitionnement flou sur la reconstruction de la densité de probabilité des données. Cette étude pourrait permettre, par exemple, d'établir un critère déterminant si la partition est en adéquation avec les données.

L'extension des histogrammes approchés à un espace de dimensions deux (et plus) n'est pas triviale. La technique de projection utilisée, pour ramener l'estimation du mode sur un espace de dimension N , à N estimations sur un espace de dimension 1, est de plus en plus complexe à mesure que N augmente. De plus, la complexité de l'implantation logicielle et le temps de calcul des algorithmes sont eux aussi dépendants de la dimension des histogrammes. Nous pensons que l'utilisation de méthodes plus appropriées [SH74] pour les aspects géométriques (liés aux calculs d'intersection de quantités floues dans le processus d'accumulation ou de restitution) nous permettra d'étendre à des dimensions plus élevées le principe des histogrammes approchés sans pour autant rendre prohibitif le temps de calcul.

L'estimation du mouvement majoritaire dans une séquence d'images s'appuie sur la théorie des histogrammes approchés présentés en première partie de ce mémoire. Notre approche consiste à associer un histogramme approché à l'espace des paramètres du mouvement entre deux images consécutives. A la fin du processus d'accumulation le mouvement apparent majoritaire est recherché comme étant le mode de la distribution sous-jacente des mouvements représentée par l'histogramme approché.

Pour réaliser l'accumulation, nous avons défini une nouvelle représentation de l'intensité des pixels à l'aide de deux classes duales floues. Chaque pixel est caractérisé par son appartenance aux classes de pixels blancs et noirs. Les cellules de l'histogramme approché correspondent chacune à un déplacement envisagé, c'est-à-dire à une combinaison des paramètres du modèle de mouvement. Un déplacement envisagé amène un pixel P_1 de l'image 1 en une position sur l'image 2 : P_2 . L'éventualité de déplacement est calculée de façon duale en fonction des appartenances aux classes de pixels blancs et noirs de P_1 et P_2 : si les deux pixels appartiennent à la même classe, P_1 votera pour le déplacement, sinon il votera contre. L'éventualité de déplacement de chaque pixel de l'image 1 (la valeur du vote) est alors accumulée dans l'histogramme approché, et ce pour tous les déplacements envisagés.

Cette approche a été validée par un grand nombre d'expérimentations sur des séquences générées artificiellement. Celles-ci nous ont permis d'évaluer les performances de notre algorithme en termes de précision et de robustesse. Nous avons pu rajouter des variations de luminosité globale ainsi que des mouvements parasites sur les séquences étudiées. Bien que les perturbations ne soient pas prises en compte par notre

modélisation, les performances de notre estimateur de mouvement sont très satisfaisantes en termes de précision. Les erreurs d'estimation sont largement inférieures en moyenne à la granularité de l'échantillonnage de l'espace paramétrique. De plus, cette méthode s'est montrée robuste vis-à-vis des variations de l'illumination globale ; cette robustesse est due à la modélisation en classes floues des niveaux de gris. Nous avons aussi constaté que la méthode était robuste vis-à-vis des occlusions et des mouvements parasites. Ceci est dû à notre méthode d'estimation du mode majoritaire de la densité de probabilité du mouvement. Les mouvements parasites sont apparentés à des modes secondaires et ne sont pas retenus pour la détermination du mouvement apparent dominant.

Afin de valider notre approche, nous avons mené des expérimentations sur des séquences d'images réelles (scènes intérieures et extérieures). Celles-ci permettent, entre autres, de tester les performances de l'algorithme dans des conditions d'exploitation réalistes. Nous avons ainsi pu montrer que notre algorithme est robuste vis-à-vis des perturbations telles que les variations d'illumination non-uniformes, la présence de mouvements parasites ou d'occlusions sur l'image (ou d'une combinaison de ces perturbations). Nous avons aussi pu montrer la robustesse de notre approche à l'égard du modèle de mouvement considéré. En effet, l'estimation du mouvement reste acceptable visuellement même si le mouvement de la caméra ne correspond pas rigoureusement au modèle de mouvement envisagé.

Nous avons appliqué notre algorithme à une tâche de construction de mosaïques d'images. Cet aspect de l'estimation de mouvement est souvent utilisé dans les processus de cartographie. Il vise à construire une carte de grande envergure, sorte de vue panoramique, à l'aide des images acquises par une seule caméra. Nous avons montré que l'estimation de mouvement apparent majoritaire se prêtait bien à ce type de problématique, et que notre algorithme donnait des résultats particulièrement satisfaisants.

Cependant, la technique employée pour construire la mosaïque ("use last") n'est pas forcément la plus appropriée. Il faudrait encore comparer les résultats obtenus avec les méthodes de construction de type médiane ou moyenne, notamment dans le cas des images sous-marines où les variations d'intensité liées à l'éclairage artificiel sont fréquentes.

Nous avons proposé une application de la version temps réel de notre algorithme à la stabilisation de séquences d'images. Cela consiste à filtrer les mouvements parasites visibles sur l'image (dus au tremblement du support de la caméra) lors de l'acquisition d'une séquence vidéo. Le mouvement apparent est donc estimé, puis les composantes de hautes et basses fréquences de ce mouvement sont séparées. La composante hautes fréquences du mouvement est affectée au cadre de l'image. L'effet obtenu est de rendre plus confortable la visualisation de la séquence. L'utilisation de cet algorithme peut permettre de restaurer des films anciens ou encore de stabiliser les films récents. Il reste cependant en suspens la question de la détermination automatique de la bande passante du filtre permettant de séparer les hautes et les basses fréquences. Le problème est assez complexe puisque les notions de hautes et basses fréquences dépendent grandement de

la nature du film.

Les perspectives de ce travail sont multiples. Le fait que la taille de l'espace paramétrique soit imposée est une limitation. Les mouvements dont l'amplitude est plus grande que la taille de l'espace paramétrique ne peuvent pas être estimés. Nous avons donc défini un processus d'estimation du mouvement apparent multi-échelle. Celui-ci utilise la représentation multi-résolution des images à l'aide des deux classes duales floues de pixels. Nous avons étudié les performances de ce processus multi-résolution. Nous nous sommes avant tout préoccupé de celles concernant le gain en temps de calcul, et nous avons montré que son utilisation devenait indispensable à mesure que la taille des images ou de l'espace paramétrique augmentait.

Malgré ses bonnes performances, l'aspect algorithmique de notre méthode nécessite de nombreuses améliorations. Le problème lié au temps d'exécution reste crucial. Nous envisageons, afin de réduire ce temps de calcul, de repenser l'algorithme en incluant un maximum de pré-calculs, dès que cela est possible. Une ébauche de ce travail, dans le cas simple d'un mouvement à deux translations, a été présentée dans le paragraphe intitulé "vers le temps réel". Il nous reste à généraliser cette approche, si c'est possible, pour des modèles paramétriques de mouvement plus complexes.

L'amélioration de la précision de l'estimation du mouvement est une des perspectives de notre travail. Nous avons été obligé de dégrader volontairement la précision des quantités floues intervenant dans les possibilités et nécessités conjointes (vote des pixels) pour des raisons de complexité de calculs. Une des pistes envisagées, afin de ne pas dégrader la précision des quantités floues et de garder une complexité des calculs raisonnable, est l'utilisation des méthodes de "polygon clipping" de Southerland-Hodgman.

Finalement, une voie qui nous semble intéressante à approfondir est l'extension aux images couleurs de notre processus d'estimation de mouvement. Nous pensons que l'ajout de la couleur peut permettre de lever certaines ambiguïtés au niveau des éventualités de déplacement des pixels et rendre ainsi l'estimation plus robuste. Cependant, le passage de deux classes de pixels (noir et blanc) à trois classes (rouge, vert, bleu) risque de compliquer les relations entre les classes. Les relations de type complémentaires $\mu_n = 1 - \mu_b$ ne sont plus applicables et les équations des éventualités de déplacement en seront rendues plus complexes.

Concernant l'aspect applicatif, des améliorations doivent être envisagées. Il semble indispensable d'augmenter le nombre de paramètres du modèle de mouvement afin de pouvoir utiliser l'algorithme dans des applications plus variées. Nous pensons intégrer au moins un facteur de changement d'échelle et peut être un facteur correspondant aux changements de perspectives. Cependant ceci doit s'accompagner d'une optimisation du temps d'exécution de l'algorithme. En effet, l'algorithme est déjà relativement lent ; le fait de rajouter des paramètres le ralentira certainement encore plus.

Bibliographie

- [AB86] H. Asada and M. Brady. The curvature primal sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1) :2–14, January 1986.
- [AG92] M. Abidi and R. Gonzalez. Data fusion in robotics and machine intelligence, academic press, 1992.
- [AG93] P. Ashwanden and W. Guggenbühl. Experimental results from a comparative study on correlation-type registration algorithms. *Robust Computer Vision*, pages 268–289, 1993.
- [Ana89] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2 :283–310, 1989.
- [Bas88] M. Basseville. Distance measures for signal processing and pattern recognition. Technical report, INRIA, Septembre 1988.
- [BBV⁺95] J.F. Blacquiere, J.J.P. Blijd, M.E. Visser, Supervision, M. Korsten, and L. Spreeuwers. Stabilization of a camera platform. Report : 003M95, Laboratory for Measurements and Instrumentation, University of Twente, January 1995.
- [BCSD97] A. Branca, G. Cicirelli, E. Stella, and A. Distanto. Mobile vehicle's egomotion estimation from time varying image sequences. In *Proceedings of the IEEE Int. Conf. on Robotics and Automation*, pages 1986–1991, April 1997.
- [Bea78] P. R. Beaudet. Rotationally invariant image operators. In *Proceedings of the 4th Int. Conf. on Pattern Recognition*, pages 579–583, November 1978.
- [Bei93] H. Beigi. An overview of handwriting recognition. In *Proceedings of the 1st Annual Conference on Technological Advancements in Developing Countries*, pages 30–46, July 1993.
- [Bez81] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York : Plenum, 1981.
- [BKG95] B. De Baets, E. Kerre, and M. Gupta. The fundamentals of fuzzy mathematical morphology : Part 1 and 2. *International Journal General Systems*, 23 :155–171 (part 1), 307–322 (part 2), 1995.

- [BKS⁺99] Carla E. Brodley, A. C. Kak, C. Shyu, J. Dy, L. S. Broderick, and A. M. Aisen. Content-based retrieval from medical image databases : A synergy of human interaction, machine learning and computer vision. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pages 760–767, July 1999.
- [Blo01] Isabelle Bloch. Traitement d’images. In B. Bouchon-Meunier, editor, *Utilisations de la logique floue*, chapter 9. Hermes, Paris, France, 2001.
- [Blu01] Thomas Blum. Rapport de DEA SyAM : Stabilisation d’images thermiques, Juillet 2001.
- [BM95] Isabelle Bloch and Henri Maitre. Fuzzy mathematical morphologies : A comparative study. *Pattern Recognition*, 28(9) :1341–1387, 1995.
- [BN96] Dinkar N. Bhat and Shree K. Nayar. Ordinal measures for visual correspondence. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 351–357, June 1996.
- [Bor00] Christian Borgelt. *Data Mining with Graphical Models*. PhD thesis, Otto-von-Guericke-Universität, Magdeburg, Juli 2000.
- [Bro92] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Survey*, 24(4) :325–376, 1992.
- [BT80] S. T. Barnard and W. B. Thompson. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4) :333–340, 1980.
- [CC89] M. Coster and J. L. Chermant. *Précis d’analyse d’images*. Presses du CNRS, 1989.
- [CFR99] Alberto Censi, Andrea Fusiello, and Vito Roberto. Image stabilization by features tracking. In *Proceedings of the 10th Int. Conf. on Image Analysis and Processing*, pages 665–670, September 1999.
- [CM91] J. M. Chassery and A. Montanvert. *Géométrie discrète en analyse d’images*. Hermes, 1991.
- [CS95] Claudette Cédras and Mubarak Shah. Motion-based recognition : A survey. *IVC*, 13(2) :129–155, March 1995.
- [CS96] Jin-Long Chen and George C. Stockman. Determining pose of 3D objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1) :52–57, January 1996.
- [CSA01] F. Comby, O. Strauss, and M.-J. Aldon. Possibility theory and rough histograms for motion estimation in a video sequence. In *Proceedings of the International Workshop on Visual Form IWVF4*, pages 473–483, Capri, Italy, May 28-30 2001.
- [CTS97] Michael M. Chang, A. Murat Tekalp, and M. Ibrahim Sezan. Simultaneous motion estimation and segmentation. *IEEE Transactions on Image Processing*, 6(9) :1326–1333, 1997.

- [Den97] T. Denoeux. Reasoning with imprecise belief structures. Technical Report 97/44, Universite de Technologie de Compiegne, Heudiasyc Laboratory, 1997.
- [DF90] R. Deriche and O. Faugeras. 2D curve matching using high curvature points. In *Proceedings of the Int. Conf. on Pattern Recognition*, pages 240–242, June 1990.
- [DKe97] al. D. Koller et. Automated camera calibration and 3d egomotion estimation for augmented reality applications. In *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, pages 199–205, September 1997.
- [DLZ95] C. Demko, P. Loonis, and E. Zahzah. Isomorphism of fuzzy structures : a new method for image classification. In *In Proceedings of the 9th Scandinavian Conference on Image Analysis Vol. 1*, pages 297–304, June 1995.
- [DP88] D. Dubois and H. Prade. *Possibility Theory An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.
- [DP90] D. Dubois and H. Prade. Rough fuzzy sets and fuzzy rough sets. *International journal of general systems*, 17(2-3) :191–200, 1990.
- [Fau93] O. Faugeras. *Three Dimensional Computer Vision : A Geometric Viewpoint*. MIT Press, 1993.
- [FB00] Ronan Fablet and Patrick Bouthemy. Statistical motion-based retrieval with partial query. In *Proceedings of the 4th Int. Conf. Visual Information Systems*, pages 96–107, October 2000.
- [FSA01] Cornelia Fermüller, David Shulman, and Yiannis Aloimonos. The statistics of optical flow. *Computer Vision and Image Understanding*, 82(1) :1–32, 2001.
- [GB00] Marc Gelgon and Patrick Bouthemy. A region-level graph labeling approach to motion-based segmentation. *Pattern Recognition*, 33(4) :725–740, April 2000.
- [GBCA01] R. Garcia, J. Batlle, X. Cufi, and J. Amat. Positioning an underwater vehicle through image mosaicking. In *Proceedings of the 2001 IEEE Int. Conf. on Robotics and Automation*, pages 2779–2784, May 2001.
- [GSV98] Nuno Gracias and José Santos-Victor. Automatic mosaic creation of the ocean floor. In *Proceedings of the IEEE OCEANS'98*, pages 257–262, September 1998.
- [HKR93] Daniel P. Huttenlocher, Gregory A. Klanderman, and William J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9) :850–863, October 1993.
- [HM95] Radu Horaud and Olivier Monga. *Vision par ordinateur, Outils fondamentaux*. Hermes, 1995.
- [HS81] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17 :185–203, 1981.

- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, August 1988.
- [IA98] Michal Irani and P. Anandan. Video indexing based on mosaic representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 86(5) :905–921, May 1998.
- [JBC97] M. Bennamoun J. Banks and P. Corke. Fast and robust stereo matching algorithms for mining automation. In *proceedings of JAF 97*, volume 1, pages 139–149, November 1997.
- [JBD93] P. T. Jackway, W. W. Boles, and M. Deriche. Morphological scale-space fingerprints and their use in 3-D object recognition. In *Proceedings of Digital Image Computing Techniques and Applications*, pages 382–389, December 1993.
- [JLBB94] D. J. Fleet J. L. Barron and S. S. Beauchemin. Systems and experiment : Performance of optical flow techniques. *International Journal of Computer Vision*, 12 :1 :43–77, 1994.
- [KK96] R. Krishnapuram and J. K. Keller. The possibilistic c-means algorithm : Insights and recommendations. *IEEE Trans. on Fuzzy systems*, 4(3) :385–393, 1996.
- [KKBMM98] O. Kosheleva, V. Kreinovich, B. Bouchon-Meunier, and R. Mesiar. Operations with fuzzy numbers explain heuristic methods in image processing. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98)*, pages 265–272, July 1998.
- [KN93] Włodzimierz Kasprzak and Heinrich Niemann. Visual motion estimation from image contour tracking. *Computer Analysis of Images and Patterns*, 719 :363–370, 1993.
- [KTB87] Joseph K. Kearney, William B. Thompson, and Daniel L. Boley. Optical flow estimation : An error analysis of gradient-based methods with local optimisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2) :229–244, March 1987.
- [LGJ94] J. Lay, J. Gauch, and J. Crisman. Computing optical flow in color image sequences. *Innovation and Technology in Biology and Medicine*, 15(3) :375–390, 1994.
- [Lju99] Lennart Ljung. *System Identification : Theory for the user*. PTR Prentice Hall Information and System Science Series, 1999.
- [LWKM95] Q. T. Luong, J. Weber, D. Koller, and J. Malik. An integrated stereo-based approach to automatic vehicle guidance. In *Proceedings of the Int. Conf. on Computer Vision*, pages 52–57, June 1995.
- [LZ99] Pascal Lepinay and Rene Zapata. Réalisation d'une carte de corrélation visuelle en temps réel. Rapport LIRMM numéro 6578, LIRMM, 1999.

- [MB96] Amar Mitiche and Patrick Bouthemy. Computation and analysis of image motion : A synopsis of current problems and methods. *International Journal of Computer Vision*, 19(1) :29–55, 1996.
- [MC95] Carlos Morimoto and Rama Chellappa. Fast electronic digital image stabilization. Technical report of cvl, University of Maryland, 1995.
- [MC98] Carlos Morimoto and Rama Chellappa. Evaluation of image stabilization algorithms. In *Proceedings of ICASSP*, pages 2789–2792, May 1998.
- [MN84] G. Medioni and R. Nevatia. Matching images using linear features. *IEEE Transactions on Pattern Matching and Machine Intelligence*, 6(6) :675–685, 1984.
- [Moo79] R. Moore. Methods and applications of interval analysis. In *Proceedings of SIAM Studies on Applied Mathematics*, volume 2, Philadelphia, 1979.
- [MP98] E. Mémin and P. Perez. Dense estimation and object oriented segmentation of the optical flow with robust techniques. *IEEE Transaction on Image Processing*, 7(5) :703–719, May 1998.
- [Neg98] Shahriar Negahdaripour. Revised definition of optical flow : Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9) :961–979, September 1998.
- [NK00] Shree Nayar and Amruta Karmarkar. 360 * 360 mosaics. In *Proceedings of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 388–395, June 2000.
- [OB95] Jean-Marc Odobez and Patrick Bouthemy. Robust multiresolution estimation of parametric motion models applied to complex scenes. *Journal of Visual Communication and Image Representation*, 6 :348–365, December 1995.
- [OC96] Marciej Orkisz and Patrick Clarysse. Estimation du flot optique en présence de discontinuités : une revue. *Traitement du signal*, 13(5) :489–510, 1996.
- [Paw98] Z. Pawlak. Granularity of knowledge, indiscernability and rough sets. In *Proceedings of FUZZIEEE'98 International Conference*, Anchorage, Alaska, USA, May 5-9 1998.
- [PFT94] David W. Paglieroni, Gary E. Ford, and Eric M. Tsujimoto. The position orientation masking approach to parametric search for template matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(7) :740–774, July 1994.
- [PG92] S. K. Pal and A. Ghosh. Fuzzy geometry in image analysis. *Fuzzy Sets and Systems*, 48 :23–40, 1992.
- [PH97] Shmuel Peleg and Joshua Herman. Panoramic mosaics by manifold projection. In *Proceeding of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 338–343, June 1997.

- [PMS94] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 84–91, June 1994.
- [PRAZ00] Shmuel Peleg, Benny Rousso, Alex Rav Acha, and Assaf Zomet. Mosaicing on adaptative manifolds. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10) :1144–1154, October 2000.
- [PRH00] A. M. Peacock, D. Renshaw, and J. Hannah. A fuzzy data fusion method for improved motion estimation. In *in Proceedings of European Signal Processing Conference*, pages 1485–1488, Tampere, Finland, September 2000.
- [PS98] Lech Polkowski and Andrzej Skowron. Towards adaptative calculus of granules. In *Proceedings of FUZZIEEE'98 International Conference*, pages 111–116, Anchorage, Alaska, USA, May 5-9 1998.
- [Roh92] K. Rohr. Modelling and identification of characteristic intensity variations. *Image and Vision Computing*, pages 66–76, February 1992.
- [Ros84] A. Rosenfeld. The fuzzy geometry of image subsets. *Pattern Recognition Letters*, 2 :311–317, 1984.
- [RPF97] Benny Rousso, Shmuel Peleg, and Ilan Finci. Mosaicing with generalized strips. In *DARPA Image Understanding Workshop*, pages 255–260, May 1997.
- [SA96] O. Strauss and E. Aimé. Filtering data by using three error theories together : the guess filter. In *Proceeding of EUFIT' 96*, September 1996.
- [SB97] S. M. Smith and J. M. Brady. Susan : A new approach to low level image processing. *Int. Journal of Computer Vision*, pages 45–78, 1997.
- [SCA00] O. Strauss, F. Comby, and M.-J. Aldon. Rough histograms for robust statistics. In *Proceedings of the International Conference on Pattern Recognition*, volume 2, pages 688–691. IAPR, September 2000.
- [Sch86] B.G. Schunck. Image flow continuity equations for motion and density. In *Proceedings of IEEE Workshop on Motion : Representation and Analysis*, pages 89–94, Charleston, South Carolina, May 1986. IEEE.
- [Sch96] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris*. PhD thesis, Institut National Polytechnique de Grenoble INPG, 1996.
- [Ser82] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, New york, 1982.
- [SH74] I. E. Southerland and G. W. Hodgman. Reentrant polygon clipping. *Graphics and Image Processing*, 17(1) :32–42, 1974.
- [Sha76] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, 1976.
- [Shu86] B. G. Shunck. The image flow constraint equation. *Computer Vision Graphics and Image Processing*, 35 :20–46, 1986.

- [Sin92] A. Singh. *Optical Flow Computation : a Unified Perspective*. IEEE Computer Society Press, 1992.
- [Sis00] Marc Sistiaga. *Navigation Référencée Images de Terrain Pour Engins Sous-Marins*. PhD thesis, Université de Montpellier II, Septembre 2000.
- [SJW⁺97] Vivek Swarnakar, Myoungki Jeong, Richard Wasserman, Eric Andres, and Darold Wobschall. An integrated distortion correction and reconstruction technique for digital mosaic mammography. *SPIE Medical Imaging : Image Display*, 3031 :673–681, 1997.
- [SK99] Cristoph Stiller and Janusz Konrad. Estimating motion in image sequences. a tutorial on modeling and computation of 2-D motion. *Signal Processing Magazine*, 16 :70–91, July 1999.
- [SL99] O. Strauss and E. Lavarec. Histogrammes approchés : application aux statistiques d'ordre. In *Proceedings des Rencontres Francophones sur la Logique Floue et ses Applications*, pages 161–168, Valenciennes, France, 21-22 Octobre 1999.
- [SLH00] N. Sebe, M. S. Lew, and D. P. Huijsmans. Toward improved ranking metrics. *Transaction on Pattern Analysis and Machine Intelligence*, 22(10) :1132–1143, October 2000.
- [Sme90] P. Smets. Constructing the pignistic probability function in a context of uncertainty. In M. Henrion, R. D. Schachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*, pages 29–39. Elsevier Science Publishers, North-Holland, Amsterdam, 1990.
- [Sme94] P. Smets. The transferable belief model. *Artificial Intelligence*, 66 :191–243, 1994.
- [Smi97] Stephen M. Smith. Reviews of optic flow, motion segmentation, edge finding and corner finding. Technical report TR97SMS1, Department of Clinical Neurology, Oxford University, UK, 1997.
- [THWL98] J-C. Tsai, C-H Hsieh, S-K. Weng, and M-F. Lai. Block-matching motion estimation using correlation search algorithm. *Signal Processing : Image Communication*, 13 :119–133, 1998.
- [Tis96] Massimo Tistarelli. Multiple constraints to compute optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12) :1243–1250, December 1996.
- [Vie96] R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Inc., 1996.
- [Wal91] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- [Wan97] Li-Xin Wang. *A course in fuzzy systems and control*. Prentice Hall PTR, 1997.
- [WM95] Joseph Weber and Jitendra Malik. Robust computation of optical flow in a multi-scale differential framework. *International Journal of Computer Vision*, 14 :67–81, 1995.

-
- [WW91] T. H. Wonnacott and R. J. Wonnacott. *Statistique : Economie - Gestion - Sciences - Médecine*. Economica, 1991.
- [ZALE99] N. Zahid, O. Abouelala, M. Limouri, and A. Essaid. Unsupervised fuzzy clustering. *Pattern Recognition Letters*, 20 :123–129, 1999.
- [ZFD97] I. Zoghlami, O. Faugeras, and R. Deriche. Using geometric corners to build a 2D mosaic from a set of images. In *Proceedings of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 420–425, June 1997.
- [Zha95] Zhengyou Zhang. Estimating motion and structure from correspondences of line segments between two perspective images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12) :1129–1139, December 1995.
- [ZT97] D. Ziou and S. Tabbone. Edge detection techniques - an overview. Technical report number 195, Département de Mathématique et Informatique, Université de Sherbrooke, 1997.
- [ZW94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance. In *3rd European Conference on Computer Vision*, Stockholm, May 1994.
- [ZZyJ98] Zhingang Zhu, Guangyou Zu, Yudong yang, and Jesse S. Jin. Camera stabilization based on a 2.3D motion estimation and inertial motion filtering. In *Proceedings of the IEEE int. conf. on Intelligent Vehicles*, pages 329–334, October 1998.