



HAL
open science

Une nouvelle approche de mise en correspondance stéréoscopique dense par méthodes possibilistes

Christophe Rabaud

► **To cite this version:**

Christophe Rabaud. Une nouvelle approche de mise en correspondance stéréoscopique dense par méthodes possibilistes. Traitement du signal et de l'image [eess.SP]. Université Montpellier II - Sciences et Techniques du Languedoc, 2005. Français. NNT : . tel-00262287

HAL Id: tel-00262287

<https://theses.hal.science/tel-00262287>

Submitted on 11 Mar 2008

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.

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER II

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

Formation doctorale : **Systèmes Automatiques et Microélectroniques**

École doctorale : **Information, Structures, Systèmes**

présentée et soutenue publiquement

par

Christophe RABAUD

Le 15 juillet 2005

Titre :

**Une nouvelle approche de mise en correspondance
stéréoscopique dense par méthodes possibilistes**

JURY

Mme. Marie-José ALDON	Chargée de recherche CNRS au LIRMM	Directeur de thèse
Mme. Sylvie GALICHET	Professeur au LISTIC (Annecy)	Rapporteur
M. Philippe MONTESINOS	Maître de conférence au LGI2P (Nimes)	Examineur
Mme Michèle ROMBAUT	Professeur au LIS (Grenoble)	Rapporteur
M. Olivier STRAUSS	Maître de conférence au LIRMM	Co-directeur de thèse
M. René ZAPATA	Professeur au LIRMM	Président du jury

Remerciements

En premier lieu, je tiens à remercier Michel Habib, directeur du LIRMM au début de ma thèse et Michel Robert, actuel directeur du LIRMM, pour m'avoir accueilli au sein du laboratoire pour mon DEA et pour ma thèse.

Ensuite, mes remerciements vont à Michèle ROMBAUT et Sylvie GALICHET qui ont eu la gentillesse d'accepter d'être mes rapporteurs, René ZAPATA qui a accepté d'être le président du jury et enfin Philippe MONTESINOS pour avoir accepté d'être examinateur.

Je tiens à remercier tout particulièrement Oliver STRAUSS qui m'a encadré, guidé mais a su aussi me laisser choisir mes orientations de recherche durant ces années de thèse. Cela a été un plaisir et un honneur de travailler avec quelqu'un d'aussi compétent et d'aussi cultivé, et avec qui j'ai adoré nos diverses discussions sur tout et sur rien (pour quelqu'un qui fait du flou c'est un comble). Mes remerciements vont aussi à Marie-José ALDON, mon directeur de thèse, qui a accepté d'assumer cette fonction.

J'adresse aussi toute ma sympathie aux permanents du département de robotique : Alain Liégeois (pour qui j'ai toujours eu le plus grand respect), François Prunet (un professeur comme on n'en fait plus), Alain Fournier, Etienne Dombre, André Crosnier, Bruno Jouvencel (qui m'étonnera toujours autant), Philippe Fraisse, François Pierrot (Pierrot-san), Didier Crestani, Philippe Poignet, David Andreu (marche pas Tornado!!), David Guiraud, Christine Azevedo et les thésards du LIRMM qui sont passé du côté obscur des permanents j'ai nommé Olivier Company (trop de café), Frédéric Comby (Ex-Strauss thesarus elfus), Sébastien Druon et Sébastien Krut (un gars très sympa).

Viens maintenant le tour des thésards et je commencerais par les anciens, ceux qui étaient là quand j'ai commencé mon DEA : Jérôme Albaric (dit la barrique), Julien Aragones, Vincent Creuze (un ami), Philippe Baccou (docteur es' taipan), Christophe Caby, Geovany Borges (le maître brésilien), Gilles Duchemin (une tranche de jambon SVP), Frédéric Marquet (abla espagnol), Philippe Lucidarme (2 neurones), Benoît Telle.

Une pensée pour les thésards ayant fini leurs thèses juste avant moi : Andrés Vivas, Micael Michelin (kite & salsa), Meziane Bennour (t'es encore la ?), Fabien Lydoire (Lidux V10.4), Tomas Salgado (oui oui), Amornrit Puttipsatkajorn (si si). Reste les

petits nouveaux : Abdellah El jalaoui (un ancien élève), Jean-Mathias Spiewak (non, tu ne sera jamais meilleur que moi au ping-pong), Philippe Amat (dont je n'ai vu que les contours), Cédric Agniel, Micael Sauve (achète un jean neuf sans trous), José Marconi Rodrigues (encore un brésilien), Aurélien Noce, Gaël Pages (le chef du café), Pierre Maillet, Vincent Begoc, Yousra Ben Zaida, Samer Mohammed, Hassan El Makssoud, Bastien Jacquot, Vincent Nabat (breizh boy), Stéphane Ronchi (bientôt N°), Walid Zarrad, Florence Jacquy (fuzzy-doctorante), Lounis Douadi. Je rajouterai un stagiaire brésilien fort sympathique Pedro Paim.

Je voudrais aussi remercier mes parents qui m'ont toujours laissé le choix de mes études et qui m'ont soutenu et sans qui je ne serai pas là (même si je ne le dis pas souvent, bien sur que je vous aime). Un merci à mon frère David pour tout plein de chose, ce n'est pas la peine de les dire on se comprend sans parler.

Je voudrai remercier aussi mon ami Olivier Mateu (aka Olivier M des youngsters), une amitié très précieuse qui date de plus de 15 ans. J'ai aussi une pensée pour Carla Silva Rocha Aguiar qui a su me redonner une motivation et à qui je dois beaucoup, elle sait pourquoi.

Enfin, pour finir et pour être sur d'oublier personne, je remercie tout le monde.

Table des matières

Table des matières	iii
Table des figures	vii
Liste des tableaux	xi
1 État de l'art	5
1.1 Principe de la vision par ordinateur	7
1.1.1 Le modèle sténopé	7
1.1.2 Changement de repère	10
1.1.3 Image numérique	11
1.1.3.1 Echantillonnage de l'espace, cas des caméras à balayage	11
1.1.3.2 Echantillonnage de l'espace, cas des caméras matricielles	13
1.1.3.3 Image échantillonnée	13
1.1.3.4 Modèle d'erreur de l'échantillonnage de l'espace	15
1.1.3.5 Quantification de l'information du niveau de gris	16
1.1.3.6 Modèle d'erreur de la quantification des niveaux de gris	18
1.1.3.7 Modèle d'erreur de la numérisation	18
1.2 La vision stéréoscopique	18
1.2.1 Principe de base	19
1.2.2 Relation entre les projections - Epipolarité	19
1.2.3 Etalonnage	21
1.2.4 Reconstruction	22
1.2.5 Restriction géométrique des mises en correspondance	23
1.2.5.1 La contrainte épipolaire	23
1.2.5.2 La contrainte d'ordre	23

1.2.5.3	La contrainte d'unicité	24
1.2.6	Cas particulier des images redressées	25
1.3	Mise en correspondance	26
1.3.1	Les problèmes de la mise en correspondance	27
1.3.1.1	Variation du niveau de gris	27
1.3.1.2	Variation du nombre de pixel de la projection d'un même objet dans les deux images	27
1.3.1.3	Le problème de l'occultation	27
1.3.1.4	Violation des contraintes géométriques d'appariements	28
1.3.1.5	Les zones peu texturées	28
1.3.2	Que met-on en correspondance?	29
1.3.2.1	Mise en correspondance basée sur les indices caracté- ristiques	29
1.3.2.2	Mise en correspondance basée sur les pixels	30
1.3.3	Les méthodes globales	31
1.3.3.1	Programmation dynamique	32
1.3.3.2	Théorie des graphes	34
1.3.3.3	Autres méthodes utilisant les contraintes globales . . .	37
1.3.4	Les méthodes locales	38
1.3.4.1	Choix de la zone	39
1.3.4.2	Mise en correspondance de motif	40
1.3.4.3	Méthodes basées sur le gradient, approches différentielles	43
1.3.4.4	Mise en correspondance d'indices	44
1.3.5	Les méthodes coopératives	46
1.4	Conclusion	48
2	Modèle de pixel flou pour la mise en correspondance pixel à pixel	51
2.1	Introduction	51
2.1.1	Notre objectif	51
2.1.1.1	Sur quels critères peut-on se baser pour établir les mises en correspondance?	51
2.2	Pixel, aspect géométrique	53
2.2.1	Hypothèse sur la luminance	53

2.2.2	Mesure de luminance échantillonnée : rapport entre pixel et photosite	54
2.3	Modèles classiques de représentation d'une information aléatoire spatialisée : le pixel	56
2.3.1	Modélisation par noyau sommatif	57
2.3.1.1	Le noyau gaussien	60
2.3.1.2	Le noyau uniforme	61
2.3.2	Modélisation par un intervalle	62
2.3.3	Modélisation mixte : les intervalles de confiance	63
2.3.4	Modélisation floue	64
2.3.4.1	Interprétation d'une distribution de possibilité	64
2.3.4.2	Modélisation d'une distribution de probabilité par une distribution de possibilité	65
2.3.5	Modélisation du pixel par une distribution de possibilité pyramidale symétrique	66
2.4	Pixel, aspect chromatique	67
2.4.1	Représentation logique de l'intensité lumineuse	68
2.4.2	Evaluation de l'information chromatique du pixel flou	70
2.4.2.1	Cas du pixel \mathcal{M}^α avec $\alpha \in]0.5, 1]$	70
2.4.2.2	Cas du pixel \mathcal{M}^α avec $\alpha \in [0, 0.5]$	71
2.5	Mise en correspondance stéréoscopique	72
2.5.1	Concept de base	72
2.5.2	Processus bipolaire	73
2.5.3	Mise en correspondance géométrique dans le cas d'un capteur stéréoscopique étalonné avec erreur	74
2.5.4	Mise en correspondance chromatique	76
2.5.5	Fusion des contraintes de mise en correspondance	77
2.5.6	Evaluation des domaines de mise en correspondance	78
2.6	Conclusion	80
3	Mise en correspondance dense de pixels dans une paire d'images stéréoscopique	83
3.1	Principe de base	83
3.2	Modélisation des contraintes stéréoscopiques	84
3.2.1	Modélisation de la contrainte d'unicité	84

3.2.1.1	Violation de la contrainte d'unicité - modification à partir des possibilités de mise en correspondance	85
3.2.1.2	Violation de la contrainte d'unicité - modification à partir des nécessités de mise en correspondance	87
3.2.1.3	Respect de la contrainte d'unicité	87
3.2.2	Modélisation de la contrainte d'ordre	88
3.2.2.1	Principe	89
3.2.2.2	Relations spatiales	89
3.2.2.3	Proposition duale de la contrainte d'ordre	91
3.2.2.4	Violation de la contrainte d'ordre - modification à partir des possibilités de mise en correspondance	92
3.2.2.5	Violation de la contrainte d'ordre - modification à partir des nécessités de mise en correspondance	92
3.2.3	Cas des images redressées - contrainte du point de visée	93
3.3	Création d'une carte de disparité spécifique, comparaison avec les méthodes classiques	93
3.3.1	Cas des images redressées	94
3.3.2	Comparaison	97
3.3.2.1	Les méthodes	97
3.3.2.2	Les images	99
3.3.2.3	Résultats des méthodes de mise en correspondance de motif	101
3.3.2.4	Résultats des méthodes par programmation dynamique	103
3.3.2.5	Résultats des méthodes par coupure de graphe	104
3.4	Conclusion	107

Table des figures

1.1	Stéréoscope de Wheatstone (a), Stéréoscope de Duboscq (b).	5
1.2	Lentille convergente.	7
1.3	Principe de la chambre noire.	8
1.4	Principe du sténopé.	8
1.5	Le modèle géométrique d'une caméra.	9
1.6	Changement de repère dans l'espace tridimensionnel.	11
1.7	Principe du tube Vidicon.	12
1.8	Principe de la numérisation d'un signal électrique vidéo.	12
1.9	Relation entre pixel et portion de l'espace échantillonné.	13
1.10	Principe de la numérisation d'un signal électrique vidéo.	14
1.11	Rétine CCD (a), Rétine CMOS (b).	14
1.12	Image échantillonnée.	15
1.13	Numérisation de l'espace.	15
1.14	Cône de visibilité d'un photosite pour CMOS ou balayage.	16
1.15	Principe de la quantification.	17
1.16	Quantification linéaire.	17
1.17	Modèle d'erreur.	18
1.18	La géométrie d'un capteur stéréoscopique.	19
1.19	Définition de la contrainte épipolaire.	20
1.20	Définition de la contrainte épipolaire du point de vue pixel.	21
1.21	Définition de la contrainte d'ordre (a), Cas d'une surface transparente (b).	24
1.22	Cas de non-respect de la contrainte d'ordre (a), Cône interdit (b).	24
1.23	Définition de la contrainte d'unicité.	25
1.24	Cas de non-respect de la contrainte d'unicité.	25
1.25	Images stéréoscopiques redressées.	26

1.26	Variation du nombre de pixel (a), problème d'occultation (b).	28
1.27	Texture uniforme.	29
1.28	Exemple d'une représentation de l'espace des disparités en représentant les lignes droite-gauche (a), ou en utilisant la représentation ligne droite-disparité (b). L'intensité représente le coût de la mise en correspondance potentielle le long de la ligne de recherche, un pixel blanc représentant un faible coût.	33
1.29	Positions possibles pour f et la structure du graphe (en pointillés, les positions sont associées à des arcs verticaux) (a), le graphe avec les capacités des arcs (b).	36
1.30	Exemple de coupe minimale. L'appariement qui minimise la fonction d'énergie est $[(m_1, d_1), (m_2, d_2), (m_3, d_2)]$	37
1.31	Représentation par flot maximum de l'estimation de la disparité tel que proposée par Roy et Cox.	38
1.32	Mise en correspondance de motif.	40
1.33	Pyramides formées à partir d'une paire d'images à mettre en correspondance.	45
1.34	Mise en correspondance avec invariants et triangulation de Delaunay.	48
2.1	Les problèmes de l'échantillonnage.	54
2.2	Influence de l'échantillonnage entre deux vues proches (a), valeurs des niveaux de gris des deux motifs (b).	55
2.3	Différents échantillonnages de l'espace.	56
2.4	Définition d'un pixel.	57
2.5	Interpolation (a), ergodicité (b).	59
2.6	Représentation de l'échantillonnage par une distribution de probabilité unimodale centrée.	60
2.7	Représentation de l'échantillonnage par une distribution uniforme.	61
2.8	Intervalle inclus dans le pixel (a), Intervalle contenant plusieurs pixels (b).	63
2.9	Représentation d'intervalles de confiance associés à un noyau triangulaire.	65
2.10	Représentation floue d'un pixel.	66
2.11	Localisation du pixel flou par α -coupes.	67
2.12	Classe des pixels noirs (a), Classe des pixels blancs (b).	68
2.13	Compatibilité avec la classe des pixels noirs (a), Compatibilité avec la classe des pixels blancs (b).	70
2.14	Evaluation des informations chromatiques du pixel flou ($\alpha = 1$).	71

2.15	Evaluation des informations chromatiques du pixel flou ($\alpha \in]0.5, 1]$). . .	71
2.16	Evaluation de la compatibilité de l'information de luminance ($\alpha \in [0, 0.5]$). . .	72
2.17	Domaines de mise en correspondance	73
2.18	Sous-ensemble flou de distribution triangulaire pour représenter chaque paramètre de la matrice fondamentale (a), Sous-ensemble flou de distribution triangulaire pour représenter les coordonnées ligne et colonne d'un pixel (b), Intersection entre le bandeau épipolaire flou et un pixel flou (c).	75
2.19	Couple d'images stéréoscopique.	78
2.20	Lignes épipolaires Inf, Moy et Sup (a), bandeau épipolaire (b),	79
2.21	Possibilité et impossibilité chromatique.	80
2.22	Distributions de possibilité et de nécessité.	80
2.23	Coupe des distributions de possibilité et de nécessité.	81
3.1	Violation de la contrainte d'unicité.	85
3.2	Respect de la contrainte d'ordre.	89
3.3	Définition des relations spatiales dans le cas des images redressées (a), dans le cas général (b).	90
3.4	Violation de la contrainte d'ordre.	91
3.5	Un cas particulier de stéréovision.	93
3.6	Images stereoscopiques.	95
3.7	Répartition des possibilités de mise en correspondance (P.M.E.C.) et des possibilités de non mise en correspondance (P.N.M.E.C.) initiales. . . .	95
3.8	Répartition des possibilités de mise en correspondance (P.M.E.C.) et des possibilités de non mise en correspondance (P.N.M.E.C.) modifiées. . .	95
3.9	Distribution de la mesure de vérité d'appariement de la ligne x (a), distribution de la vraie mesure de d'appariement de la ligne x (b), image de disparité vraie (c), image de disparité calculée à partir de notre mesure (d).	96
3.10	Images stéréoscopiques 1 : Corridor de Lena, image gauche (a), image droite (b), image de disparité vraie (c).	100
3.11	Images stéréoscopiques 2 : Cônes de Scharstein et Szeliski, image gauche (a), image droite (b), image disparité vraie (c).	100
3.12	Images stéréoscopiques 3 : Statue de Tsukuba, image gauche (a), image droite (b), image disparité vraie (c).	101

3.13 Résultats des méthodes de mise en correspondance de motifs, images stéréoscopique1 : SAD(a), ZSAD(b), SSD(c), ZSSD(d), NCC(e), ZNCC(f), CENSUS(g).	102
3.14 Résultats des méthodes de mise en correspondance de motifs, images stéréoscopiques2 : SAD(a), ZSAD(b), SSD(c), ZSSD(d), NCC(e), ZNCC(f), CENSUS(g).	102
3.15 Résultats des méthodes de mise en correspondance de motifs, images stéréoscopiques3 : SAD(a), ZSAD(b), SSD(c), ZSSD(d), NCC(e), ZNCC(f), CENSUS(g).	103
3.16 Résultats des méthodes de mise en correspondance par programmation dynamique, images stéréoscopiques 1 : DP1 (a), DP2-Fuzzy (b).	104
3.17 Résultats des méthodes de mise en correspondance par programmation dynamique, images stéréoscopiques 2 : DP1 (a), DP2-Fuzzy (b).	104
3.18 Résultats des méthodes de mise en correspondance par programmation dynamique, images stéréoscopiques 3 : DP1 (a), DP2-Fuzzy (b).	105
3.19 Résultats des méthodes de mise en correspondance par coupure de graphe, images stéréoscopiques 1 : GC1 (a), GC2 (b), GC3 (c), GC4-Fuzzy (d), GC5-Fuzzy (e), GC6-Fuzzy (f).	105
3.20 Résultats des méthodes de mise en correspondance par coupure de graphe, images stéréoscopiques 2 : GC1 (a), GC2 (b), GC3 (c), GC4-Fuzzy (d), GC5-Fuzzy (e), GC6-Fuzzy (f).	106
3.21 Résultats des méthodes de mise en correspondance par coupure de graphe, images stéréoscopiques 3 : GC1 (a), GC2 (b), GC3 (c), GC4-Fuzzy (d), GC5-Fuzzy (e), GC6-Fuzzy (f).	106

Liste des tableaux

1.1	Méthodes de mise en correspondance de motifs dans le cas des images redressées.	41
1.2	Exemple de transformation de rang (à gauche) et de recensement (à droite).	42
3.1	Paramètres de simulation.	98

Introduction générale

On situe les débuts de la vision artificielle dans les années 1960. On appelle vision artificielle ou vision par ordinateur un système couplant une caméra à un ordinateur. Le capteur ainsi obtenu permet une perception sans contact de son environnement à l'instar de la vision humaine. Initialement utilisée dans la restauration des images spatiales, la vision par ordinateur a maintenant un spectre d'application très large (surveillance, télédétection, contrôle qualité, multimédia, art, jouet, etc.).

Les systèmes de vision sont utilisés pour obtenir des informations de bas niveau de complexité (mesure de luminance, vision 2D) jusqu'à des niveaux de complexité élevés (reconnaissance de formes, création de carte de navigation, etc.) et ce grâce à l'augmentation des vitesses et des capacités de calcul des ordinateurs.

La reconstruction d'une carte de profondeur fait partie des méthodes de grande complexité. Une caméra donne une mesure projective de la luminance des objets de l'environnement d'un système. Si on souhaite reconstruire une information tridimensionnelle à partir d'une image issue d'une caméra, il faut :

- soit utiliser des connaissances a priori sur les objets perçus,
- soit adjoindre à la caméra un système (télémètre, ladar, caméra, etc.) permettant de rajouter une information à l'image traitée.

Nous nous intéressons dans ce manuscrit au cas où l'autre source d'information est également une caméra.

La vision stéréoscopique, c'est-à-dire la perception d'une scène à partir de deux caméras ou plus, permet de reconstruire le relief apparent. L'obtention d'informations tridimensionnelles à partir de vues bidimensionnelles est directement liée à la possibilité d'apparier, dans chaque image, les projections 2D d'un même objet 3D. Ce sont l'information de disparité (c'est-à-dire la variation de position de la projection sur les différentes images) et la connaissance des différents paramètres des capteurs utilisés qui permettent de reconstruire les coordonnées d'un point tridimensionnel. Dans certaines applications (métrologie, reconstruction d'objet, etc.), la reconstruction tridimensionnelle nécessite l'utilisation de cartes de disparité dense. A l'inverse des méthodes qui appartiennent uniquement des points caractéristiques, nous nous sommes intéressés à la problématique de l'appariement de la totalité des pixels de chaque image.

L'appariement dans les images stéréoscopiques s'appuie généralement sur l'hypothèse que la luminance émise par tout point de l'environnement tridimensionnel peut être considérée comme uniforme si la variation de point de vue est faible. On pourrait penser naïvement que l'appariement des pixels des images stéréoscopiques devrait pouvoir être obtenu par une mesure de distance sur les luminances. Cependant, le processus d'intégration spatial et temporel de la luminance au niveau des rétines, et particulièrement l'échantillonnage et la quantification de cette mesure, invalident cette possibilité. De plus, des phénomènes tels que l'occultation partielle, le changement de point de vue, la variation de la spécularité, etc. gênent les procédures d'appariement.

La majeure partie des méthodes classiques de mise en correspondance utilise une représentation statistique de l'erreur d'appariement. Cependant, on est en droit de se poser des questions sur l'utilisation des outils de statistiques lorsque le nombre d'informations à mettre en correspondance ne dépasse pas deux. Pour pallier le biais de cette approche, il est d'usage d'utiliser des techniques de régularisation pour estimer les moments statistiques de la distribution de luminance locale. Ces techniques s'appuient sur l'hypothèse d'ergodicité du bruit de mesure en utilisant le voisinage de chaque pixel.

Ce type d'approche utilise des voisinages (ou fenêtres) pour obtenir une mesure statistique de la distance entre les deux fonctions d'illuminations échantillonnées. L'appariement est alors obtenu en recherchant, dans la seconde image, le correspondant du voisinage de chaque pixel de la première image et inversement. Cette procédure induit des quantités de calculs importants. Elle ne prend pas en compte les particularités des systèmes stéréoscopiques. En effet, si le système est étalonné, la connaissance du modèle de formation des images ainsi que de la position relative des caméras permettent de restreindre a priori les domaines de correspondance. Cette restriction utilise les outils de la géométrie épipolaire qui permettent d'astreindre le correspondant d'un pixel d'une image à une ligne définie dans l'autre image. L'utilisation de cette contrainte épipolaire permet en outre de rejeter des appariements ambigus provenant de motifs répétitifs fréquents dans les images naturelles. Cependant, la restriction a priori des domaines d'appariement peut biaiser considérablement la mise en correspondance si les paramètres d'étalonnage ne sont pas fiables (ou si le modèle utilisé est inadéquat).

D'autre part, l'aspect échantillonné de l'information de luminance influe directement sur la robustesse des méthodes de mesure de similarité. Comme il existe différents types de technologies de capteurs vidéos (caméras à balayage, rétines C.C.D. ou C.M.O.S.) et donc différents types d'échantillonnages, il semble difficile de fixer un modèle de représentation de l'image numérique suffisamment général pour que les traitements effectués dans le cadre de la mise en correspondance puissent s'abstraire de la technologie d'obtention de l'image. Même si la connaissance de la technologie utilisée est disponible, elle est difficile à utiliser sans rajouter dans le traitement de l'information un haut niveau de complexité de représentation ainsi que beaucoup d'a priori sur les propriétés de diffusion de la lumière. A ce défaut s'ajoute tous les problèmes liés aux bruits des systèmes, aux différences de points de vue, etc. et surtout à la quantification qui induit imprécision et saturation de l'information de luminance. La modélisation de la luminance par une variable aléatoire ne permet pas de présenter de façon simple ces

défauts de connaissance du signal.

Plutôt que de tenter d'augmenter la fiabilité de l'hypothèse de similarité de la luminance pour deux pixels correspondants et d'ergodicité du bruit dans l'utilisation des voisinages, nous avons opté pour une approche visant à minimiser l'influence des apriorismes dans le processus de mise en correspondance. Nous avons choisi de restreindre notre étude à l'appariement dense de pixels dans des images stéréoscopiques en niveau de gris. Le plan de cet ouvrage est détaillé ci-dessous.

Nous présentons tout d'abord un état de l'art de la vision par ordinateur en reprenant toute la chaîne d'acquisition de l'information de luminance depuis sa projection continue jusqu'aux données numériques (échantillonnées et quantifiées) représentées par les pixels. Le volume de ce premier chapitre est important au regard de la taille de l'ensemble du manuscrit. Cette disproportion a deux causes. La première est que la problématique de l'appariement stéréoscopique n'est pas neuve et que la littérature sur ce sujet est abondante et variée. La seconde est liée au fait que nous proposons une approche nouvelle et que la justification de notre démarche nécessite de réviser un certain nombre de points qui sont généralement admis sans discussion. Ainsi nous voyons les différents modèles proposés par la littérature pour représenter la numérisation de la luminance ainsi que leur implication sur les processus d'appariement. Nous présentons ensuite les bases de la vision stéréoscopique, c'est-à-dire les rudiments de géométrie épipolaire, le principe du calibrage des capteurs stéréoscopiques ainsi que les méthodes de reconstruction tridimensionnelle. Enfin, nous proposons un état de l'art sur les méthodes de mise en correspondance. L'aptitude des méthodes d'appariement à délivrer une carte de disparité fiable est liée à la robustesse de la méthode d'estimation de la similarité des mesures de luminance. Ces mesures de similarité sont elles-mêmes très influencées par la modélisation du passage continu-discret. Un bon modèle de représentation des informations de luminance serait un modèle suffisamment général pour que les traitements effectués puissent s'abstraire de la technologie d'obtention de l'image et suffisamment simple pour être utilisable sur un ordinateur ordinaire et éviter les biais de modélisation. De notre état de l'art, il est ressorti qu'aucune méthode ne semblait satisfaire à cette caractéristique. L'élaboration d'un nouveau modèle de représentation de l'illumination échantillonnée nous est apparue nécessaire.

Ce modèle, nous le présentons dans le chapitre 2. Le nouveau modèle de pixel que nous proposons tente de mieux représenter le phénomène d'échantillonnage de l'espace ainsi que les phénomènes de quantification de la valeur de luminance associée. L'échantillonnage spatial est modélisé par une partition floue tandis que l'information de luminance est associée à deux classes duales. Cette modélisation nous permet une nouvelle approche de la mise en correspondance des informations discrètes quantifiées de luminance par le biais de la théorie des possibilités. Ainsi nous associons, à chaque pixel de chaque image, un domaine pondéré des pixels de l'autre image pouvant lui être appariés.

Pour augmenter la spécificité des domaines d'appariement, nous proposons, dans le chapitre 3, un algorithme coopératif et itératif. Nous avons adapté trois contraintes

d'appariements stéréoscopiques classiques en les formulant de façon sémantique. Nous validons notre approche en estimant la carte de disparité sur des images de référence puis en la comparant avec celles obtenues par les approches classiques présentées dans le premier chapitre.

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

Chapitre 1

État de l'art

Au III^{ème} siècle avant J-C, Euclide, géomètre grec, définit la perception du relief comme "recevoir, au moyen de chaque oeil, l'impression simultanée de deux images dissemblables d'un même objet". En 1484, Léonard de Vinci précise les principes de la vision binoculaire mais l'état d'avancement technologique de l'époque ne lui permettait pas de mettre ces principes en application. La vraie histoire de la stéréoscopie commence en 1832 avec l'Anglais Charles Wheatstone. Il étudie des moyens d'observation de dessins en relief. En 1838, il fait breveter le premier "stéréoscope" (figure 1.1a). C'est un appareil muni de deux miroirs dans lequel on observe des dessins placés aux extrémités du stéréoscope. Ces deux images donnaient, à l'observateur, une impression de relief. En 1844, David Brewster applique le principe de Wheatstone à la photographie : la photographie stéréoscopique est née. En 1850, Duboscq construit le premier stéréoscope en série (figure 1.1b).

De nombreuses équipes composées de biologistes, de psychologues et autres scientifiques ont tenté de dégager une modélisation de la vision humaine sous forme d'un système stéréoscopique binoculaire [Jul60], [BBP67], [CDW76]. Se basant sur l'ensemble de ces travaux, Marr et Poggio [MP79] ont proposé en 1979 une méthode de reconstruction tridimensionnelle utilisant deux caméras. Le principe de cette méthode consiste à utiliser plusieurs vues d'une même scène et d'en reconstituer la géométrie tridimensionnelle grâce à la disparité des images. Cette méthode porte naturellement le nom de

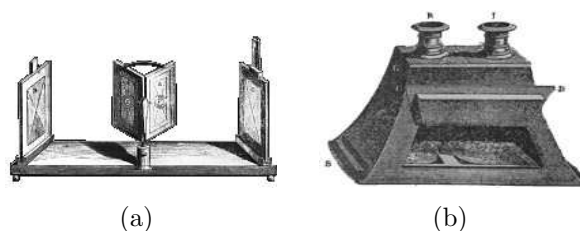


Figure 1.1 – Stéréoscope de Wheatstone (a), Stéréoscope de Duboscq (b).

vision stéréoscopique ou stéréovision. Une implémentation partielle de cette méthode a été réalisé par Grimson [Gri81] en 1981. A. Lux [Lux85], en 1985, fixe comme objectif ambitieux à la vision par ordinateur "la construction d'un capteur visuel capable de remplir les fonctions de l'oeil humain dans un système artificiel". La stéréovision est donc vue comme un moyen, à l'image de la vision humaine, de reconstruire le relief à partir de deux images prises au même instant mais à partir de points de vue différents.

Le principe de la reconstruction stéréoscopique est basé sur l'aptitude à retrouver, dans chaque image, la projection d'un même objet (point, ligne, motif). En effet, l'information de profondeur d'un objet est liée d'une part à la disparité de ses projections dans les différentes images et d'autre part, à la position relative des deux capteurs d'images. La reconstruction stéréoscopique fait donc apparaître deux problèmes différents. D'une part le calcul de la disparité qui fait intervenir des problèmes de mise en correspondances et d'autre part, l'aptitude à inverser le problème de la projection stéréoscopique, c'est à dire de retrouver l'information tridimensionnelle à partir de la connaissance de la position relative des capteurs d'images et de celle de la disparité. Les études basées sur la géométrie projective proposé par O. Faugeras et son équipe [Fau93] semblent avoir débouché sur une méthode générique de reconstruction tridimensionnelle. Concernant les problèmes liés aux estimations (étalonnage, reconstruction de surface) et aux appariements, il semble qu'aucune méthode ne soit suffisamment fiable, robuste et efficace pour permettre une utilisation simple des mesures obtenues par stéréovision. Il semble entre autre difficile d'obtenir des appariements denses (c'est à dire la mise en correspondance de la totalité des points), ce qui limite l'utilisation de la stéréovision pour la métrologie, la reconstruction 3D, etc. C'est dans ce cadre que s'inscrit notre travail. Nous nous intéressons à la reconstruction dense d'une scène tridimensionnelle à partir d'un couple d'images stéréoscopiques connaissant une estimation de la position relative des deux caméras.

Dans le premier paragraphe, nous allons revenir sur les bases de la vision par ordinateur en reprenant le modèle sténopé afin d'établir un modèle complet de caméra. Nous nous intéressons plus particulièrement aux problèmes d'échantillonnage de l'image ainsi qu'à la quantification des niveaux de gris. Ensuite, nous verrons les principes de base de la vision stéréoscopique ainsi que la géométrie épipolaire résultante. Ces deux premiers paragraphes vont nous permettre d'introduire les différents termes utilisés ainsi que les différentes notations employées. Enfin, nous verrons les méthodes habituellement employées pour la mise en correspondance de pixels dans les différentes caméras. Pour chacune de ces méthodes, nous présenterons le principe de base ainsi que les diverses améliorations apportées. Nous avons classé ces méthodes en deux groupes : les approches utilisant des contraintes locales regroupant les méthodes basées sur le traitement du signal d'une part et les approches utilisant des contraintes globales utilisant des informations " haut niveau " d'autre part. Le paragraphe 1.3.3 traitera des méthodes basées sur l'utilisation de contraintes globales. L'une des principales méthodes, la programmation dynamique, y sera présenté ainsi qu'un ensemble d'outils plus récent. Traitant des méthodes locales, le paragraphe 1.3.4 recense les différentes méthodes apparentées à la corrélation ainsi qu'aux méthodes basées sur un calcul différentiel du flot

optique. Les méthodes basées sur la mise en correspondance d'indices remarquables y seront aussi présentées. Nous poursuivrons par une présentation des méthodes à cheval entre les méthodes locales et les méthodes globales, utilisant ces contraintes de façon coopératives. Nous finirons par une conclusion justifiant une nouvelle approche utilisant une nouvelle représentation des pixels ainsi que des contraintes stéréoscopiques moins contraignantes.

1.1 Principe de la vision par ordinateur

1.1.1 Le modèle sténopé

Une image, au sens usuel, est une projection optique de l'univers tridimensionnel. La modélisation du phénomène de formation des images a fait l'objet de nombreux travaux. Le modèle le plus connu est celui de Gauss, dont l'approximation dite "des lentilles minces" (rayons paraxiaux) est l'un des plus utilisés. Dans cette approximation, la position de l'image est indépendante de la direction des rayons lumineux et la lentille est stigmatique (c'est dire qu'à tout point objet correspond un point image et non une tache). Une lentille mince possède 2 foyers placés symétriquement sur l'axe optique par rapport au centre S de la lentille (figure 1.2, F est le foyer objet et F' est le foyer image). Pour construire l'image de l'objet AB formée à travers une lentille mince, on utilise les propriétés de la lentille en traçant deux rayons particuliers issus de B :

- le rayon (1) parallèle à l'axe optique passant à la sortie de la lentille par le foyer image F' ,
- le rayon (2) passant par le centre de la lentille ne subissant pas de déviation.

L'intersection de ces deux rayons permet de déterminer B' , image de B . On déduit de cette construction la position et la taille de l'image $AB : A'B'$.

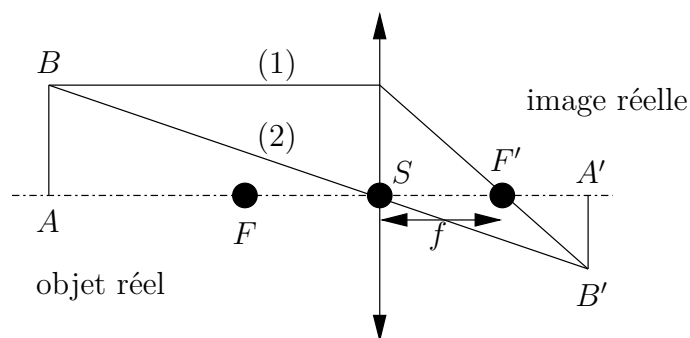


Figure 1.2 – Lentille convergente.

Le modèle de Gauss reste complexe pour une utilisation en vision par ordinateur. Cependant, on peut montrer que, si la distance focale de l'objectif utilisé est petite par rapport à la distance de l'image à l'objet, le modèle de Gauss peut être approché par un modèle projectif homographique simple que l'on appelle modèle sténopé ou trou d'épingle. L'invention du sténopé (venant du grec signifiant "vue ou oeil étroit") re-

monte à l'antiquité. Aristote (384-322 avant J-C) percevait un trou dans un mur d'une pièce sombre pour obtenir une image renversée de tous les objets placés devant cet orifice. Le mathématicien arabe Alhazen (965-1039) a établi les formules mathématiques

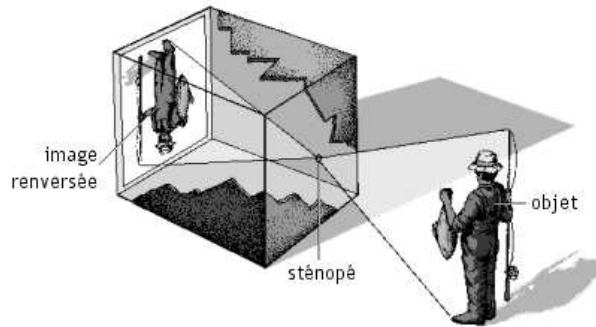


Figure 1.3 – Principe de la chambre noire.

du sténopé. Il a pu ainsi expliquer la vision binoculaire. En 1515, Léonard de Vinci décrit la *camera obscura*. Le principal défaut du sténopé était de donner une image peu lumineuse. Si on agrandissait le trou pour avoir plus de lumière alors l'image devenait floue puisqu'on s'éloignait du cas idéal d'une source ponctuelle. La solution de remplacer le trou par une lentille est attribuée à tort à Giambattista della Porta (1535-1613) qui a publié cette idée en 1589 dans le dix-septième volume de *Magia Naturalis*. Vingt ans plutôt, Daniello Barbero avait déjà publié l'idée dans *La Pratica della Perpettiva*.

Ce modèle sténopé a fait l'objet de nombreuses études [Rob65], [HZ00]. Considérons le cas monodimensionnel de la figure 1.4. Notre caméra est alors représentée par :

- le sténopé ou centre optique c de l'objectif où convergent les rayons lumineux et par lequel passe la droite focale (cx),
- le point o , point principal de la rétine,
- le plan rétinien situé à distance focale f (ici le segment $[c, o]$) par rapport au centre optique.

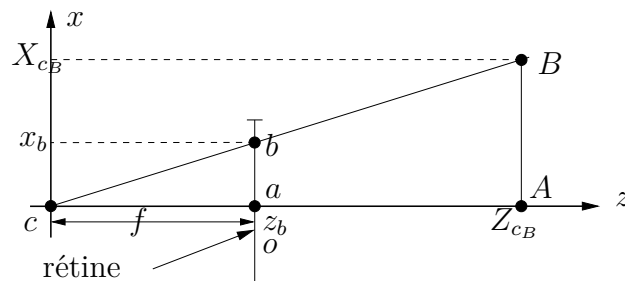


Figure 1.4 – Principe du sténopé.

Le centre optique peut être représenté devant ou derrière le plan rétinien. Dans notre cas, il sera placé derrière. Le point B de coordonnées $(X_{cB}, Z_{cB})^t$ exprimé dans le repère caméra rayonne dans toutes les directions mais seul un pinceau lumineux (hypothèse

que l'objet a une surface Lambertienne) passe le centre optique c se projetant ainsi sur le plan rétinien en b . De simples considérations géométriques nous permettent de calculer x_b (z_b étant trivialement égale à f). Le triangle BcA est semblable au triangle bca donc $\frac{ba}{BA} = \frac{ca}{cA}$. Autrement dit, la taille de la projection $cb = x_b$ est proportionnelle à la distance sténopé-rétine (la distance focale $ca = z_b = f$) et inversement proportionnelle à la distance sténopé-point Z_{cB} . La projection du point B dans le plan rétinien (ou plan image) exprimée dans le repère du sténopé (ou repère caméra) s'écrit alors :

$$\begin{cases} x_b = f \frac{X_{cB}}{Z_{cB}} \\ z_b = f \end{cases}$$

L'extension de ce principe à une image plane est triviale. Sur la figure 1.5 sont représentés le plan I sur lequel se forme l'image, ou plan rétinien, et l'axe optique (co) que nous prenons par convention identique à l'axe z du repère caméra. La distance entre o et c est la distance focale f . Un point B se projette en b dans le plan rétinien le long d'une droite passant par B et c . Soit $(X_{cB}, Y_{cB}, Z_{cB}, 1)^t$ les coordonnées du point B dans le repère caméra (avec t l'opérateur de transposition). Les coordonnées de la projection de B dans le plan rétinien sont :

$$\begin{cases} x_b = f \frac{X_{cB}}{Z_{cB}} \\ y_b = f \frac{Y_{cB}}{Z_{cB}} \\ z_b = f \end{cases}$$

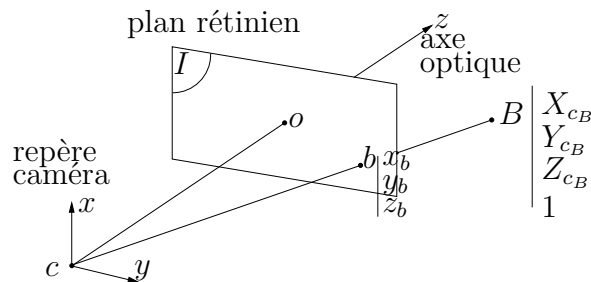


Figure 1.5 – Le modèle géométrique d'une caméra.

Si le modèle sténopé est plus simple que le modèle de Gauss, il reste quand même non linéaire. La géométrie projective, initiée par Desargues puis Pascal, mais principalement développé au 19^{ème} siècle par Monge repose sur la notion de projection centrale (dite aussi conique). On doit à Olivier Faugeras [Fau93] d'avoir popularisé la notion de géométrie projective ramenant ainsi le modèle sténopé à un modèle presque linéaire. Selon la géométrie projective, un point dans \mathcal{P}^2 (plan projectif) est définie par trois variables formant un vecteur de coordonnées à un facteur d'échelle près. En posant $s = \frac{Z_{cB}}{z_b} = \frac{X_{cB}}{x_b} = \frac{Y_{cB}}{y_b}$ alors les coordonnées de la projection de B dans le plan rétinien peuvent faire l'objet d'une écriture linéaire :

$$\begin{pmatrix} sx_b \\ sy_b \\ sz_b \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_{c_B} \\ Y_{c_B} \\ Z_{c_B} \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} sx_b \\ sy_b \\ sz_b \end{pmatrix} = P \begin{pmatrix} X_{c_B} \\ Y_{c_B} \\ Z_{c_B} \\ 1 \end{pmatrix} \quad (1.1)$$

La matrice P est la matrice de projection perspective qui relie les coordonnées tridimensionnelles d'un point à ses coordonnées bidimensionnelles projectives exprimées dans le plan rétinien.

1.1.2 Changement de repère

Nous venons de voir qu'il est possible de connaître les coordonnées de la projection d'un point $B = (X_{c_B}, Y_{c_B}, Z_{c_B})^t$ dans le plan rétinien à partir de ses coordonnées exprimées dans le repère caméra. Dans la plupart des cas, ce repère n'est pas le repère de travail. Les coordonnées des éléments de la scène (reconstruction [LH81], [Xie89]) ou des organes effecteurs (robotique [AEH98]) sont données dans un repère appelé repère de travail ou repère "monde" (Figure 1.6). Soient $(X_{w_B}, Y_{w_B}, Z_{w_B}, 1)^t$ les coordonnées du point B dans le repère de travail et K la transformation entre le repère de travail et le repère caméra tel que :

$$\begin{pmatrix} X_{c_B} \\ Y_{c_B} \\ Z_{c_B} \\ 1 \end{pmatrix} = K \begin{pmatrix} X_{w_B} \\ Y_{w_B} \\ Z_{w_B} \\ 1 \end{pmatrix}$$

avec :

$$K = \begin{pmatrix} R & T \\ 0_3^t & 1 \end{pmatrix} \quad (1.2)$$

ou R est la matrice de rotation entre les deux repères et T la matrice de translation entre les deux centres des deux repère et $0_3 = (0, 0, 0)^t$.

La position de B dans le plan rétinien est donc :

$$\begin{pmatrix} sx_b \\ sy_b \\ sz_b \end{pmatrix} = PK \begin{pmatrix} X_{w_B} \\ Y_{w_B} \\ Z_{w_B} \\ 1 \end{pmatrix} \quad (1.3)$$

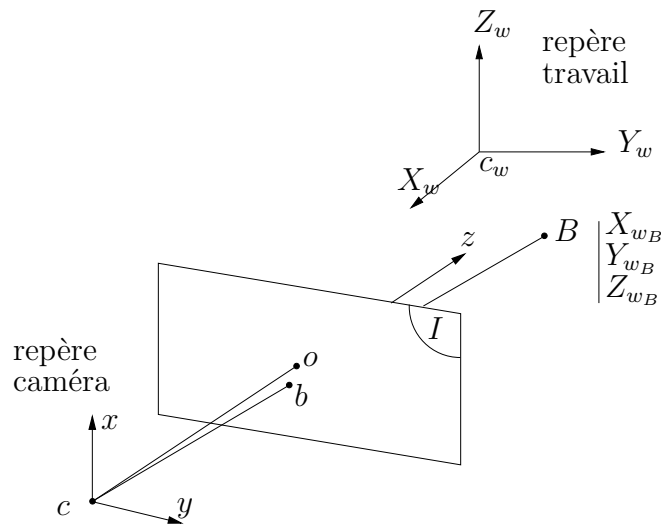


Figure 1.6 – Changement de repère dans l'espace tridimensionnel.

1.1.3 Image numérique

Pour qu'une image puisse être traitée par ordinateur, il faut qu'elle soit disponible sous format numérique, c'est à dire échantillonnée et quantifiée. Cette transformation induit une adaptation de la représentation mathématique du signal mesuré. Pendant de nombreuses années, ce sont les techniques issues de la théorie de l'échantillonnage de Shannon qui ont assuré cette adaptation. Ces méthodes sont souvent limitées par le fait qu'elles font appel à des techniques statistiques pour traiter les informations en faible nombre alors que l'objet des statistiques est de traiter des grands nombres. D'autres méthodes, purement discrètes, ont fait leur apparition sous le nom de "géométrie discrète des images" [JR70], [Thi92]. L'aspect peu nuancé de ces méthodes les rend difficilement utilisables dans le cadre qui est le nôtre. L'élément de base en imagerie numérique est le pixel. De l'anglais "picture element", il est souvent considéré l'équivalent du point dans le monde continu. Un pixel regroupe les deux transformations appliquées à l'image réelle : il est repéré par un jeu de coordonnées entières et un niveau de gris quantifié lui est associé. Utilisés sous forme de mosaïque, les pixels, vu de loin, donnent l'impression d'une image continue.

1.1.3.1 Echantillonnage de l'espace, cas des caméras à balayage

Depuis l'ICONOSCOPE de Vladimir Kosma Zworykin en 1923 jusqu'à la très répandue VIDICON (figure 1.7), le principe des caméras à balayage reste le même. Il s'agit d'utiliser une surface sensible à la lumière. Grâce à un système optique, l'image à analyser est projetée sur une couche photoconductrice. Un faisceau électronique, commandé par le système de balayage, explore successivement chaque point de cette image électrique. On extrait ainsi un signal électrique dont l'amplitude est représentative de l'éclairement du point analysé sur l'image. Il va être déchargé en permanence par un système de lecture donnant naissance à un courant électrique : le signal vidéo. La sur-

face sensible est lue ligne par ligne, des impulsions rajoutées sur le signal vidéo viennent signaler les changements de ligne et de trame.

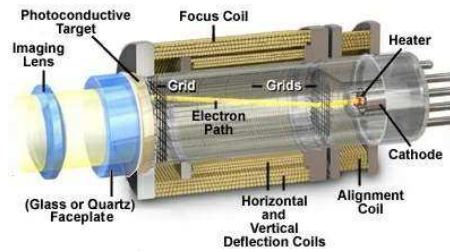


Figure 1.7 – Principe du tube Vidicon.

La discrétisation de l'espace des colonnes s'obtient par échantillonnage temporel du signal vidéo entre deux signaux de retour de ligne. Le système d'échantillonnage temporel se comporte comme un circuit associant un interrupteur et une capacité. Pendant un temps τ (très inférieur à la période d'échantillonnage T), le circuit est fermé chargeant ainsi le composant capacitif. Puis le circuit est ouvert, la charge est lue et numérisée. Ce principe est repris sur la figure 1.8. Si i est le numéro de la ligne de l'image et t le temps écoulé depuis le signal de retour de ligne, la valeur lue sur le composant capacitif est affectée au pixel de coordonnée (i, j) avec $i = \text{trunc}(\frac{t}{T})$, ou $\text{trunc}(x)$ est la troncature de x .

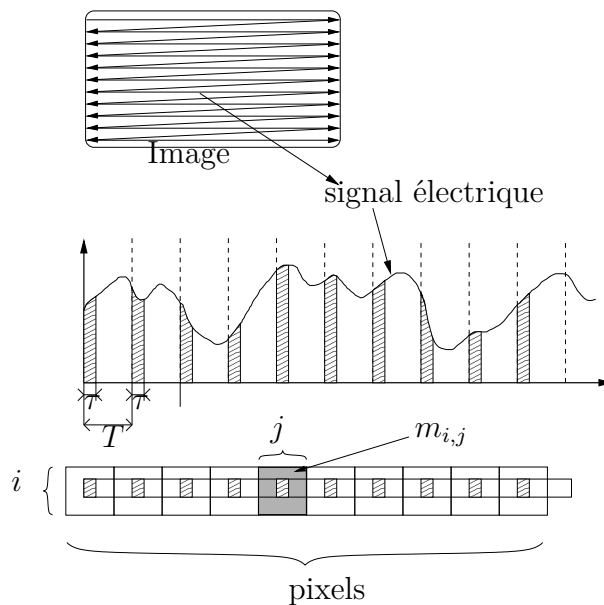


Figure 1.8 – Principe de la numérisation d'un signal électrique vidéo.

Lors de la visualisation de l'image numérisée, cette valeur sera interprétée par l'oeil comme étant la valeur de toute la surface du pixel. L'interpolation produite par le système oculaire donnera une impression de continuité. Cette valeur ne représente

cependant que l'intégration du signal lumineux sur une surface beaucoup plus petite (figure 1.9).

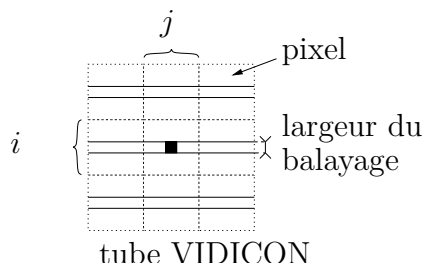


Figure 1.9 – Relation entre pixel et portion de l'espace échantillonné.

1.1.3.2 Echantillonnage de l'espace, cas des caméras matricielles

En 1970, W. S. Boyle et G. G. AMELIO du laboratoire BELL inventent un nouveau composant électronique : le C.C.D. pour Coupled Charge Device que l'on peut traduire par "dispositif à transfert de charge". Un an plus tard E. Arnold de la société PHILIPS invente le C.I.D.I.S. pour Charge Injection Device Image Sensor ou capteur d'image à injection de charge dont les photosites (plus petit élément homogène qui réagit à la lumière dans un capteur électronique) sont des capacités M.O.S. Les versions ultérieures du C.I.D.I.S. portent le nom générique de rétines C.M.O.S.

Les C.C.D. comme les C.M.O.S. sont, comme la plaque photographique, des récepteurs bidimensionnels composés par une mosaïque de minuscules récepteurs à semi-conducteur. Les photosites sont disposés comme les cases d'un échiquier en rangs et en colonnes. Chaque photosite accumule un nombre d'électrons proportionnel à la quantité de lumière reçue. La lecture des charges électriques accumulées va permettre de reconstituer l'image.

Le signal issu de ce type de composant est, par nature, échantillonné. Cependant, sa transmission est souvent assurée par un câble vidéo. Dans ce cas, le processus de numérisation est identique à celui des caméras à balayage, à la différence près que le signal échantillonné est transformé en un signal analogique au mieux par un interpolateur, mais généralement par un bloqueur (figure 1.10). Dans ce cas, les modèles d'erreur sont très éloignés des bruits gaussiens et uniformes couramment employés.

Si la transmission du signal est numérique, alors à chaque pixel peut être associé un photosite. Dans le cas des C.C.D., les photosites sont jointifs (figure 1.11a). Par contre, dans le cas des C.M.O.S., à chaque photosite étant associé un circuit de traitement, la mosaïque des photosites n'est pas jointive (figure 1.11b).

1.1.3.3 Image échantillonnée

L'objet de ce paragraphe est de relier les coordonnées métriques d'un point de l'espace 3D aux coordonnées du pixel sur lequel il se projette. L'origine du plan rétinien (ou plan image) est définie par o , intersection entre l'axe optique et le plan rétinien.

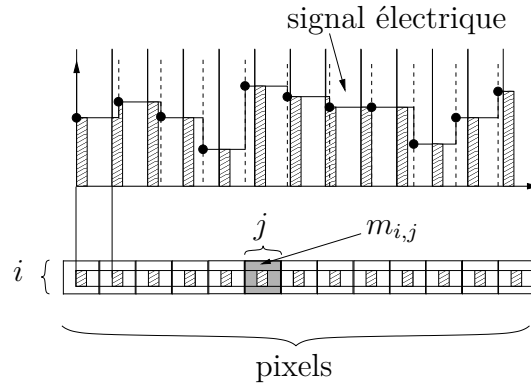


Figure 1.10 – Principe de la numérisation d'un signal électrique vidéo.

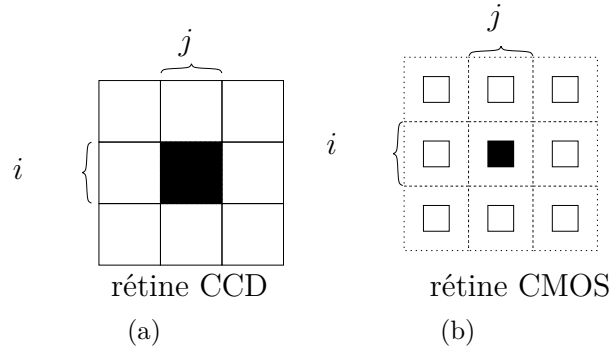


Figure 1.11 – Rétine CCD (a), Rétine CMOS (b).

Dans l'image échantillonnée, l'origine est le coin supérieur gauche de la rétine. Soient (u_o, v_o) les coordonnées du point o dans ce nouveau repère, Ni et Nj le nombre de lignes et le nombre de colonnes de l'image, h_r la dimension de la rétine selon l'axe x et l_r la dimension de la rétine selon l'axe y . Dans le repère (u, v) , les coordonnées de b sont :

$$\begin{pmatrix} u_b \\ v_b \end{pmatrix} = \begin{pmatrix} -k_u & 0 \\ 0 & k_v \end{pmatrix} \begin{pmatrix} x_b \\ y_b \end{pmatrix} + \begin{pmatrix} u_o \\ v_o \end{pmatrix} \quad (1.4)$$

avec $k_u = h_r/Ni$ et $k_v = l_r/Nj$.

La projection b de B sur la rétine est considérée comme participant à la valeur de niveau de gris du pixel m de coordonnées $(i, j)^t$ tel que :

$u_b \in [i - 0.5, i + 0.5[$ et $v_b \in [j - 0.5, j + 0.5[$. On écrit aussi :

$$\begin{pmatrix} i \\ j \end{pmatrix} = \begin{pmatrix} \text{ent}\{u_b\} \\ \text{ent}\{v_b\} \end{pmatrix} \quad (1.5)$$

avec $\text{ent}\{x\}$ représentant la partie entière de x .

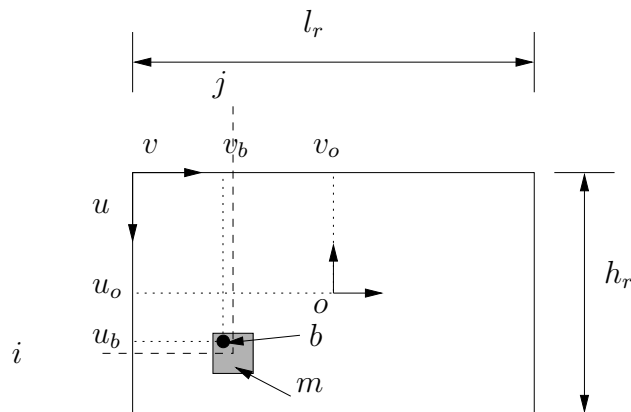


Figure 1.12 – Image échantillonnée.

1.1.3.4 Modèle d'erreur de l'échantillonnage de l'espace

Pour bien comprendre l'impact de l'échantillonnage sur les variations des niveaux de gris de l'image, il est important de différencier les photosites (entité physique intégrant le flux lumineux pendant le temps d'ouverture pour les caméras matricielles ou la surface du faisceau échantillonnée pour les caméras à balayage) et le domaine de l'espace rétinien affecté à chaque pixel que nous appellerons "pixel physique" (figure 1.13).

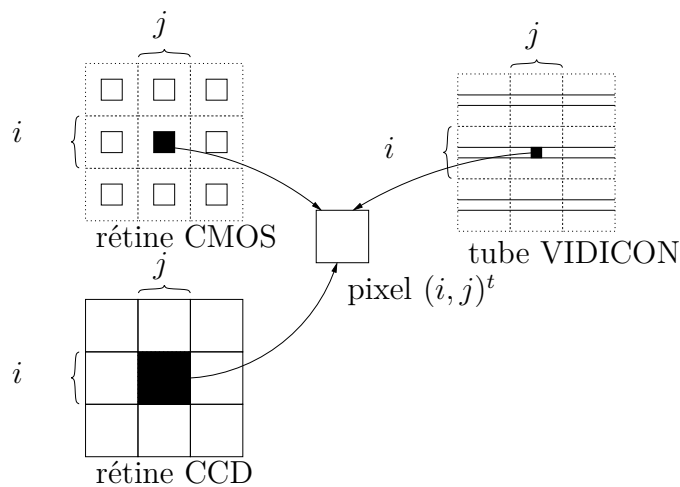


Figure 1.13 – Numérisation de l'espace.

Dans le cas des rétines C.C.D., on peut considérer qu'il y a coïncidence entre photosite et pixel physique (du moins lorsque le signal a été transmis sous sa forme échantillonnée). Dans le cas des caméras à balayage et des rétines C.M.O.S., le photosite est, a priori, inclus dans le pixel physique. Dans ce dernier cas, certains points de l'espace ne sont pas "vus" ; c'est à dire qu'ils ne participent pas directement à l'estimation de la valeur du niveau de gris des pixels qui leur sont associés (figure 1.14). Néanmoins, sous l'hypothèse de continuité de la matière (pas de discontinuité forte le long d'une surface) ainsi que sous l'hypothèse de surface ayant une granularité faible, alors on

admet qu'il y a une forte corrélation entre les points de l'espace qui sont à l'extérieur du cône de visibilité du photosite et ceux qui sont à l'intérieur. Lorsque la dimension du photosite est très inférieure à la dimension du pixel physique, la modélisation de l'échantillonnage par la convolution avec une distribution de Dirac est envisageable. Sinon, il faut supposer que cette distribution est convoluée avec un noyau représentant le défaut d'adéquation de l'échantillonnage.

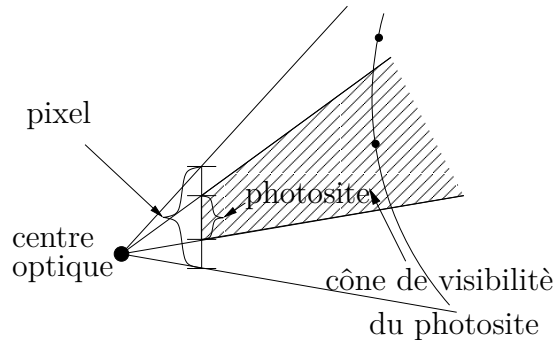


Figure 1.14 – Cône de visibilité d'un photosite pour CMOS ou balayage.

Même si la position relative de la caméra et l'environnement ne bougent pas, et que l'illumination globale de la scène ne varie pas, la valeur des niveaux de gris des pixels peut varier. Les différentes sources de variations sont multiples : variation du temps d'intégration, aléas électroniques, couplages électriques, méthode de lecture, ré-échantillonnage, etc. Aux vues de toutes ces sources de perturbation, on constate qu'il serait très difficile d'avoir un modèle complet, et si ce modèle complet pouvait être établi, son identification semble impossible à maîtriser.

L'échantillonnage de l'espace induit un bruit. Ce bruit est, au minimum, de type uniforme si la surface observée est peu granuleuse. Si on considère que la taille du photosite est très inférieure à la taille du pixel physique alors la distribution du bruit peut être considérée comme une distribution centrée autour d'un mode (comme par exemple un bruit gaussien). Cette distribution de bruit permet de prendre en compte les aléas d'illumination. Enfin, l'hypothèse de continuité de la matière ainsi que du couplage entre les valeurs (dans le cas des caméras matricielles) induit une corrélation entre les valeurs donc un bruit coloré. Traditionnellement, l'erreur induite par l'échantillonnage est modélisée par des noyaux gaussiens [Sch01], des distributions uniformes [KPKP89],[KK01] ou des intervalles d'erreur [Tel03].

1.1.3.5 Quantification de l'information du niveau de gris

La quantification est assurée par des composants électroniques appelés C.A.N. (Convertisseur Analogique Numérique). Quantifier un signal échantillonné consiste à associer, à chaque échantillon, une valeur numérique entière. Chaque valeur numérique représente une portion de l'espace de définition de la variable à quantifier.

Pour quantifier V_e , la grandeur électrique associée à la lumière reçue par un photosite, il faut donc partitionner son domaine de définition et associer une grandeur numérique

à chaque cellule de la partition qui est généralement un intervalle (Figure 1.15)



Figure 1.15 – Principe de la quantification.

Le nombre de cellules de la partition est N . Il est défini par la résolution numérique choisie pour la quantification. Les systèmes les plus répandus représentent les niveaux de gris d'une image sur 8 bits. Dans ce cas $N = 2^8 = 256$.

La partition représentée sur la figure 1.15 n'est pas régulière. Cette irrégularité peut avoir une signification particulière (par exemple représente une courbe de sensibilité du capteur qui serait non-linéaire). Cependant, un composant réalisant une telle quantification n'est pas simple à réaliser et n'est pas simple à paramétrer.

La quantification la plus courante est la quantification linéaire. Elle utilise trois paramètres qui sont la valeur de saturation haute V_{\max} , la valeur de saturation basse V_{\min} et N le nombre de cellules de la partition (figure 1.16). La valeur de quantification n associée à V_e est obtenue de la façon suivante :

$$n = \begin{cases} 0 & \text{si } V_e < V_{\min} \\ N - 1 & \text{si } V_e > V_{\max} \\ \text{ent} \left\{ \frac{V_e - V_{\min}}{V_{\max} - V_{\min}} N \right\} & \text{sinon} \end{cases}$$

La quantité V_{\min} est souvent appelée offset tandis que l'écart $V_{\max} - V_{\min}$ est appelé gain. Des ajustements sur le gain peuvent parfois permettre de compenser de mauvaises ouvertures de diaphragme. A chaque entier $e \in [1, N - 2]$ est associé l'intervalle $\left[e\gamma - \frac{\gamma}{2}, e\gamma + \frac{\gamma}{2} \right[$ avec $\gamma = \frac{V_{\max} - V_{\min}}{N}$ le pas de quantification. Les intervalles associés à 0 et $N - 1$ sont des intervalles ouverts respectivement à gauche et à droite.

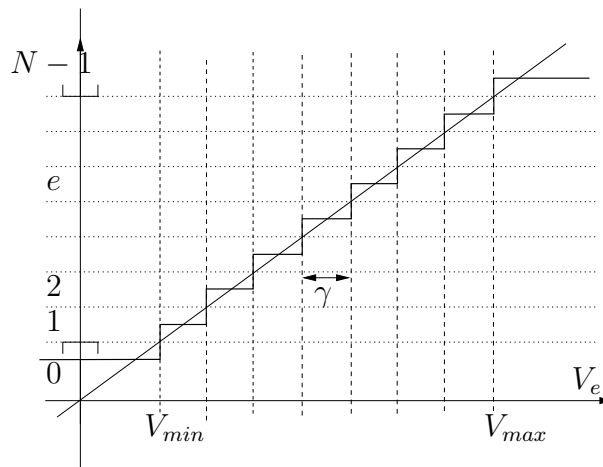


Figure 1.16 – Quantification linéaire.

1.1.3.6 Modèle d'erreur de la quantification des niveaux de gris

La quantification du niveau de gris induit une erreur puisqu'on dispose d'un nombre limité de niveaux différents pour représenter l'amplitude d'un signal continu. La valeur maximale de l'erreur est de $\frac{\gamma}{2}$. On constate que, plus la résolution est importante, plus l'erreur est petite [WK96]. On remplace généralement cette erreur par un bruit de quantification qui est modélisé soit par une distribution uniforme [Kol84] soit par des intervalles d'erreurs [JKDW01].

On peut aussi remarquer que la présence des bornes V_{\min} et V_{\max} provoquent un phénomène de saturation qui est difficilement modélisable par une distribution probabiliste.

1.1.3.7 Modèle d'erreur de la numérisation

Comme tout composant électronique, le C.A.N. est soumis aux bruits et aux aléas électroniques. Le signal de mesure M à quantifier est donc bruité par un bruit B_P par rapport au signal S originel. La quantification rajoute un bruit B_Q sur le signal quantifié M_Q (figure 1.17).

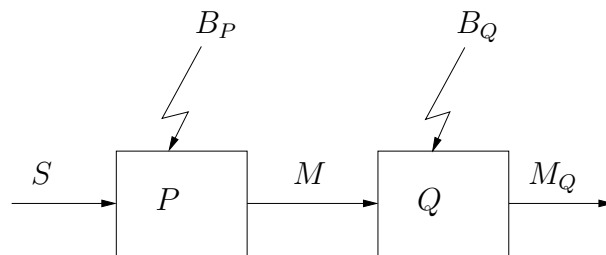


Figure 1.17 – Modèle d'erreur.

Dans le cas où l'on représente le bruit de quantification B_Q par une distribution supposée uniforme, le bruit associé au signal quantifié M_Q est un mélange (Bayes, noyaux, etc.) entre le bruit du signal de mesure B_P et le bruit de quantification B_Q . Dans le cas où l'on représente le bruit de quantification par un intervalle d'erreur, on doit comparer les deux bruits. Si le bruit du signal à quantifier est très inférieur au bruit de quantification, on considère le bruit associé au signal à quantifier négligeable devant le bruit de quantification et on représente le bruit du signal M_Q par un intervalle d'erreur. Si, par contre, les deux bruits sont d'importance équivalente ou si le bruit du signal à quantifier est plus grand que le bruit de quantification, alors il faut mélanger les deux bruits.

1.2 La vision stéréoscopique

A cause de l'aspect projectif de la création des images, il est impossible de recréer la géométrie tridimensionnelle d'une scène à partir d'une seule image [Xie89]. La stéréovision, c'est à dire la perception d'une scène à partir, au minimum, de deux caméras, permet de reconstruire un relief apparent. Dans la plupart des cas, seules deux caméras

sont utilisées. Ito [II86] et Ayache [Aya89] ont proposé d'utiliser un système trinoculaire (trois caméras) afin de mettre à profit la redondance d'informations apportée par la troisième caméra, d'une part pour vérifier les résultats obtenus par le premier couple d'images, et d'autre part afin de lever certaines indéterminations.

Les applications de la stéréovision sont diverses : construction de carte et reconnaissance aérienne [CW81], [Hou93], [BB96], navigation pour la robotique [BFZ93], [Las96], [KOY00], [ML00], reconstruction d'objets [FL95], [SDJ96], modélisation d'environnements [ML04], etc. La reconstruction volumétrique est une autre utilisation des vues multiples [FB03], [SCM⁺04]. Dans [Day00], Dayer recense les méthodes utilisant une approche volumétrique à base de vues multiples pour reconstruire une scène.

1.2.1 Principe de base

Soit le système stéréoscopique représenté sur la figure 1.18. Ce système comporte trois systèmes de coordonnées, deux attachés à chaque plan rétinien $(u_1, v_1)^t$ et $(u_2, v_2)^t$ et un tridimensionnel $(X_w, Y_w, Z_w, 1)^t$. Soit B un point de la scène et b_1 et b_2 ses deux projections dans chaque plan image. La connaissance de la géométrie relative des deux capteurs permet de reconstruire la position tridimensionnelle du point B à partir de la connaissance de ses deux projections. Pour faire cette reconstruction, il est nécessaire d'être capable d'apparier les deux projections.

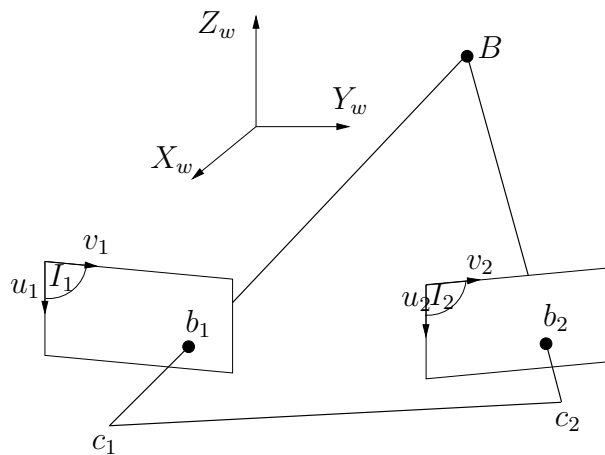


Figure 1.18 – La géométrie d'un capteur stéréoscopique.

1.2.2 Relation entre les projections - Epipolarité

Un point de l'espace tridimensionnel possède trois degrés de liberté. Un point dans un plan possède deux degrés de liberté. Les deux projections b_1 et b_2 du point B dans les deux plans images du système stéréoscopique sont donc liées. Cette relation porte le nom d'épipolarité.

Pour comprendre simplement cette relation, plaçons nous dans le cas représenté par la figure 1.19. Si B a comme projection b_1 dans la première image, il appartient à la

demi-droite Δ_1 passant par c_1 et b_1 . Cette demi-droite s'étend vers l'infini depuis b_1 dans la direction de l'axe optique. Lorsque B parcourt Δ_1 , sa projection sur l'image 2 parcourt aussi une demi-droite qui est la projection perspective de Δ_1 sur l'image 2. Cette droite porte le nom de droite épipolaire.

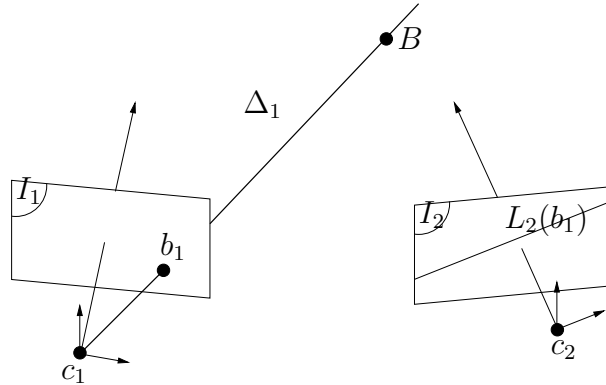


Figure 1.19 – Définition de la contrainte épipolaire.

A chaque point b_1 de l'image 1 correspond une droite épipolaire $L_2(b_1)$ sur l'image 2. Cette droite peut être paramétrée par le vecteur $l_2(b_1)$ tel que si b_2 est un point de l'image 2, b_2 appartient à la droite $L_2(b_1)$ si et seulement si $b_2^t l_2(b_1) = 0$. La projection de la demi-droite Δ_1 dépend de la transformation géométrique entre les deux caméras. La relation entre b_1 et la droite épipolaire $L_2(b_1)$ fait intervenir une matrice appelée matrice essentielle E [HZ00] qui est une matrice 3×3 mais de rang 2 :

$$l_2(b_1) = Eb_1$$

Ce que l'on peut aussi écrire :

$$b_2^t Eb_1 = 0$$

Cette relation est entièrement symétrique. A chaque point b_2 de l'image 2 correspond une droite épipolaire $L_1(b_2)$ sur l'image 1 de paramètres $l_1(b_2)$. La relation entre b_2 et $l_1(b_2)$ fait intervenir la transposée de la matrice essentielle :

$$l_1(b_2) = E^t b_2$$

La relation d'épipolarité relie la projection d'un point tridimensionnel dans une image à une ligne dans le plan de la deuxième image. Dans le cas des images numériques, le travail s'effectue sur les pixels. Soit m_1 un pixel de l'image 1 de coordonnées $(i_1, j_1, 1)^t$ (figure 1.20). Faugeras et son équipe [Fau93] ont montré que la relation entre m_1 et la droite épipolaire $L_2(m_1)$ faisait intervenir une matrice appelée matrice fondamentale F qui est aussi une matrice 3×3 de rang 2 qui, en plus de dépendre de la

transformation géométrique entre les deux caméras, dépend des paramètres d'échantillonnage de l'espace :

$$l_2(m_1) = Fm_1$$

Si m_2 un pixel de l'image 2 de coordonnées $(i_2, j_2, 1)^t$ appartient à la droite $l_2(m_1)$ tel que $m_2^t l_2(m_1) = 0$, alors on peut écrire :

$$m_2^t F m_1 = 0 \quad (1.6)$$

Cette relation est entièrement symétrique. A chaque pixel m_2 de l'image 2 correspond une droite épipolaire $L_1(m_2)$ sur l'image 1 de paramètre $l_1(m_2)$. La relation entre m_2 et $l_1(m_2)$ fait intervenir cette fois-ci la transposée de la matrice fondamentale :

$$l_1(m_2) = F^t m_2$$

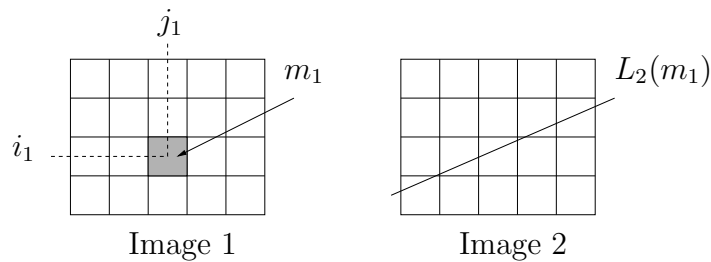


Figure 1.20 – Définition de la contrainte épipolaire du point de vue pixel.

1.2.3 Etalonnage

Pour pouvoir utiliser la relation d'épipolarité entre les projections, la connaissance de la matrice fondamentale F est indispensable. L'estimation de la matrice fondamentale a fait l'objet de très nombreux travaux [XZ96], [HZ00], [Fau93], [Aya89]. Cette étape est connue sous le nom d'étalonnage du capteur stéréoscopique. Elle vise à estimer les paramètres intrinsèques et extrinsèques du système stéréoscopique en utilisant une mire dont on connaît le modèle.

Le principe est le suivant : dans chaque image, on détermine la position des éléments caractéristiques composant la mire. Ensuite, ces éléments sont appariés en utilisant l'une des méthodes classiques de mise en correspondance d'indices caractéristiques (méthodes qui seront présentées dans la suite de ce chapitre). La phase d'appariement est facilitée par la connaissance de relations géométriques entre les éléments de la mire. Enfin, on cherche à résoudre le système surdéterminé formé par la relation d'épipolarité 1.6.

Le nombre minimum de point requis pour établir un ensemble de solutions est de cinq points [FM90]. Il existe une méthode de calcul de solution unique proposée par

Longuet-Higgins [LH81] basée sur huit points mis en correspondance. Usuellement, on utilise plus de huit points afin d'augmenter la robustesse du calcul en résolvant l'équation 1.6 par des méthodes basées sur les moindres carrés [HZ00].

Depuis quelques années, un certain nombre de méthodes ont été proposées pour effectuer l'étalonnage du système stéréoscopique sans utiliser de mire [LF93], [LF97], [XZ96]. Connu sous le nom d' "auto-calibration", ces méthodes estiment tout d'abord les paramètres intrinsèques d'une caméra par l'estimation de la matrice fondamentale associée au déplacement de la caméra. Généralement, trois déplacements suffisent pour trouver une solution. L'estimation se base sur l'appariement de points d'intérêt préalablement extraits. Ensuite, grâce à deux déplacements supplémentaires du capteur stéréoscopique, les paramètres extrinsèques (relation géométrique entre les deux caméras) sont estimés.

1.2.4 Reconstruction

Reconstruire les coordonnées d'un point tridimensionnel nécessite la connaissance des coordonnées de ses projections dans les différents plans images, des paramètres de projection, de changement de repère et des paramètres d'échantillonnage de l'espace.

La reconstruction s'obtient par triangulation. En reprenant l'exemple de la figure 1.18, le point B se trouve à l'intersection des deux droites (c_1b_1) et (c_2b_2) . Or, les seules informations disponibles sont les coordonnées des pixels supposés correspondants m_{b_1} et m_{b_2} . L'imprécision de la position de b_1 et b_2 induit une imprécision dans la position estimée du point B . En effet, le point b_1 (resp. b_2) appartient au pixel m_{b_1} (resp. b_2). Si on considère les positions de m_{b_1} et m_{b_2} comme précises, la plupart du temps $(c_1m_{b_1})$ et $(c_2m_{b_2})$ ne s'intersectent pas (ceci est due à l'échantillonnage de l'espace).

En utilisant l'équation 1.3 du modèle de caméra et I_{c_x} la matrice des paramètres intrinsèques de la caméra x définie par les équations 1.4 et 1.5, on peut établir les relations suivantes :

$$\begin{cases} m_{b_1} = I_{c_1}PK_1B = A_1B \\ m_{b_2} = I_{c_2}PK_2B = A_2B \end{cases}$$

On peut écrire le système reliant le point B avec les pixels dans lesquels il s'est projeté dans les différents plans image par :

$$B = \left[\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}^t \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \right]^{-1} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}^t \begin{pmatrix} m_{b_1} \\ m_{b_2} \end{pmatrix}$$

Ce système est sur-déterminé : B est définie par trois inconnues et chacune de ses projections sont définies par deux variables soit quatre au total dans le cas d'un système stéréoscopique à deux caméras. Il est généralement résolu par minimisation d'un critère de distance. Par exemple, nous pouvons citer le critère basé sur la somme des écarts absolus entre m_{b_i} les coordonnées de la projection du point B dans le plan

image i et \widehat{m}_{b_i} les coordonnées de la projection de l'estimation du point B dans le plan image i ($c = \sum_{i=1}^2 |\widehat{m}_{b_i} - m_{b_i}|$).

1.2.5 Restriction géométrique des mises en correspondance

Sans autre information (si le système stéréoscopique n'est pas étalonné et si on ne s'intéresse pas aux niveaux de gris des pixels), le correspondant du pixel m_{b_1} de l'image 1 peut être n'importe quel pixel appartenant à l'image 2. Pour pallier cette difficulté, on utilise un certain nombre de contraintes pour réduire le nombre de correspondants potentiels. Ces contraintes sont de deux natures différentes : géométrique due à la nature du capteur stéréoscopique et physique due aux propriétés d'illumination. Dans la suite de ce paragraphe, nous étudions seulement les contraintes géométriques.

1.2.5.1 La contrainte épipolaire

Cette contrainte géométrique nous permet de réduire l'espace de recherche du correspondant à une seule dimension. Grâce à la géométrie épipolaire de l'équation 1.6, le correspondant du point m_{b_1} dans l'image 2 est astreint à la ligne épipolaire $l_{m_{b_1}}$.

1.2.5.2 La contrainte d'ordre

En vision stéréoscopique, on considère usuellement que l'ordre des projections dans les différents plans images reste le même. On peut effectivement constater sur la figure 1.21a que l'ordre des projections dans le plan image 1 est $[a_1, b_1, c_1]$ et que l'ordre des projections dans le plan image 2 est $[a_2, b_2, c_2]$. Dans leur livre [HM95], Horaud et Monga ont démontré que si la scène observée contient des objets ayant des surfaces transparentes fortement inclinées par rapport au plan des images, l'ordre des projections est inversé. Sur la figure 1.21b, on peut constater que l'ordre des projections dans l'image 1 $[a_1, b_1]$ est l'inverse de l'ordre des projections dans le plan image 2 $[b_2, a_2]$. On peut aussi constater que si l'objet 2 avait été opaque alors le point A n'aurait pas été visible dans l'image 1 (occultation), et la contrainte d'ordre n'aurait pas été violée.

Il existe un deuxième type de scène observée pour lequel l'ordre des projections est inversé entre les deux plans images : c'est lorsque les distances entre les objets et le capteur stéréoscopique sont très différentes (figure 1.22a).

On constate sur cet exemple que la différence de distance entre les points A et B et le capteur stéréoscopique étant très faible, l'ordre des projections n'est pas changé ($[a_1, b_1]$ et $[a_2, b_2]$). Par contre, le point C étant très loin du capteur stéréoscopique (par rapport aux points A et B), l'ordre des projections est inversé ($[c_1, (a_1, b_1)]$ et $[(a_2, b_2), c_2]$). Partant de ces observations, Faugeras [Fau93] a défini un cône représentant une région interdite définie à partir du point A . Cette région modélise le fait que tout point appartenant au cône aura l'ordre de ses projections inversé dans les deux plans images par rapport aux projections du point A .

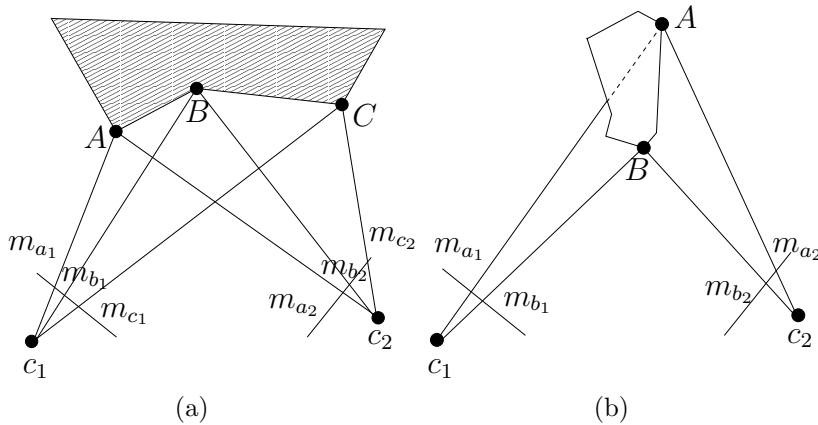


Figure 1.21 – Définition de la contrainte d'ordre (a), Cas d'une surface transparente (b).

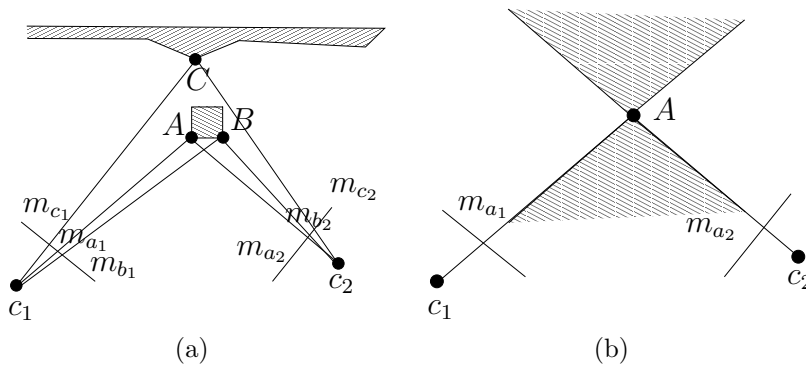


Figure 1.22 – Cas de non-respect de la contrainte d'ordre (a), Cône interdit (b).

1.2.5.3 La contrainte d'unicité

Si on se limite à observer des objets opaques dont la disparité n'est pas très forte, alors un objet dont la projection est un point sur la première image a une projection qui est aussi un point sur la seconde image. Ceci n'est pas vrai pour des objets transparents. Cette contrainte est un cas limite de la contrainte d'ordre (figure 1.23). Les points A et C se projettent en un même point sur l'image 1 et en deux points dans l'image 2. Cette contrainte peut être utilisée afin de réduire le nombre de correspondances possibles [CHR96], [OK85].

Cette contrainte étant un cas particulier de la contrainte d'ordre, elle peut être aussi violée. Ce cas est représenté sur la figure 1.24. Faugeras [Fau93] a démontré que si le bord d'un objet est confondu avec le bord du cône représentant la région interdite, la contrainte d'unicité peut être violée.

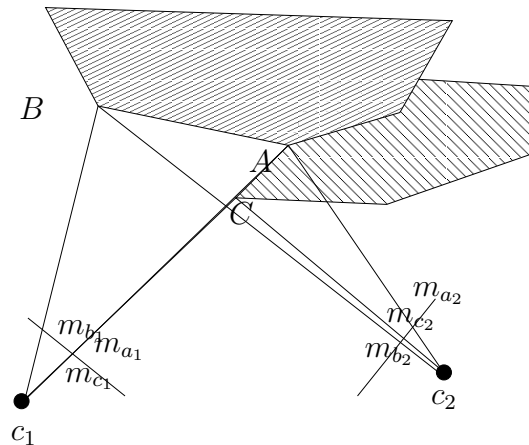


Figure 1.23 – Définition de la contrainte d'unicité.

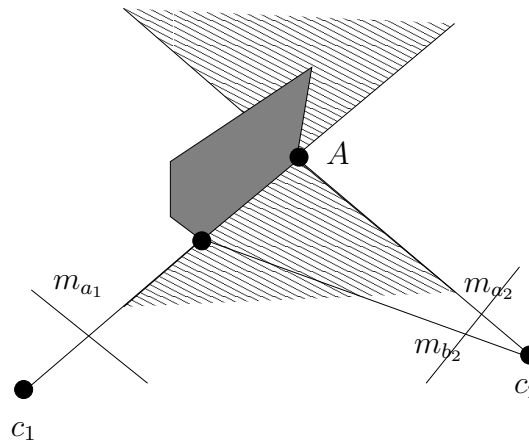


Figure 1.24 – Cas de non-respect de la contrainte d'unicité.

1.2.6 Cas particulier des images redressées

Lorsqu'un capteur stéréoscopique est correctement étalonné, c'est à dire lorsque l'on connaît la géométrie relative des deux caméras et les paramètres qui gouvernent la projection des images, il est possible de "redresser" les images [Aya89]. Redresser les images consiste en la création, par interpolation, des deux images qu'aurait acquies un capteur stéréoscopique si les deux caméras étaient disposées de telle façon que leurs axes optiques (z_1 et z_2) soient parallèles et que la droite (c_1, c_2) soit confondue avec les axes horizontaux (y_1 et y_2) (figure 1.25).

Lorsqu'une paire d'images est redressée, la contrainte épipolaire se simplifie : à toute ligne i_1 de l'image 1 correspond une ligne i_2 . Il est même possible de redresser les images de façon à ce que les deux lignes soient identiques sur les deux images.

Enfin, si m_1 est un pixel de l'image 1 de coordonnées (i_1, j_1) , m_2 appartient forcément à un segment $[a_2, m_{21}]$ de la ligne épipolaire de coordonnée i_2 de l'image 2. a_2 est l'intersection de cette ligne avec le bord gauche de l'image et m_{21} est le point de

coordonnée (i_2, j_1) . Cette contrainte géométrique est symétrique.

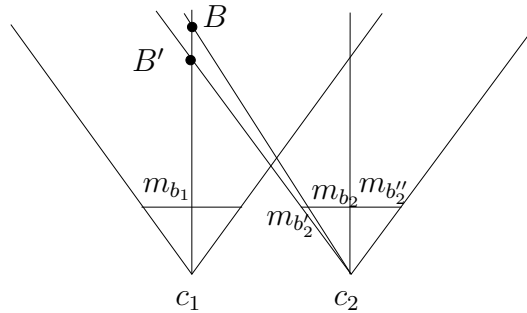


Figure 1.25 – Images stéréoscopiques redressées.

1.3 Mise en correspondance

Le problème de la mise en correspondance stéréoscopique est un sujet de recherche complexe qui occupe la communauté scientifique depuis plus de quarante ans [Jul62]. Marr et Poggio [MP79], en 1979, ont identifié l'ambiguïté naturelle du problème de la mise en correspondance. Ils ont défini deux contraintes permettant de réduire cette ambiguïté qui sont l'unicité et la continuité. Nous avons défini l'unicité au paragraphe précédent. La contrainte de continuité s'énonce ainsi :

si la matière est séparée en objets dont les surfaces sont suffisamment lisses, alors la disparité (variation de position de la projection sur les deux images) est une fonction continue par morceaux.

Dans [MP79], Marr et Poggio ont proposé de baser leur technique de mise en correspondance sur une étude de la stéréoscopie humaine. Dans cette approche, seuls sont appariés des indices faiblement ambigus tels que les bords et les coins. Les cartes de disparités produites par cette technique ont le désavantage d'être éparées. Au début des années 1990, les recherches se sont focalisées sur des approches différentes de la théorie de Marr avec comme objectif d'obtenir une carte de disparité dense.

La mise en correspondance de deux indices (pixel, coin, imagerie, ...) peut être vue comme un algorithme de minimisation du coût d'une fonction de la disparité. Cette fonction d'appariement base son coût sur une distance à la fois chromatique et géométrique. De façon plus explicite, la distance de deux indices est liée à la dissemblance de leur fonction d'illumination locale ainsi qu'aux contraintes géométriques liées à la disposition relative des deux caméras (épipolarité).

Les algorithmes de mise en correspondance stéréoscopique ont été caractérisés suivant les critères qu'ils emploient [DA89], [Zha93], [MDG03]. On peut définir deux familles de contraintes : les contraintes locales qui portent sur un petit nombre de pixels autour d'un pixel d'intérêt et les contraintes globales qui portent sur une ligne de recherche ou sur l'image toute entière. Les méthodes utilisant des contraintes locales, aussi appelées méthodes locales, poursuivent la théorie de Marr et se concentrent sur l'amélioration des techniques de sélection du correspondant en tentant d'augmenter

les caractères discriminants des éléments remarquables. Les méthodes utilisant des contraintes globales, aussi appelées méthodes globales, tentent de définir un modèle global de la scène observée et de minimiser une fonction de coût globale. En marge de ces deux catégories, il existe un certain nombre d'algorithmes utilisant ces deux approches.

1.3.1 Les problèmes de la mise en correspondance

1.3.1.1 Variation du niveau de gris

Marr et Poggio [MP77] ont établi que les projections d'un même point tridimensionnel devaient avoir des intensités semblables. Cette contrainte physique impose l'utilisation d'un modèle d'illumination ainsi que d'un modèle de réflectance des surfaces des objets. Le modèle le plus simple et le plus utilisé est le modèle Lambertien [Hor86]. Ce modèle explicite le fait que la surface des objets reflète la lumière identiquement suivant toutes les directions. Malgré l'utilisation d'un tel modèle, les projections d'un même point tridimensionnel dans chaque image peuvent avoir des intensités différentes. Les sources de variation de niveau de gris sont multiples : aléas lumineux, échantillonnage de l'espace, quantification du niveau de gris, sensibilités différentes entre les caméras utilisées. De plus, certains types de surfaces mettent le modèle de réflectance Lambertien en échec.

1.3.1.2 Variation du nombre de pixel de la projection d'un même objet dans les deux images

En fonction de la différence de point de vue entre chaque caméra, la projection d'un même objet dans les deux images peut être représentée par un nombre de pixels différents. C'est une conséquence directe de l'échantillonnage de l'espace. La figure 1.26a représente ce cas. On constate que, dans l'image 1, la projection de la ligne définie entre les points A et C est représentée par 5 pixels alors que, dans l'image 2, elle n'est représentée que par 3 pixels. Or ces pixels représentent le même objet. La mise en correspondance pixel à pixel est alors impossible.

1.3.1.3 Le problème de l'occultation

Certains points présents dans une image peuvent être absents dans l'autre image. C'est à dire qu'une partie de la scène observée n'est visible que par une seule caméra : il s'agit d'un problème d'occultation. Nous avons représenté un tel cas sur la figure 1.26b. On peut constater que le point B n'est vu que par la caméra 1. Ce phénomène de disparition/apparition de points pose énormément de problèmes lors du calcul des coûts d'appariement. Il existe trois manières de traiter ce problème :

- La première est de détecter les régions occultées avant ou après la mise en correspondance [HA89], [Wil91]. Dans le cas de mise en correspondance dense, on remplace la zone occultée par une interpolation basée sur la disparité du voisinage de la zone. Dans le cas de mise en correspondance éparse, la zone est toute

simplement exclue de l'algorithme de calcul des disparités.

- La deuxième façon est de réduire la sensibilité des algorithmes aux occultations. Sara [SB97] et Stewart [Ste97] ont proposé d'utiliser des algorithmes robustes pour réduire la sensibilité de la mise en correspondance aux occultations, aux bruits et aux distorsions géométriques. Zabih et Woodfill [ZW94] et Bahat et Nayar [BN98] ont proposé d'effectuer des transformations non paramétriques appliquées aux intensités des pixels avant d'utiliser des mesures de corrélation. Enfin, Kanade et Okutomi [KO94] et Mansouri et Mitiche [MM98] ont proposé de modifier la taille et la forme de la zone de recherche pour optimiser la mesure de ressemblance au voisinage d'une occultation.
- La troisième façon est de calculer un modèle géométrique d'occultation afin de l'intégrer au processus d'appariement [Bel96], [BT98a].

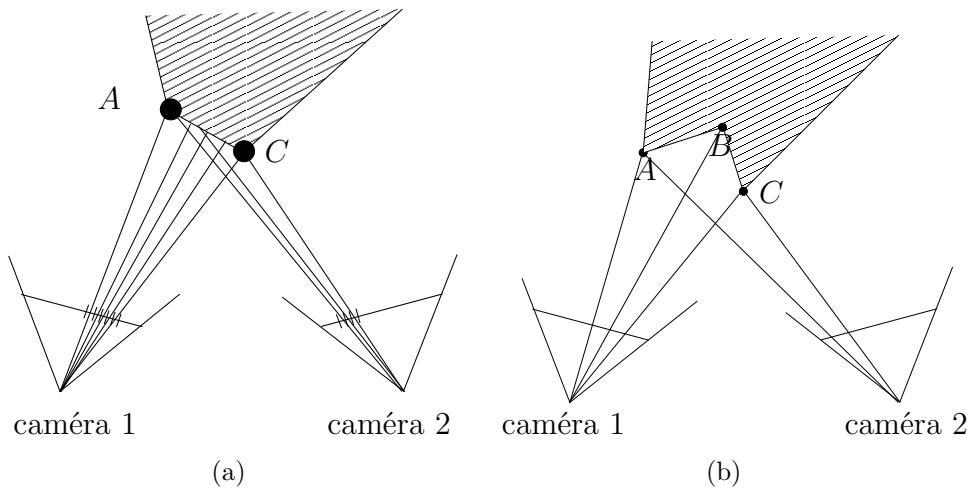


Figure 1.26 – Variation du nombre de pixel (a), problème d'occultation (b).

1.3.1.4 Violation des contraintes géométriques d'appariements

Nous avons vu dans le paragraphe précédent qu'il existe des contraintes dépendant de la nature de la scène et des objets observés : il s'agit de la contrainte d'unicité et de la contrainte d'ordre. Ces contraintes ne sont pas violées si la scène ne contient pas d'objet ayant des surfaces fortement inclinées. Il se peut néanmoins que certaines scènes observées contiennent ce genre de surfaces. Dans ces cas là, ces contraintes ne sont plus respectées faussant ainsi les résultats des algorithmes de mise en correspondance.

1.3.1.5 Les zones peu texturées

Le fait de vouloir mettre en correspondance deux zones de pixels appartenant à différentes images impose à ces zones d'être texturées. En effet, elles doivent offrir un pouvoir suffisamment discriminant pour que la mesure de distance entre les deux zones soit significative. Sur la figure 1.27 sont représentés les niveaux de gris de deux images. Que l'on s'intéresse au niveau de gris d'un pixel (a), d'un groupe de quatre pixel (b) ou

de neuf pixels (c), on est incapable de déterminer un correspondant dans l'autre image du fait que la texture est trop uniforme entre les deux images.

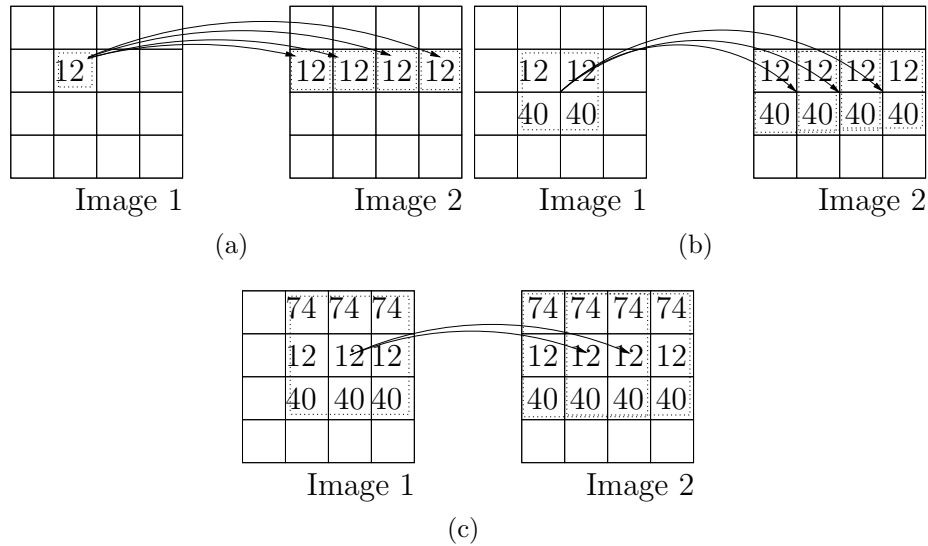


Figure 1.27 – Texture uniforme.

1.3.2 Que met-on en correspondance ?

Dans la plupart des travaux qui composent notre bibliographie se dégagent deux approches d'appariements : les méthodes qui appariant des pixels et qui sont donc très proches de la structure échantillonnée des images, et les méthodes qui appariant des indices caractéristiques. Un indice caractéristique est un élément de structure de l'image dont la signature d'illumination présente peu d'ambiguïté pour l'appariement comme pour la localisation. Cette séparation en deux classes est plus ou moins historique et ne reflète pas les approches actuelles. Tout d'abord, un pixel est aussi un indice caractéristique. Ensuite, il y a un certain nombre de nouveaux types d'indices caractéristiques récemment proposés ayant des propriétés intéressantes. Ces éléments ne sont plus des coins, des bords ou des lignes mais des points [Sch96] ou des régions [TG00] aux caractéristiques invariantes aux transformations géométriques entre les deux caméras.

1.3.2.1 Mise en correspondance basée sur les indices caractéristiques

Les méthodes de mise en correspondance basées sur les indices caractéristiques font appel à l'extraction d'éléments remarquables dans les images suivie d'une mise en correspondance. Ces méthodes nécessitent une interprétation de l'information contenue dans les images. D'après Marr [MP79], un point caractéristique est un lieu (dans l'image) où il y a une discontinuité de la variation d'intensité. Ces indices sont en fait des éléments remarquables dans les images 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 ces indices est qu'ils ont des propriétés qui se conservent

entre les deux images. L'inconvénient est que le nombre de ces indices est en général très restreint générant au final une carte des disparités éparées.

Le principe des méthodes de mise en correspondance stéréoscopiques basées sur les indices caractéristiques repose sur deux étapes : tout d'abord, il faut extraire les indices caractéristiques, ensuite il faut les mettre en correspondance.

Les indices sont généralement classés en deux groupes :

Le premier groupe est composé des indices du type contours. Les méthodes les plus utilisées pour les détecter sont les méthodes dérivatives [HM95]. Elles sont divisées en trois catégories :

- Les méthodes utilisant le calcul du gradient. Dans un premier temps, ces méthodes déterminent les gradients directionnels de l'image en chaque point. Ensuite, elles calculent la norme du gradient. Enfin, le calcul du maximum local de la norme du gradient permet de définir les points du contour.
- Les méthodes basées sur le calcul des dérivées secondes. Ces méthodes utilisent la norme du gradient pour calculer la dérivée seconde dans la direction du gradient. Les points du contour sont obtenus en cherchant les passages par zero de la dérivée seconde.
- Les méthodes reposant sur le calcul du Laplacien. Ces méthodes sont basées sur le calcul du Laplacien de l'image lissée et sur la recherche de ses passages par zero.

Toutes ces méthodes substituent à l'image en niveau de gris une image binaire (noir et blanc) contenant uniquement les contours détectés. Puisque les images ne contiennent que des contours, l'identification de droites ou de segments est simplifiée.

Le deuxième groupe d'indices caractéristiques concerne les points d'intérêts. L'inconvénient des contours est leur imprécision de localisation. Les points d'intérêts correspondent à des caractéristiques bidimensionnelles telles que la plus forte courbure, les coins, les jonctions entre les segments, etc. Ces points sont beaucoup moins nombreux que les points de contours mais ils sont plus fiables.

Les principaux détecteurs de points d'intérêts sont ceux de Canny [Can86], de Marr et Hildreth [MH80] ou de Harris [HS88]. Le lecteur pourra trouver d'autres détecteurs dans les documents suivants : [Sch96], [DF90], [HM95].

Une fois les indices caractéristiques obtenus, il faut les mettre en correspondance entre les différentes images. A chaque indice caractéristique on associe un vecteur d'attributs à partir duquel sera calculé l'appariement. Les méthodes de mise en correspondance basées sur les indices caractéristiques seront présentées dans le paragraphe suivant concernant les méthodes de mise en correspondance locales.

1.3.2.2 Mise en correspondance basée sur les pixels

L'objet de ce type de méthode est de produire des cartes de disparité denses. Le coût d'appariement de deux pixels est basé sur une mesure de similarité de la fonction d'illumination locale. Pour être le plus proche possible de la structure échantillonnée

de l'image, ce sont les pixels de l'image 1 que l'on tente d'apparier avec les pixels de l'image 2.

La méthode la plus simple consiste à comparer les intensités des pixels [CHR96]. Mais cette mesure est très sensible aux bruits ainsi qu'aux distorsions de l'image et présente beaucoup d'ambiguïtés d'appariement. Des approches plus fiables utilisent des fenêtres de comparaison contenant le voisinage du pixel considéré. La similitude entre deux fenêtres est obtenue par une mesure statistique de la distance entre les deux fonctions d'illuminations échantillonnées. Les mesures les plus couramment utilisées sont la somme des écarts quadratiques (Sum of Squared Differences ou S.S.D.) [CHR96], [OKO01], la somme des écarts absolus (Sum of Absolute Differences ou S.A.D.) [Hir01], l'intercorrélation normalisée (Normalized Cross-Correlation ou N.C.C.) [CM99], [Sar02] ou les méthodes de rangs et de recensements [BN98].

Les méthodes basées sur la mise en correspondance de pixels sont très sensibles aux occultations, aux défauts de réflexions et aux variations d'illuminations entre les différentes vues. De plus, l'apparition d'artefacts provenant de l'échantillonnage de l'image peut perturber la mesure de ressemblance. L'évaluation d'un coût de mise en correspondance défini sur une fenêtre peut alors assigner une mauvaise valeur d'appariement à deux pixels correspondant, même dans des images ne contenant pas d'effets dégradants tels que des occultations ou des surfaces non-uniformes. Peu d'articles ont portés sur ces problèmes.

Dans [BT98b], Birchfield et Tomasi ont proposé une méthode insensible à l'échantillonnage de l'image. La mesure de similarité utilise des fonctions d'interpolations linéaires basées sur l'intensité du voisinage des pixels dans les deux images. Les coûts de mise en correspondance sont évalués sur des positions de pixels entières et sur des demi-pixels (là où la fonction de l'image est interpolée de façon symétrique dans chaque image).

Szeliski et Scharstein [SS02] ont proposé une méthode se basant sur les travaux de Birchfield et Tomasi [BT98b] mais utilisant une interpolation sur les images en vue d'obtenir des fonctions continues (et ainsi diminuer la sensibilité à l'échantillonnage). A l'inverse de [BT98b], le coût de la mise en correspondance est évalué sur un espace de disparité continu.

Plus récemment, Clerc [Cle02] a proposé de représenter le signal continu de l'image par des coefficients d'ondelettes calculés sur chaque pixel de l'image. De ce fait, les coûts de mise en correspondance sont calculés sur ces coefficients. Cette méthode permet de corriger les appariements même dans les structures répétitives de taille différentes dans les deux images.

1.3.3 Les méthodes globales

Une méthode d'appariement est dite globale lorsque la fonction de coût est évaluée sur l'ensemble de l'image. L'utilisation de ce type de méthode implique la définition d'un modèle de la scène observée pour régulariser (contraindre) l'ensemble des ap-

pariements. Certaines méthodes s'appuient sur la contrainte épipolaire (stéréovision calibrée) pour ramener ce problème bidimensionnel à un problème monodimensionnel [Bel96], [CHMR92]. Tandis que d'autres méthodes traitent directement du problème bidimensionnel [BVZ99], [IG98].

La régularisation globale a pour but de réduire la sensibilité des algorithmes aux ambiguïtés provoqués par des occultations, une trop faible texture locale ou des défauts d'illumination. Le prix à payer pour cette amélioration est un accroissement de la complexité des calculs ainsi qu'une sensibilité du modèle de régularisation.

1.3.3.1 Programmation dynamique

La programmation dynamique est une méthode mathématique qui réduit la complexité de calcul d'un algorithme d'optimisation en le décomposant en sous systèmes plus simples. Appliquée à la mise en correspondance stéréoscopique, cette méthode cherche un chemin de coût minimum à travers une matrice composée de tous les appariements possibles. Dans la totalité des travaux que nous avons référencé, cette technique est utilisée pour une mise en correspondance restreinte par la contrainte épipolaire. Le problème de la mise en correspondance se réduit à un problème de mise en correspondance des pixels d'une ligne de l'image 1 avec ceux de la ligne de l'image 2 qui est son épipolaire.

Pour que les techniques de programmation dynamique soient envisageables, il est nécessaire de supposer que les coûts sont additifs, c'est à dire que le coût du chemin global est la somme des coûts des chemins partiels obtenus récursivement. Le coût local pour chaque point dans l'espace de recherche est défini en utilisant une méthode de mise en correspondance locale (S.A.D., S.S.D., etc.). Les occultations sont modélisées en assignant un groupe de pixels dans une image à un seul pixel dans l'autre image et en pénalisant cette solution par un coût d'occultation appliqué au coût global du chemin. On peut voir sur la figure 1.28 une représentation de la matrice de recherche ainsi qu'une modélisation de l'espace monodimensionnel des disparités. Les axes sont définis par les lignes de recherche des images droite et gauche comme proposé par Ohta et Kanade [OK85] ou Cox et al. [CHR96]. Les problèmes liés à la formulation de la mise en correspondance grâce à la programmation dynamique sont le choix du coût d'une occultation, la difficulté de garder une consistance inter-ligne de recherche [BI99], [OK85] et le respect des contraintes d'ordre et de continuité. Avec N pixels composant une ligne de recherche, la complexité de calcul en utilisant la programmation dynamique est en $O(N^4)$ à laquelle il faut ajouter le temps requis pour les fonctions des coûts locaux.

Il existe un certain nombre de variations en vue de diminuer la complexité du calcul et de réduire les ambiguïtés de mise en correspondance. Par exemple, Baker et Binford [BB81] proposent de calculer la disparité indépendamment pour chaque ligne d'une paire d'images stéréoscopiques redressées [Fau93]. Ils cherchent et corrigent les mises en correspondance qui violent la contrainte d'ordre. Ohta et Kanade [OK85] ont proposé d'intégrer à cette méthode des contraintes de continuités verticales entre les

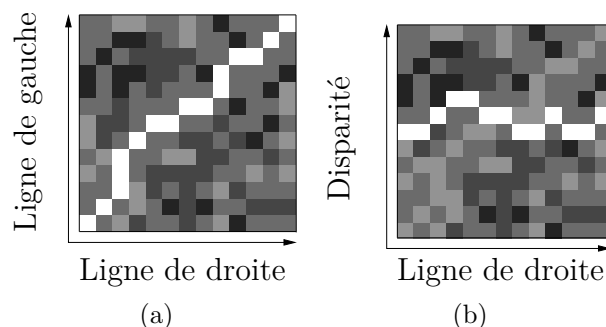


Figure 1.28 – Exemple d’une représentation de l’espace des disparités en représentant les lignes droite-gauche (a), ou en utilisant la représentation ligne droite-disparité (b). L’intensité représente le coût de la mise en correspondance potentielle le long de la ligne de recherche, un pixel blanc représentant un faible coût.

différentes lignes épipolaires.

Dans [CHMR92], [Cox94], [CHR96], Cox présente un algorithme de vraisemblance maximum qui respecte les contraintes d’ordre et d’unicité. Le coût local dépend du coût de la mise en correspondance des pixels ainsi que d’un coût de pénalité fixe pour les pixels non appariés. Les résultats obtenus dans [CHMR92] montrent qu’il existe plusieurs minima globaux de la fonction de coût résultant de l’accumulation de plusieurs hypothèses. Dans [CHR96], Cox propose un algorithme amélioré qui recherche la solution avec le minimum de discontinuités possible dans les directions verticales et horizontales de l’espace des disparités, se conformant ainsi à la contrainte de continuité de la matière.

Remarquant qu’une discontinuité de la disparité le long d’une ligne épipolaire était la signature d’une occultation, Geiger et al.[GLY95] ont mis à profit cette propriété pour définir un nouveau modèle d’occultation. Le coût d’une mise en correspondance est évalué sur des fenêtres paramétrables permettant de s’adapter aux variations d’intensité et de localisation. La représentation de la fonction de disparité optimale est obtenue par l’estimation du maximum a posteriori en utilisant la programmation dynamique.

Belhumeur [Bel96] a développé un modèle basé sur un réseau Bayésien où les régions occultées sont explicitement représentées et calculées. Dans cet article, Belhumeur présente un certain nombre de modèles de scène, des configurations les plus simples aux scènes les plus complexes. Il discute sur le fait que les modèles fiables ne contiennent pas seulement des profondeurs simples mais aussi des discontinuités de profondeur localisées sur les contours des objets ainsi que des surfaces bosselées différemment inclinées. Ce modèle est utilisé pour calculer la fonction de disparité optimale par l’estimation du maximum a posteriori. Belhumeur présente aussi une nouvelle stratégie de programmation dynamique qui calcule simultanément la disparité, les régions occultées et les orientations des surfaces.

Dans [BI99], Bobick et al. ont présenté un algorithme qui recherche les apparie-

ments et les régions occultées simultanément, mais sans que les contraintes de surface lisse ou de continuité soient utilisées. Ils ont introduit une structure de donnée appelée "image de l'espace des disparités" (Disparity Space Image ou D.S.I.), dans laquelle on recherche le meilleur chemin par programmation dynamique. La sensibilité aux coûts d'occultation ainsi que la complexité de calcul sont réduits grâce à l'utilisation d'éléments de mise en correspondance fiables qui orientent la recherche de la solution. Cette méthode exploite la relation entre la largeur de la zone interdite d'une occultation dans le D.S.I. avec l'intensité des bords des objets. Cette méthode permet d'obtenir des résultats relativement fiables malgré la présence de très grandes régions d'occultation.

L'un des principaux avantages de la mise en correspondance en utilisant la programmation dynamique est de fournir un support global pour des régions qui sont localement faiblement texturées et qui autrement seraient appariées de façon incorrecte. L'utilisation de cette méthode permet aussi de résoudre les problèmes d'appariement liés aux occultations. Par contre, les pixels, au voisinage d'une occultation, posent problème. En effet, le coût d'une mise en correspondance pour des pixels proches d'une occultation est élevé. Des méthodes pour pallier à ces difficultés ont été proposées dans [BT98a]. Ces méthodes remplacent le coût d'une mise en correspondance dans le voisinage d'une discontinuité par un coût d'occultation fixe.

Le principal inconvénient des méthodes par programmation dynamique, en plus de la complexité du calcul, est qu'une erreur locale peut être propagée tout le long de la ligne de recherche corrompant ainsi d'autres mises en correspondance potentiellement correctes.

1.3.3.2 Théorie des graphes

Si certaines méthodes utilisant la programmation dynamique tentent d'imposer une certaine régularité dans la fonction de disparité dans toutes les directions, la plupart n'exploitent pas totalement la cohérence bidimensionnelle. Ce défaut vient du fait que la majorité des méthodes utilisant la programmation dynamique appariant les pixels appartenant à la même ligne épipolaire sur les deux images sans prendre en compte une éventuelle continuité de l'image tridimensionnelle à reconstruire. La théorie des graphes permet de généraliser cette technique en deux dimensions. Ce type de méthode a été proposé par Roy et Cox [RC98] puis a été formalisé par Veksler [Vek99] et Kolmogorov et Zabih [KZ01], [KZ02a], [KZ02b].

Le principe de flot dans un graphe peut être vu comme un problème d'écoulement d'eau dans un réseau de tuyaux. Considérons une source d'eau s de débit infini, un puits t de contenance infinie et un réseau de tuyaux reliant la source au puits. Le flot maximum que l'on peut faire passer dans ce réseau de tuyaux est contraint par le réseau. Certains de ces tuyaux séparant la source du puits se comportent comme un goulot d'étranglement limitant à eux seuls l'ensemble du flot. C'est l'ensemble de ces tuyaux que l'on nomme "goulot d'étranglement" ou "coupure minimale" et la somme de leur capacité est appelé "capacité minimale". Si la capacité d'un de ces tuyaux augmente, la capacité minimale augmente et donc le flot maximum augmente. La valeur du flot

maximum est égale à la coupure minimale [FF62]. Donc trouver l'ensemble des tuyaux réalisant le goulot d'étranglement minimum est un problème analogue à celui consistant à trouver la valeur du flot maximum.

Lorsqu'il s'agit d'utiliser ce type de méthode pour réaliser une mise en correspondance de pixels, le graphe représente un réseau reliant une source et un puits par l'intermédiaire de tous les pixels d'une des images stéréoscopiques. Chaque "tuyau" est segmenté en autant de tronçons que de disparités envisageables. La capacité de chaque tronçon représente une fonction de coût d'attribution de la disparité envisagée au pixel auquel le tronçon est relié. Chacun des tronçons est relié aux tronçons voisins par des tuyaux dont la capacité d'écoulement permet de contraindre l'ensemble du réseau.

Pour expliquer simplement cette approche, nous allons présenter l'exemple à une seule dimension proposé par Paris et Sillion [PS02]. Soit trois pixels m_1 , m_2 et m_3 appartenant à l'image 1 et une fonction f représentant la fonction de disparité associée à chaque pixel qui peut prendre deux valeurs discrètes d_1 et d_2 . Associer une disparité à un pixel de l'image 1 revient à apparier le pixel de l'image 1 avec un pixel de l'image 2. Pour définir une énergie sur f , on introduit une fonction de coût $c(m, d) > 0$ qui représente l'énergie du choix $f(m) = d$ (c'est à dire le fait d'associer au pixel m la disparité d) et une fonction de pénalité $p(f(m_i), f(m_{i+1}))$ pour $i \in \{1, 2\}$ qui représente notre souhait d'avoir des valeurs de f similaire pour des points proches (cette fonction permet de modéliser la contrainte de continuité de la matière). $p(f(m_i), f(m_{i+1})) = 0$ si $f(m_i) = f(m_{i+1})$ et vaut $p_0 > 0$ sinon.

On cherche f qui minimise :

$$\varepsilon(f) = \sum_{i=1}^3 c(m_i, f(m_i)) + \sum_{i=1}^2 p(f(m_i), f(m_{i+1}))$$

On construit un graphe tel qu'indiqué sur la figure 1.29a. La base du graphe est une grille dont les axes verticaux représentent chaque pixel m_i et les axes horizontaux chaque niveau de disparité possible d_i . A chaque position (m, d) on associe un arc vertical avec la capacité $c(m, d)$, les arcs horizontaux correspondent à la fonction de pénalité p_0 . Les arcs qui relient la source à la grille et la grille au puits ont une capacité infinie (figure 1.29b). Paris et Sillion ont démontré dans [PS02] qu'en calculant la coupure minimale de ce graphe (figure 1.30), on trouve la fonction $f_0 = \arg \min_f \varepsilon(f)$. Dans cette même publication, ils expliquent le passage à des cas en deux et trois dimensions.

La première utilisation de la théorie des graphes pour le problème de la mise en correspondance stéréoscopique a été proposée par Roy et Cox [RC98], [Roy99]. Dans cette approche, le graphe de représentation du problème (figure 1.31) est très proche de celui utilisé par Bobick et al. pour représenter l'espace des disparités (D.S.I.). Le maillage de base est un maillage 3D composé de points (i, j, d) , où (i, j) sont les coordonnées des pixels d'une des images et d les disparités possibles, d'une source s et d'un puits t . En interne, le maillage est connecté par des arcs dans les six directions permettant

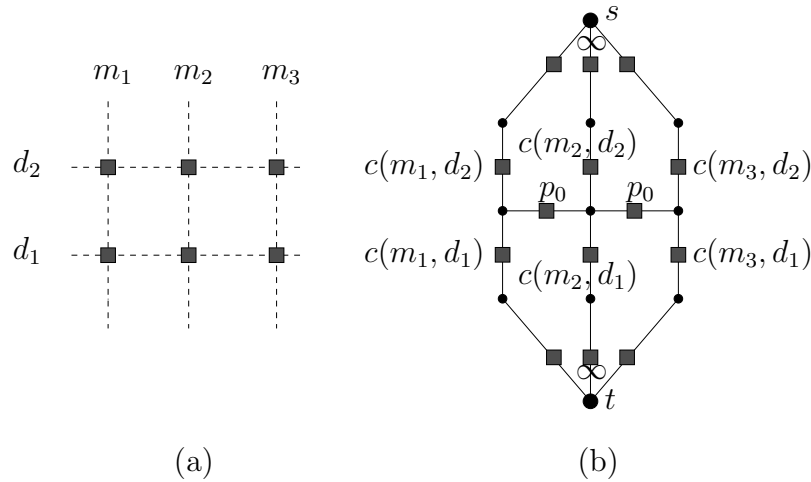


Figure 1.29 – Positions possibles pour f et la structure du graphe (en pointillés, les positions sont associées à des arcs verticaux) (a), le graphe avec les capacités des arcs (b).

d'inclure la contrainte de continuité. Roy et Cox cherchent dans ce graphe une coupe minimum séparant la source du puits. La carte de profondeur est construite à partir de la coupe minimum en associant pour chaque point (i, j) la plus grande disparité de l'arc associé le long de la coupe minimum choisie.

Dans [IG98], Ishikawa et Geiger ont présenté une autre formulation garantissant les contraintes d'ordre et d'unicité. Pour expliciter le modèle des discontinuités et des occultations, le modèle du graphe inclut des contraintes géométriques qui requièrent des correspondances mutuelles entre les discontinuités de disparité le long de la ligne épipolaire dans une image avec une région occultée dans l'autre image. La construction du graphe est quelque peu modifiée par rapport au modèle de Roy et Cox. Leur graphe est composé d'un double maillage 3D de tous les appariements possibles (x, y, r) où (x, r) sont les coordonnées d'un pixel dans l'image 1 et (y, r) sont les coordonnées dans l'image 2, d'une source et d'un puits. Les arcs sont orientés. Le modèle choisi encourage la régularité de surface le long des lignes épipolaires. La mise en correspondance stéréoscopique optimale correspond à la mise en correspondance des arêtes en utilisant la coupe minimale.

Boykov et al [BVZ99], [BVZ98], dans leur première approche, ont basé la fonction d'énergie sur un champ aléatoire de Markov (M.R.F.) qui est capable de préserver les discontinuités. Le coût associé à chaque arc correspond aux termes de la fonction d'énergie. Dans le cas général de la théorie des graphes, le but est de trouver une coupe minimale. Due à la grande complexité de cette recherche, Boykov et al ont proposé un algorithme qui cherche seulement une approximation de cette solution. Kolmogorov [KZ01] a poursuivi les travaux de Boykov en améliorant la fonction d'énergie dans laquelle les occultations étaient explicitement représentées.

Buehler et al. [BGCM02] ont repris la même approche basée sur la théorie des

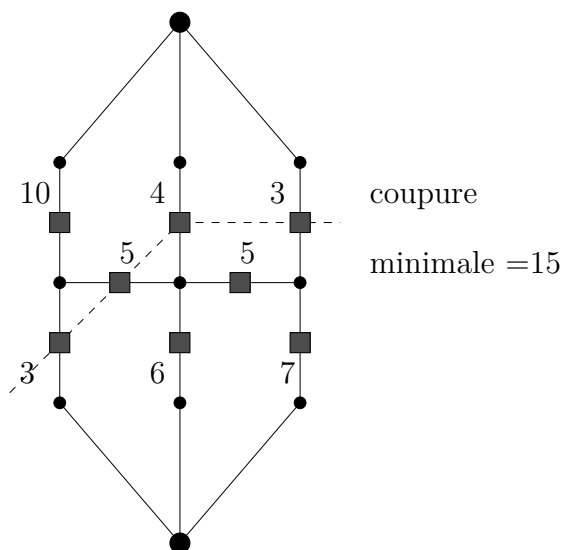


Figure 1.30 – Exemple de coupure minimale. L'appariement qui minimise la fonction d'énergie est $[(m_1, d_1), (m_2, d_2), (m_3, d_2)]$.

graphes pour traiter d'un cas de mise en correspondance avec un système trinoculaire.

Les méthodes basées sur la théorie des graphes ont deux limitations. La première est que, dû à la contrainte de continuité, elles ont tendance à construire des cartes de disparité qui ont toujours le défaut d'aplatir les objets. La deuxième est que, comme la fonction de pénalité entre deux pixels voisins de disparités différentes n'est pas nécessairement convexe, la minimisation de l'énergie est un problème NP-complet [KZ01] dont on obtient finalement qu'une approximation. Ishikawa [IG98] propose l'étude du cas où cette fonction est convexe et décrit un graphe qui permet d'obtenir le résultat exact.

1.3.3.3 Autres méthodes utilisant les contraintes globales

La programmation dynamique et la théorie des graphes utilisent le plus souvent des méthodes de minimisation d'énergie pour des mises en correspondance globales. Un certain nombre d'autres approches ont été développées. Parmi les plus notables, il y a la diffusion non linéaire et la propagation de croyance.

La méthode de Kanade et Okutomi [KO94] utilise une zone rectangulaire de taille variable qui, en croissant, exclue les directions qui augmentent une mesure d'incertitude. Scharstein et Szeliski [SS98] ont généralisé les travaux de Kanade et Okutomi en proposant d'utiliser une diffusion non uniforme (c'est à dire une agrégation locale pondérée de support plutôt que d'utiliser des fenêtres rectangulaires) pour augmenter la région support. Mansouri et al. [MM98] ont proposé une méthode similaire qui permet aux équations de diffusions d'être anisotropes dans les régions proches des bords des objets afin de mieux localiser les discontinuités de profondeur.

Dans [SSZ02], Sun et al modélisent le problème de la mise en correspondance glo-

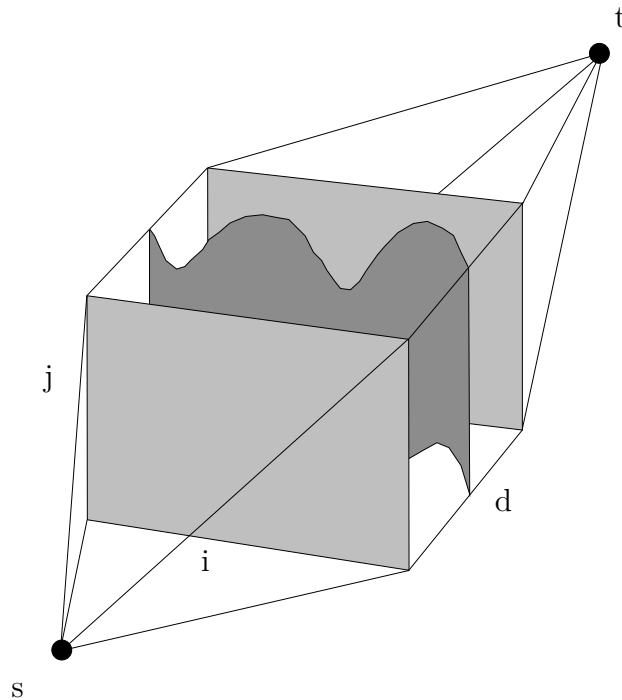


Figure 1.31 – Représentation par flot maximum de l'estimation de la disparité tel que proposée par Roy et Cox.

bale par une structure basée sur un réseau de Markov et le résolvent en utilisant la propagation de croyance. Cette approche obtient des résultats comparables aux méthodes basées sur la théorie des graphes. Dans [LQ02], Lhuillier et Quan proposent d'utiliser des points remarquables appariés grâce à l'intercorrélation normalisée centrée (Zero mean Normalized Cross-Correlation ou Z.N.C.C.) comme points de départ à une mise en correspondance quasi-dense en propageant la confiance d'appariement à leur voisinage. Ils ont inclus une contrainte de limite de gradient de disparité 2D permettant d'utiliser des images non calibrées.

Dans [YT96], Yvinec et Thiesse utilisent une représentation en fréquence dans le cas d'une géométrie épipolaire inconnue et relient le décalage spatiale entre les caméras par un déphasage dans l'espace fréquentielle de l'image.

1.3.4 Les méthodes locales

Les méthodes locales ont été supplantées par les méthodes globales au début des années 1990. Moins de dix ans plus tard, les recherches se sont de nouveau concentrées sur le concept de Marr en raison principalement des problèmes liés aux méthodes globales que sont la complexité de calcul élevée ainsi que la définition d'un modèle précis de la scène observée et l'intégration des contraintes stéréoscopiques. Il est apparu qu'il n'est pas nécessaire d'effectuer une optimisation globale pour formuler correctement le problème de la mise en correspondance et ainsi obtenir l'information de disparité mais qu'au contraire les informations locales de l'image étaient suffisantes. Dans les

méthodes locales, la mise en correspondance est basée sur l'appariement d'éléments les plus semblables. A chaque élément à mettre en correspondance est assigné une liste de mesures caractéristiques. Une distance statistique basée sur ces caractéristiques évalue leur similarité en vue de sélectionner des éléments appariés. Le coût algorithmique de la caractérisation d'un élément dépend uniquement des propriétés locales de l'image (avec un voisinage prédéfini). Le principe des méthodes locales repose alors sur la définition d'une fenêtre appropriée contenant l'élément à mettre en correspondance ainsi que son voisinage, sur le choix de caractéristiques discriminantes et une mesure de similarité calculée entre ces éléments. Chaque position de la fenêtre génère une signature particulière et on retient au final la position de la fenêtre obtenant le meilleur score d'appariement [BI99], [GLY95], [OKO01].

1.3.4.1 Choix de la zone

Le choix de la taille et de la forme de la zone est crucial pour garantir le succès des méthodes locales. La zone doit être assez large pour contenir suffisamment d'éléments afin de rendre l'appariement fiable, la fiabilité étant liée au nombre de pixels utilisés pour évaluer l'indice d'appariement. A contrario, la zone doit être suffisamment petite pour éviter les problèmes liés aux occultations et aux fortes variations de disparité. Enfin, la zone doit fournir des caractéristiques suffisamment discriminantes pour garantir un appariement unique.

Dans une démarche consistant à mettre en correspondance des pixels, l'approche classique consiste à définir un voisinage rectangulaire de taille fixe. Ce choix de taille fixe présente de nombreux inconvénients particulièrement si il existe de grandes variations de profondeur dans la scène analysée : une taille de fenêtre réalisant un bon compromis entre fiabilité et risque sur une zone de l'image correspondant à des objets proches ne conviendra pas dans une zone où la scène perçue est très éloignée. La différence de point de vue entre les caméras peut engendrer des motifs différents même pour une zone correspondante.

Pour pallier ces inconvénients, il a été envisagé d'adapter la taille de la fenêtre aux caractéristiques détectées dans la fenêtre.

Dans [KO94], Kanade et Okutomi augmentent la taille des fenêtres dans les zones de l'image où la surface est peu texturée afin d'augmenter le pouvoir discriminant de la zone de recherche.

Certains auteurs [TSK01], [ZK02] ont proposé de segmenter l'image afin d'être sûr que la fenêtre de recherche ne contienne pas d'élément extérieur à l'objet étudié et donc éviter ainsi les problèmes de variation de disparité liés aux bords d'objets.

Dans [SS02], Szeliski et Scharstein modélisent les fenêtres de recherche dans l'espace des disparités afin de rendre leur algorithme insensible aux problèmes géométriques liés aux positions des différentes caméras.

1.3.4.2 Mise en correspondance de motif

Les méthodes de mise en correspondance de motifs cherchent à estimer la disparité en un pixel de l'image en mettant en correspondance une petite région centrée sur le pixel (le motif de référence) dans la première image avec une série d'autres petites images de même taille extraites d'une zone de recherche dans l'autre image (Figure 1.32). La contrainte épipolaire permet généralement de réduire la zone de recherche à une seule dimension. Un grand nombre de distances statistiques ont été proposées pour estimer la ressemblance entre deux sous-images. Certaines sont basées sur la corrélation, d'autres utilisent des statistiques de moment ou d'ordre. Les distances les plus couramment utilisées sont présentées dans le tableau 1.1. Dans [AG93], Aschwanden et Guggenbuhl ont comparé ces différentes mesures.

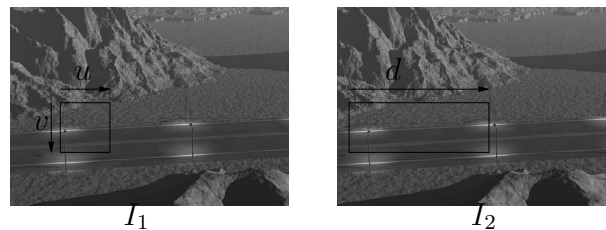


Figure 1.32 – Mise en correspondance de motif.

Le principe des distances basées sur les moments est le suivant : il s'agit de faire la somme des différences (quadratique ou absolue) pixel à pixel des deux motifs (S.A.D. ou S.S.D.). L'inconvénient de ces deux distances est qu'elles sont sensibles aux variations de l'illumination globale entre les deux images. C'est pourquoi on leur préfère généralement leurs versions centrées (Z.S.A.D. ou Z.S.S.D.) qui sont des distances invariantes aux translations uniformes de l'illumination globale. La mesure de distance statistique normalisée (N.Z.S.S.D.) permet de régulariser toutes les différences pixel à pixel au sein d'un même motif. L'intercorrélation normalisée (N.C.C.) est basée sur le même principe que les méthodes basées sur les moments mais effectue un calcul de corrélation entre les deux motifs. Cette méthode statistique est celle qui est la plus couramment employée pour déterminer une ressemblance.

Zabih et Woodfill [ZW94] ont proposé une méthode alternative utilisant les statistiques de rang pour assurer la mise en correspondance. Les statistiques de rang font partie du groupe des statistiques robustes. Les auteurs attendent donc de ce type de méthode une réduction de la sensibilité de la mise en correspondance vis à vis du modèle de bruit utilisé. Le calcul de la distance statistique qu'ils proposent se déroule en deux temps : tout d'abord, ils appliquent des transformations non paramétriques locales aux images. Cette transformation de rang est appliquée aux régions (le motif de référence et la zone de recherche) dans les deux images. La transformation de rang pour une petite région autour d'un pixel est définie comme le nombre de pixel composant la région pour lesquels l'intensité est plus petite que le pixel centrale (tableau 1.2). Les valeurs résultantes sont basées sur le nombre de pixels ayant un niveau de gris

Mesure de distance	Définition
Somme des écarts absolus (SAD)	$\sum_{u,v} I_1(u, v) - I_2(u + d, v) $
Somme des écarts absolus centrés (ZSAD)	$\sum_{u,v} (I_1(u, v) - \bar{I}_1) - (I_2(u + d, v) - \bar{I}_2) $
Somme des écarts quadratiques (SSD)	$\sum_{u,v} (I_1(u, v) - I_2(u + d, v))^2$
Somme des écarts quadratiques centrés (ZSSD)	$\sum_{u,v} ((I_1(u, v) - \bar{I}_1) - (I_2(u + d, v) - \bar{I}_2))^2$
Somme des écarts quadratiques centrés normalisés (NZSSD)	$\sum_{u,v} \left(\frac{I_1(u, v) - \bar{I}_1}{\sqrt{\sum_{u,v} (I_1(u, v) - \bar{I}_1)^2}} - \frac{I_2(u + d, v) - \bar{I}_2}{\sqrt{\sum_{u,v} (I_2(u + d, v) - \bar{I}_2)^2}} \right)^2$
Intercorrelation normalisée (NCC)	$\frac{\sum_{u,v} (I_1(u, v) \cdot I_2(u + d, v))}{\sqrt{\sum_{u,v} (I_1(u, v))^2 \cdot (I_2(u + d, v))^2}}$
Intercorrelation normalisée centrée (ZNCC)	$\frac{\sum_{u,v} (I_1(u, v) - \bar{I}_1) \cdot (I_2(u + d, v) - \bar{I}_2)}{\sqrt{\sum_{u,v} (I_1(u, v) - \bar{I}_1)^2 \cdot (I_2(u + d, v) - \bar{I}_2)^2}}$
Rang	$\sum_{u,v} (I'_1(u, v) - I'_2(u + d, v))$ $I'_k(u, v) = \sum_{m,n} I_k(m, n) < I_k(u, v)$
Recensement	$\sum_{u,v} \text{HAMMING}(I'_1(u, v), I'_2(u + d, v))$ $I'_k(u, v) = \text{BITSTRING}_{m,n}(I_k(m, n) < I_k(u, v))$

Tableau 1.1 – Méthodes de mise en correspondance de motifs dans le cas des images redressées.

inférieur à un seuil plutôt que sur les niveaux de gris eux-même. Ils ne retiennent au final que le nombre de pixels dont l'intensité est inférieure sans retenir leur localisation. Ensuite, après que la transformation de rang ait été appliquée, ils calculent la mise en correspondance de motif sur ces nouvelles valeurs en utilisant les méthodes classiques tel que la somme des écarts absolus.

Les statistiques de rang, si elles sont plus robustes, sont moins précises que les statistiques de moment. Dans le cas de la méthode de Zabih et Woodfill, la transformation non-linéaire permettant de passer d'un motif en niveaux de gris à deux proportions fait perdre beaucoup d'information. Pour pallier ce problème, les auteurs ont proposé une autre méthode sous le nom de "recensement" [ZW94] permettant de conserver l'information d'ordre relatif des pixels. Le principe de cette méthode est le suivant : ils comparent toujours le niveau de gris des pixels entourant le pixel central mais ils codent le résultat dans une chaîne booléenne permettant ainsi que conserver la distribution spatiale des rangs (tableau 1.2).

89 63 72		89 63 72	
67 55 64	$\Rightarrow 2$	67 55 64	$\Rightarrow 00000011$
58 51 49		58 51 49	

Tableau 1.2 – Exemple de transformation de rang (à gauche) et de recensement (à droite).

La mise en correspondance est alors calculée en utilisant la distance de Hamming entre les chaînes booléennes. Cette transformation augmente la dimension des données image par un facteur lié à la taille de la région locale, ce qui augmente le temps de calcul d'autant. Dans leur article de référence, Banks et Corke [BC01] comparent les performances des méthodes basées sur les transformations de rang et de recensement par rapport aux méthodes basées sur les mesures de corrélation et de différences. Leurs conclusions indiquent que les méthodes basées sur des statistiques de rang offrent des performances comparables et sont plus robuste aux variations de luminosité ainsi qu'aux occultations.

Les méthodes de mise en correspondance de motifs présentent l'avantage d'une grande simplicité d'implémentation logicielle ou matérielle [FHZa93], [LZ99]. De plus, elles ont fait l'objet de nombreuses études [BN98], [ZW94] et leurs performances sont bien connues [AG93].

Le principal inconvénient est que ces méthodes sont employées dans le cas de mise en correspondance d'images stéréoscopiques redressées. En effet, la discrétisation de l'espace des mouvements entre les deux images est implicite lorsqu'il s'agit de translation entre les deux images. En revanche, elle ne l'est pas lorsque la différence de point de vue entre les images comporte en plus des rotations, et donc des déplacements non-entiers des pixels. Comment alors faire la différence pixel à pixel entre deux motifs ? L'autre inconvénient est que si le motif de référence disparaît complètement ou

partiellement dans la zone de recherche suite à une occultation, le processus de mise en correspondance fournit quand même un résultat d'appariement. La mise en correspondance retenue sera celle la plus probable au sens de la distance utilisée mais ne représentera pas obligatoirement l'appariement correct. Il est donc nécessaire de fixer arbitrairement un seuil au-delà duquel la mise en correspondance est rejetée car trop peu fiable.

1.3.4.3 Méthodes basées sur le gradient, approches différentielles

L'approche différentielle réalise une estimation de mouvement en mettant en relation les dérivées spatiales et temporelles des images. Cette méthode qui a été initiée par Horn et Shunck [HS81] repose sur une propriété de conservation de l'illumination des points entre les différentes images. Dans le cadre de la mise en correspondance stéréoscopique, les méthodes basées sur le gradient ou flot optique [LK81], [GKR92] cherchent à déterminer de petites disparités locales entre deux images en formulant une équation différentielle reliant le mouvement entre les deux images et l'intensité lumineuse. Soit $v_m = (di_m, dj_m)^t$ la disparité à estimer au pixel m , l'écriture de la conservation de l'illumination locale donne :

$$(\nabla_i E) v_m + E_t = 0$$

avec $\nabla_i E$ la composante horizontale du gradient local de l'illumination autour du pixel et E_t la variation d'illumination locale entre les deux images. Cette équation est une équation projective qui ne contraint que la composante de v_m qui est parallèle au gradient local d'illumination. Pour obtenir une estimation de v_m , il faut régulariser le problème de l'estimation de la disparité, ce qui implique de devoir utiliser une contrainte additionnelle pour obtenir des résultats complets. Généralement, on suppose une certaine uniformité du champ de vitesse. En effet, sous réserve de ne pas se trouver sur des zones de discontinuité de mouvement, les pixels voisins du pixel considéré possèdent le même déplacement. Cette contrainte nous permet de nous rapprocher de la mise en correspondance de motif. La ligne épipolaire modélise alors le déplacement dans une seule direction.

Bien que les approches différentielles aient peu intéressé les recherches en mise en correspondance stéréoscopique (en comparaison du nombre d'articles portant sur les mise en correspondance de motifs), nous pouvons quand même citer les travaux de Scharstein et de Wei et al. Dans [Sch94b] et [Sch94a], Scharstein propose d'utiliser les champs de gradient de chaque image afin d'effectuer une mise en correspondance pixel à pixel en accumulant localement des votes sur différentes disparités possibles.

Dans [WBH98], Wei et al. calculent une fonction de coût de mise en correspondance de deux fenêtres basée d'une part sur une mesure (S.S.D.) basée sur les intensités des pixels et, d'autre part, sur des mesures basées cette fois ci sur les gradients des intensités des pixels.

L'un des principaux inconvénients de l'approche différentielle est qu'elle nécessite

une image suffisamment texturée sinon les gradients spatiaux et temporels tendent vers zéro. Cependant, l'estimation des gradients s'effectue au travers de filtres lisseurs (passe-bas) qui requièrent une faible texture de l'image pour être valide. Si la texture est trop importante, on aura de fortes variations d'intensité et la valeur estimée de la dérivée sera fautive du fait du lissage. Notons aussi que l'hypothèse d'illumination globale constante entre les deux images est fréquemment violée (cf. problèmes de la mise en correspondance). De plus, cette approche utilise l'hypothèse que l'ensemble des points utilisés pour lisser le champ de vitesse appartiennent à un même objet. Cette hypothèse est fautive dans le cas de discontinuités du mouvement, c'est à dire à la frontière entre les différents objets. Enfin, cette méthode estime de faibles mouvements et donc de faibles disparités. Cet algorithme est donc mis en défaut pour des scènes comportant de fortes disparités.

1.3.4.4 Mise en correspondance d'indices

Les méthodes de mise en correspondance de motif et les méthodes basées sur le gradient sont sensibles aux discontinuités de profondeur. Cette sensibilité vient du fait que deux zones centrées sur deux pixels correspondants peuvent contenir des pixels différents. Par exemple, si une zone est centrée sur le coin d'un objet, certains pixels appartiennent à l'objet et d'autres appartiennent à d'autres objets situés à une profondeur différente. Ces autres objets seront vus de manière différente par les deux caméras à cause de la différence de point de vue. Ces méthodes sont aussi très peu précises sur les zones peu texturées.

Les méthodes basées sur la mise en correspondance d'indices cherchent à palier ces problèmes en limitant les zones à mettre en correspondance à des éléments remarquables dans les images (coins [BHF96], lignes, courbes [SZ98], etc.). Dhond et Aggarwal [DA89] ont recensé un grand nombre de méthodes basées sur la mise en correspondance d'indices. Dans [VL01], Vincent et Laganier ont comparé les performances de différents algorithmes de mise en correspondance d'indices.

Nous allons maintenant présenter un algorithme de mise en correspondance d'indice utilisant une approche hiérarchique. Le principe de cette approche a été proposé par Hannah [Han84], [Han89] dans le cadre de la reconstruction de terrain à partir de deux vues aériennes. A partir d'une paire d'images stéréoscopiques, il génère un ensemble de paires d'images de résolution décroissante. Une partie de cet ensemble est montrée sur la figure 1.33. La diminution de la taille des images est obtenue par décimation en moyennant les niveaux de gris des pixels dans un carré de 2 pixels sur 2 pixels. Ces quatre pixels sont remplacés par un seul pixel dans l'image de résolution inférieure, et ainsi de suite.

Après la création de ces images, il cherche des points d'intérêts (coins, lignes, points lumineux, etc.) dans l'image de gauche de plus haute résolution. Il cherche ensuite les coordonnées d'un de ces points jusqu'à l'image de plus basse résolution (Fig 1.33). Ensuite, il recherche un correspondant à ce point en utilisant son vecteur d'attribut ainsi que la contrainte épipolaire dans l'image de droite au niveau de résolution le plus

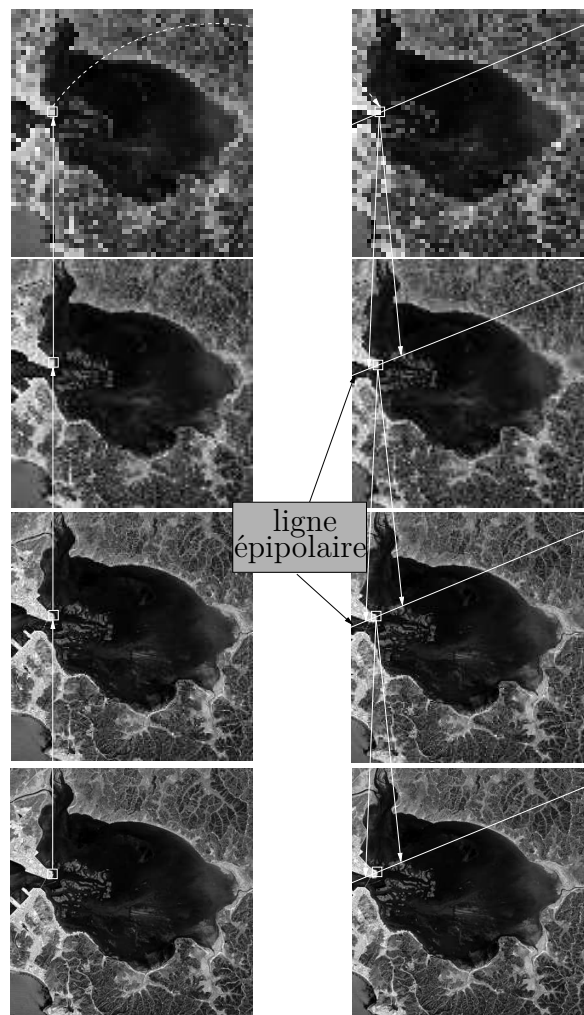


Figure 1.33 – Pyramides formées à partir d’une paire d’images à mettre en correspondance.

faible. L’image étant de taille réduite, la recherche n’est pas très coûteuse en temps de calcul. Ensuite, il recherche les coordonnées du correspondant depuis la plus faible résolution à la plus haute résolution. Le point sélectionné dans l’image de résolution inférieure permet d’initialiser une zone de recherche pour trouver le correspondant du point recherché à cette résolution.

L’approche hiérarchique dans le domaine de la mise en correspondance a souvent été reprise que ce soit dans le domaine de l’imagerie médicale [Oli00], [She04] ou dans le domaine de la robotique [KKCP96]. Sun [Sun98] propose d’utiliser une mise en correspondance hiérarchique ainsi qu’une mesure de mise en correspondance par corrélation mais en adaptant la fenêtre de recherche du correspondant en fonction du contenu de la zone. La mise en correspondance hiérarchique reste efficace même si la contrainte épipolaire est mal connue car la recherche du correspondant se fait sur la totalité d’une image de très faible taille.

Cependant, le nombre d'éléments remarquables dans les images est généralement très faible ce qui limite d'autant le nombre de pixels pour lequel la disparité va être estimée. Certains travaux plus récents basés sur les indices caractéristiques permettent quand même d'obtenir l'information de disparité sur la totalité de l'image. On distingue deux types d'approches différentes : la mise en correspondances d'indices hiérarchique et la mise en correspondance par segmentation.

La mise en correspondance d'indices hiérarchique proposée par Venkateswar et Chellappa [VC95] utilise quatre types d'éléments : lignes, sommets, bords et surfaces. Elle part du plus haut niveau d'indice (c'est à dire les surfaces) qui sont les plus faciles à apparier jusqu'au plus bas niveau (c'est à dire les lignes). Chaque appariement d'un niveau est utilisé comme région de recherche pour les éléments de niveau inférieur. Les contraintes classiques de restrictions d'appariement sont utilisées pour grouper les mises en correspondance en éléments certains. Cette méthode permet de réduire la complexité de calcul et les auteurs proposent d'utiliser cet algorithme comme estimateur de mouvement.

Birchfield et Tomasi [BT99] ont proposé une approche basée sur une pré-segmentation [RG91] des deux images avant la mise en correspondance. La segmentation permet d'obtenir des surfaces dont les pixels les composant appartiennent au même plan objet. Ces différents plans sont mis en correspondance en utilisant la méthode de Shi et Tomasi [ST94] qui détermine les paramètres de la transformation affine de chaque région entre les deux images.

Les méthodes de mise en correspondance d'indices classiques favorisent des cartes de disparités éparses. Les méthodes améliorées que nous venons de présenter permettent d'obtenir une information de disparité sur la totalité de l'image. La qualité des résultats est fortement tributaire de la qualité des algorithmes de détection de points d'intérêts (dans le cas de la méthode de Venkateswar et Chellappa) ou de la qualité de la segmentation (dans le cas de la méthode de Birchfield et Tomasi). Enfin, les cartes de disparités obtenues ne sont qu'une approximation simpliste de la fonction de disparité réelle en raison principalement du nombre limité de régions segmentées et du faible nombre de surfaces identifiées. Néanmoins, un certain nombre de méthodes [TM96], [Sch96], [LQ02] se base sur le principe des méthodes de mise en correspondance d'indices comme étape initiale à la mise en correspondance dense .

1.3.5 Les méthodes coopératives

Les algorithmes coopératifs sont inspirés du modèle de vision stéréoscopique humain et utilisent à la fois les méthodes locales et les méthodes globales.

Zitnick et Kanade [ZK00] ont proposé un algorithme itératif qui suit l'approche de Marr [MP77]. Ils définissent un espace de disparité comme une structure tridimensionnelle (x, y, d) représentant pour chaque pixel (x, y) dans une image de référence (par exemple l'image de gauche) toutes les disparités plausibles d (représentation similaire aux méthodes basées sur la théorie des graphes). Dans cette structure, deux zones sont définies : une zone support représentant la contrainte de continuité et une zone

interdite représentant la contrainte d'unicité. L'algorithme est initialisé en calculant, pour tous les appariements possibles, un score de mise en correspondance initial grâce à des méthodes basées sur l'intensité des images (S.S.D., S.A.D.). A chaque itération et pour certains appariement présents dans la zone support, le score de mise en correspondance est amélioré pendant que d'autres candidats à l'appariement, présents dans la zone interdite, sont supprimés. Ce processus est répété jusqu'à ce que la contrainte d'unicité ne soit plus violée. La carte des disparités résultante est construite de la façon suivante : pour chaque couple mis en correspondance, le score le plus haut est sélectionné. Si ce score est plus haut qu'un seuil, la disparité correspondante est acceptée. Il est à noter que cette approche heuristique ne produit pas de très bons résultats. Récemment, une autre approche coopérative a été publiée par Zhang et Kambhamettu [ZK02]. Leur algorithme adopte l'approche de Zitnick [ZK00] et l'étend dans deux directions. Tout d'abord, ils ont inclus une méthode qui garantit que les appariements initiaux de référence soient corrects. Ensuite, les régions supports sont définies sur une image segmentée. L'utilisation d'appariements supposés sûrs pour l'initialisation de l'algorithme augmente la robustesse du calcul et l'utilisation d'images segmentées assure que les régions supports ne contiennent pas d'élément extérieur à la surface des objets. Cependant, il n'existe aucune assurance que la segmentation des images soit correcte, ce qui peut conduire l'algorithme à générer des résultats non valables.

Pour pallier ces problèmes de segmentation, Tomasi et Manduchi [TM96] ont proposé de baser la mise en correspondance sur des caractéristiques locales de l'illumination invariantes aux translations géométriques et aux variations d'illumination (par exemple la dérivée de l'intensité lumineuse). On caractérise ainsi chaque point de chaque image avec une liste d'invariants et, à l'aide de la contrainte épipolaire, on cherche à mettre en correspondance les points qui offrent une liste d'invariants similaire. Pour cela, on choisit les points les plus remarquables et donc les plus faciles à mettre en correspondance. Le problème est que, en général, le nombre de ces points est restreint. De plus, dans le cas des images en niveau de gris, il faut considérer des caractéristiques invariantes d'ordre trois [Sch96]. Gouet et Montesinos [GMP98] ont proposé d'utiliser des images couleur afin d'utiliser des invariants d'ordre un. Une fois que l'on a mis en correspondance ces points grâce à leurs listes d'invariants, on crée une triangulation de Delaunay. Il s'agit de créer un triangle dont les sommets sont les points remarquables précédemment mis en correspondance. Ainsi on sait que les points qui se trouvent à l'intérieur d'un triangle formé par les trois points caractéristiques dans la première image doivent avoir leurs correspondants dans un triangle dual formé par les trois même points caractéristiques mis en correspondance dans la deuxième image. Dans chaque triangle, on cherche alors de nouveaux points caractéristiques. On les met en correspondance grâce à la contrainte épipolaire et on recrée de nouveaux triangles (Figure 1.34). Cet algorithme itératif permet de déterminer au final une mise en correspondance quasi-dense car l'ensemble des points de l'image de gauche a un correspondant dans l'image de droite du fait que chaque point appartient à un triangle.

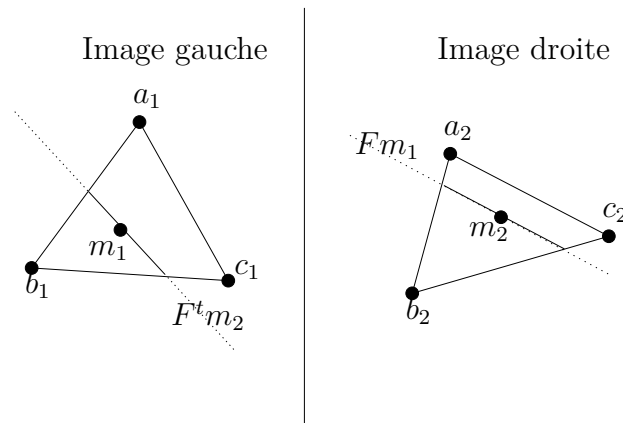


Figure 1.34 – Mise en correspondance avec invariants et triangulation de Delaunay.

1.4 Conclusion

Dans ce chapitre, nous avons proposé un état de l'art sur la stéréovision ainsi que sur les méthodes de mise en correspondance. Nous avons tout d'abord montré les phénomènes et les problèmes de modélisation d'un pixel liés aux méthodes de numérisation de l'image. Qu'il s'agisse de l'échantillonnage de l'espace ou de la quantification des niveaux de gris, les modèles d'erreur généralement utilisés sont des modèles au mieux incomplets ou tout simplement faux. Concernant l'aspect géométrique d'un capteur stéréoscopique, nous avons mis en évidence les contraintes qui sont généralement utilisées lors de la phase d'appariement.

Les méthodes de mise en correspondance se divisent en trois principales classes : les méthodes locales, les méthodes globales et les méthodes coopératives. Dans les méthodes locales, on utilise le voisinage d'un point dans une image pour trouver son correspondant dans l'autre image en utilisant une mesure de similarité du voisinage. Les méthodes globales, quant à elles, envisagent le problème soit le long d'une ligne de recherche soit sur l'image en entier.

Il ressort de notre étude plusieurs observations :

- Les méthodes utilisant les contraintes locales nécessitent la définition d'une fenêtre de taille et de forme appropriée pour définir une distance statistique entre deux motifs dans le cas des mises en correspondance de motif et des approches différentielles. La mise en correspondance de motifs tente d'estimer la disparité en un point de l'image en comparant des motifs de pixels centrés autour d'un point. L'inconvénient de ces approches est que la discrétisation de l'espace des mouvements les restreint aux images stéréoscopiques redressées. De plus, la présence d'occultations conduit l'algorithme d'appariement à fournir des résultats erronés.

Les méthodes basées sur le calcul de gradient utilisent une approche différentielle reliant le mouvement entre les deux caméras et la variation de l'intensité lumineuse. Ces méthodes peuvent s'appliquer sur tout type d'image stéréoscopique

mais l'emploi de filtres deriveurs induit deux hypothèses contradictoires quand 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. L'emploi de ces filtres pour calculer les gradients de l'image impose que la disparité entre les deux images soit faible. De plus, la présence de pixels appartenant à différents objets qui sont à différentes profondeurs fausse l'estimation du champ de vitesse.

Les méthodes basées sur la mise en correspondance d'indices limitent l'appariement à des points particuliers afin de diminuer la sensibilité des algorithmes aux fortes variations de disparité dans la scène. Elles présentent comme avantage de réduire le volume de données à traiter et ne nécessitent pas la détermination d'une fenêtre de travail. Cependant, le fait que seuls quelques points soient utilisés génère au final une carte de disparité éparse. Souvent, ces méthodes sont utilisées pour initialiser d'autres algorithmes. Au final, les méthodes locales sont relativement efficaces mais elles sont sensibles localement aux régions ambiguës (occultations, zones peu texturées., etc.).

- La principale approche utilisant les contraintes globales est basée sur la programmation dynamique. Cette méthode cherche, le long d'une ligne de recherche, à mettre en correspondance l'ensemble des pixels de façon à optimiser une fonction de coût global. Les variations de cette approche sont nombreuses. Cependant, toutes ces méthodes sont très coûteuses en temps de calcul et une erreur locale a de très fortes répercussions sur la solution globale.

Une autre approche modélisant de manière globale le problème de la mise en correspondance utilise la théorie des graphes. Cette approche permet d'inclure une cohérence bidimensionnelle par rapport à la scène observée. Le problème est que ce type d'approche construit des cartes de disparité aplaties et la complexité de calcul incite presque toujours à déterminer une estimation de la solution désirée. Un certain nombre d'autres méthodes utilisant les contraintes globales ont été développées (propagation de croyance, diffusion non-uniforme).

L'avantages des méthodes globales est qu'elles sont moins sensibles aux problèmes des régions ambiguës que les méthodes locales. En effet, les contraintes globales apportent des informations supplémentaires afin de mettre en correspondance localement les régions semblables. Cependant, les méthodes globales sont plus coûteuses en temps de calcul et l'élaboration d'un modèle global n'est pas toujours possible.

- Les méthodes coopératives utilisent aussi bien les contraintes locales que les contraintes globales, parfois de façon itérative. L'objectif des ces méthodes est de reprendre les avantages des deux approches.

De cet état de l'art nous pouvons déduire que chaque approche possède ses propres avantages et inconvénients. Chacune a fait l'objet de modifications et d'améliorations en vue de palier les lacunes de chacune. Néanmoins, il n'existe pas à l'heure actuelle une méthode unique qui marche pour tout type de scène (vue aérienne, environnement

structuré, scène à forte disparité, etc.) et pour une paire d'images stéréoscopiques quelconque (non redressée). On peut tout de même envisager un processus se rapprochant de la vision humaine, c'est à dire basé sur l'utilisation d'un réseau de neurones comme proposé par Garcia [Gar97] ou par Wang and Hsiao [WH99].

Plutôt que de tenter d'améliorer les performances des techniques déjà existantes, nous avons préféré nous tourner vers une nouvelle approche. Comby [Com01] a élaboré au cours de sa thèse une méthode pour apparier des pixels utilisant leur niveau de gris modélisé grâce à la théorie du flou afin de rendre son algorithme d'estimation de mouvement majoritaire plus robuste vis à vis des variations de niveau de gris entre les images. Nous allons nous en servir comme base à notre approche de mise en correspondance stéréoscopique dense. De plus, nous élaborons un nouveau modèle de pixel permettant d'inclure les problèmes liés à la numérisation de l'image.

Chapitre 2

Modèle de pixel flou pour la mise en correspondance pixel à pixel

2.1 Introduction

2.1.1 Notre objectif

Le but ultime de ce travail est d'obtenir une mise en correspondance dense de deux images stéréoscopiques, c'est à dire d'être capable de trouver, pour tout pixel de l'image 1, un domaine correspondant dans l'image 2 le plus spécifique possible et vice-versa. La situation optimale donnerait une bijection entre les deux images (à tout pixel de l'image 1 correspondrait un et un seul pixel de l'image 2) ce qui permettrait une reconstruction tridimensionnelle précise.

Cette bijection ne peut être obtenue dans la pratique. En effet, les images numériques sont échantillonnées et quantifiées, les informations de luminance sont sujettes à des bruits divers et les informations dont on dispose sont insuffisantes pour lever toute ambiguïté d'appariement, même si, disposant de l'étalonnage du capteur stéréoscopique, on peut réduire ces ambiguïtés par l'utilisation de la contrainte épipolaire.

C'est pourquoi nous proposons de relaxer un peu la contrainte de mise en correspondance pour se tourner vers une méthode imprécise mais fiable, c'est à dire que si un point de l'espace tridimensionnel se projette sur un pixel de l'image 1, sa projection sur l'image 2 appartient au domaine de l'image 2 défini par la mise en correspondance.

2.1.1.1 Sur quels critères peut-on se baser pour établir les mises en correspondance ?

De quelles informations dispose-t-on finalement pour assurer cette mise en correspondance et quelles sont les hypothèses minimales permettant cette association ?

Une image numérique naturelle est une représentation échantillonnée et quantifiée de la luminance projetée d'une scène dans un cône de visibilité défini par le système optique du capteur. Dans le chapitre précédent, nous avons présenté un certain nombre de technologies permettant d'obtenir de telles images. Nous avons vu que ces technologies ne délivraient pas la même information sur la luminance projetée. Les surfaces sur lesquelles la luminance est intégrée (ou photosite) diffèrent suivant les capteurs, tandis que la forme sous laquelle l'information est disponible, elle, est identique (pixel). La connaissance de la technologie utilisée semble nécessaire pour traiter l'information numérique de la luminance, particulièrement lorsqu'il s'agit de réaliser des mises en correspondance. Cependant, cette connaissance, si elle est disponible, est difficile à utiliser sans rajouter dans le traitement de l'information un haut niveau de complexité de représentation ainsi que beaucoup d'a priori sur les propriétés de diffusion de la lumière.

La plupart des méthodes de mise en correspondance utilisées en stéréovision s'appuient sur une hypothèse de diffusion uniforme de la luminance projetée. Cette hypothèse peut être formulée ainsi :

un élément de surface de la scène perçue retro-diffuse la lumière de façon similaire dans des directions d'observation proches.

Si on oublie les problèmes liés aux occultations et aux différences d'aspect, cela veut dire que la distribution de luminance projetée pour cet élément de surface est identique sur les photorécepteurs des deux caméras du système stéréoscopique. Le problème de base de la mise en correspondance stéréoscopique vient du fait que l'on ne connaît qu'un échantillonnage quantifié de cette distribution de luminance projetée. A ce problème s'ajoute le fait que le processus d'obtention de l'information échantillonnée varie, ce qui veut dire que, en fonction du capteur utilisé, les informations de luminance sont différentes.

Enfin, à ce défaut, s'ajoutent tous les problèmes liés aux bruits des systèmes, aux différences de point de vue, aux occultations, à la non-vérification de l'hypothèse de retro-diffusion uniforme, aux différences d'échantillonnage sur les deux images, etc... et surtout à la quantification qui induit imprécision et saturation de l'information de luminance. La modélisation de la luminance par une variable aléatoire ne permet pas de représenter de façon simple ces défauts de connaissance du signal.

Nous introduisons dans ce chapitre une nouvelle façon de modéliser les images numériques. Nous attendons de cette modélisation une représentation des informations de luminance suffisamment générale pour que les traitements effectués dans ce cadre puissent s'abstraire de la technologie d'obtention de l'image. Nous souhaitons que les méthodes de traitement qui en découlent soient simples d'utilisation et permettent l'intégration d'informations telles que les contraintes épipolaires et les défauts d'étalement des capteurs. Nous souhaitons que la modélisation des hypothèses de travail ajoute un apriorisme minimal sur la mise en correspondance. Enfin, nous souhaitons rester dans le cadre du traitement numérique, c'est à dire que nous ne souhaitons pas passer par une phase de reconstruction de l'information de la luminance continue. Nous

adressons pour ce faire la théorie des sous-ensembles flous.

Après ce premier paragraphe introductif, le deuxième paragraphe de ce chapitre présente la dégradation de l'hypothèse de conservation de la distribution de luminance par l'échantillonnage. Le troisième paragraphe dresse un panorama de différentes méthodes pour représenter cette dégradation. La fin de ce troisième paragraphe présente l'aspect géométrique de notre nouveau modèle de pixel afin de prendre en compte les problèmes liés à la phase d'échantillonnage de l'espace. Le quatrième paragraphe présente l'aspect chromatique lié au pixel. Le cinquième paragraphe introduit notre modélisation du problème de la mise en correspondance stéréoscopique. Finalement, une conclusion fera la synthèse de l'apport de ces nouvelles approches.

2.2 Pixel, aspect géométrique

2.2.1 Hypothèse sur la luminance

Nous abordons, dans ce paragraphe, la dégradation de l'hypothèse forte de conservation de la distribution de luminance par l'échantillonnage et son impact sur la mise en correspondance. Pour plus de simplicité dans les explications, nous posons le problème dans l'espace bidimensionnel.

Sur la figure 2.1(a) est représenté un point P de l'espace bidimensionnel ainsi que ses projections sur le plan rétinien de chaque caméra composant le capteur stéréoscopique. Soit p_1 et p_2 les projections perspectives de P sur les images 1 et 2. En supposant que les aléas de l'échantillonnage aient placé p_1 au centre du pixel m_1 , on comprend aisément qu'il y a peu de chance que la projection p_2 de P sur le plan de l'image 2 coïncide avec le centre d'un quelconque pixel de l'image 2 (figure 2.1(b)).

Appelons m_2 le pixel dont le centre est le plus proche de p_2 . En supposant que le capteur réalise un échantillonnage parfait, on voit que l'information d'illumination en m_1 et en m_2 sont intrinsèquement différentes : m_2 représente l'illumination d'un point au mieux proche de P si localement la surface est lisse (ou si P n'est pas un coin d'un objet) (figure 2.1(c)). Le même paradigme intervient si on considère que les photosites réalisent une intégration de la luminance. La luminance intégrée sur le photosite correspondant au pixel m_2 est par nature différente de la luminance intégrée sur le photosite correspondant au pixel m_1 (figure 2.1(d)). Dans ces deux cas, m_1 et m_2 sont les images échantillonnées de P alors que la valeur des mesures de luminance sont différentes par essence (indépendamment de tout bruit).

Ce phénomène est illustré sur la figure 2.2 par les deux vues d'une même scène obtenues par une même caméra après un faible déplacement. Dans ce cas, l'hypothèse d'uniformité de la réflexion lumineuse peut être considérée comme valide, cependant la répartition des niveaux de gris correspondant à un même détail (ici le coin d'une fenêtre) est très différente sur les deux images échantillonnées.

L'hypothèse forte utilisée dans la mise en correspondance pixel à pixel doit être reformulée de la façon suivante :

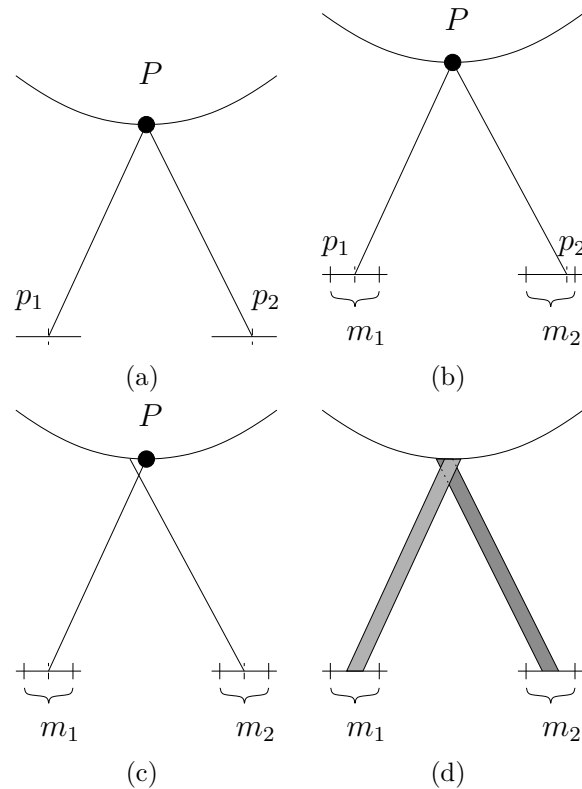


Figure 2.1 – Les problèmes de l'échantillonnage.

si p_1 et p_2 sont les projections de deux points proches de l'espace tridimensionnel, les valeurs de luminance échantillonnées qui leur sont associées sont proches ou encore, la distribution locale des luminances est comparable.

Reste à définir ce que l'on entend par proche ou comparable, et comment formuler cette hypothèse de façon à minimiser l'a-priori de représentation de la distribution de luminance et du phénomène d'échantillonnage. Cette représentation ne peut pas mettre de côté la nature intégrative de la mesure de luminance d'une part, ainsi que le fait que l'information n'est pas continue d'autre part. Pour mieux comprendre la nature même du problème, revenons à la mesure échantillonnée de luminance réalisée par les capteurs d'images.

2.2.2 Mesure de luminance échantillonnée : rapport entre pixel et photosite

Nous avons présenté dans le paragraphe 1.1.3 trois types de technologies de capteurs d'images. Les photosites associés à chaque technologie échantillonnent la luminance projetée de façon différentes. Ces différences sont résumées sur le figure 2.3.

Si on suppose que la mesure de luminance ne présente pas d'aléa, on voit que la surface d'intégration se rapproche fortement de l'échantillonnage ponctuel dans le cas des caméras à balayage, de l'échantillonnage par pavage dans le cas des capteurs CCD

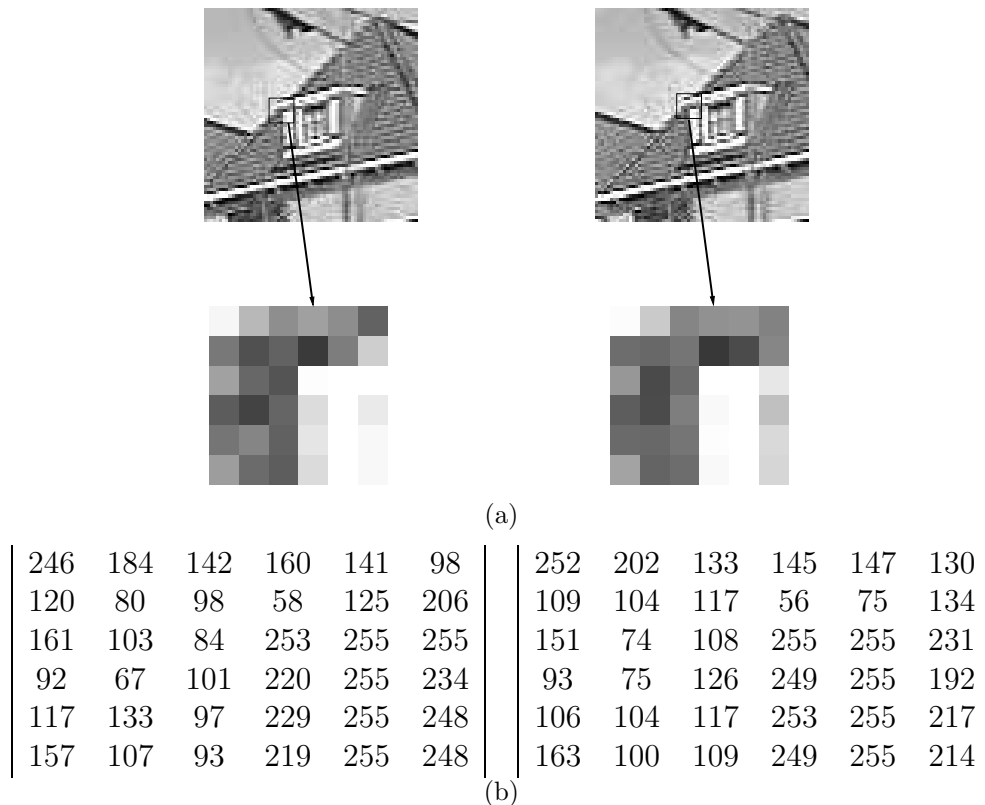


Figure 2.2 – Influence de l'échantillonnage entre deux vues proches (a), valeurs des niveaux de gris des deux motifs (b).

et d'une situation intermédiaire dans le cas des capteurs CMOS.

De façon générale, si $\phi(x, y)$ est une fonction continue de la luminance projetée sur la rétine, la valeur $I(i, j)$ de l'illumination échantillonnée peut être écrite sous la forme intégrale :

$$I(i, j) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(x, y) n_{i,j}(x, y) dx dy \quad (2.1)$$

où $n_{i,j}(x, y)$ représente le noyau d'intégration sur le pixel de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne. La forme de ce noyau dépend de la technologie utilisée. Le modèle classique de représentation de l'échantillonnage utilise des noyaux sommatifs [Sil86] pour associer, à chaque mesure de la luminance échantillonnée, un domaine probabilisé du plan image. Ce domaine probabilisé représente indifféremment : les phénomènes d'intégration donnant lieu à la mesure de la luminance, le domaine spatial sur lequel la mesure de luminance peut être considérée comme ergodique ou le domaine dans lequel cette valeur de luminance peut être prise en compte pour en obtenir une interpolation. Cependant, le choix du noyau influe considérablement sur les résultats obtenus [CP95].

Le modèle minimal d'un noyau ne représentant que le phénomène d'échantillonnage

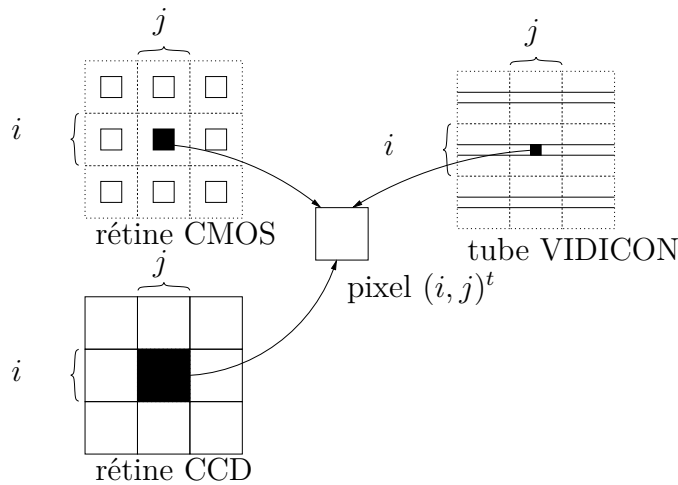


Figure 2.3 – Différents échantillonnages de l'espace.

est tel que :

- son support maximal est défini par les 8 voisins de chaque pixel,
- son support minimal est défini par le pixel "ponctuel".

La littérature propose un certain nombre de possibilités de modélisation de l'échantillonnage par noyaux sommatifs [Sch01], [KK01], [BT98b] ou par intervalles d'erreurs [Tel03]. Chacune des représentations par noyaux sommatifs s'attache plutôt à une technologie particulière sauf dans la représentation par noyau gaussien qui se veut plus générale en s'appuyant sur la théorie de la convergence vers la loi normale. Cette dernière représentation ne correspond cependant pas à la première propriété minimale des noyaux sommatifs proposée précédemment. Enfin, il semble important de prendre en compte une hypothèse de continuité de la luminance dans les zones où l'information est absente (c'est à dire entre les photosites). En effet, sous hypothèse de continuité de la matière et dans le cas de surface faiblement granuleuse, on admet qu'il y a une forte corrélation entre les points de l'espace qui se projettent à côté du photosite et ceux qui se projettent à l'intérieur du photosite.

2.3 Modèles classiques de représentation d'une information aléatoire spatialisée : le pixel

Le fait que l'information soit échantillonnée a, bien sur, une grande influence sur la nature même de la problématique de mise en correspondance basée sur la luminance. L'objet de ce paragraphe est de dresser un panorama rapide des méthodes qui sont utilisées habituellement pour représenter la dégradation de l'information imposée par l'échantillonnage.

2.3.1 Modélisation par noyau sommatif

Nous nous intéressons à la modélisation du phénomène de mesure de la luminance sur un support mal connu (figure 2.3). Dans le paragraphe précédent, nous avons vu que la surface de ce support pouvait varier depuis le point jusqu'au pavé représentant ce que nous avons convenu d'appeler pixel (pixel représenté sur la figure 2.4).

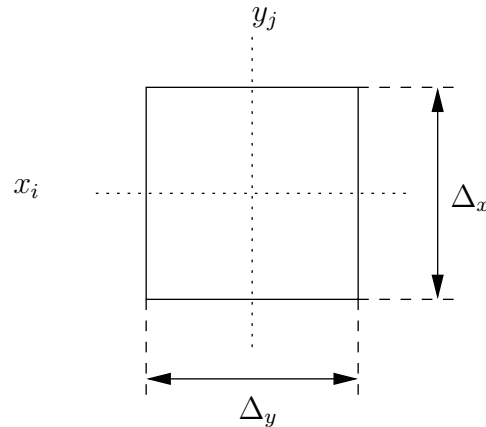


Figure 2.4 – Définition d'un pixel.

La modélisation classique par noyau sommatif consiste à considérer que l'image échantillonnée $I(i, j)$ est obtenue par la convolution de la luminance continue $\phi(x, y)$ avec un ensemble de noyaux sommatifs $n_{i,j}(x, y)$ (équation 2.1). La sommativité des noyaux s'exprime par :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} n_{i,j}(x, y) dx dy = 1$$

Les noyaux sommatifs, au sein d'une même image, sont généralement tous identiques et choisis dans une classe de noyaux centrés comme les noyaux gaussiens, uniformes, exponentiels, etc..., l'impulsion de Dirac étant un cas particulier de noyau sommatif.

On pose :

$$n_{i,j}(x, y) = n(x - x_i, y - y_j)$$

avec x_i et y_j les coordonnées du centre du pixel (figure 2.4). La relation 2.1 entre $I(i, j)$, la valeur de l'illumination échantillonnée, et $\phi(x, y)$, la fonction continue de la luminance projetée sur la rétine, devient :

$$I(i, j) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(x, y) n(x - x_i, y - y_j) dx dy \quad (2.2)$$

Il doit exister un lien fort entre les paramètres de la forme du noyau sommatif choisi et la période (ou surface) d'échantillonnage. Un des résultats remarquables de la théorie

du traitement du signal échantillonné est la possibilité qui est offerte de réaliser, par traitement discret de l'image, une opération d'extraction de l'information équivalente à ce que l'on aurait obtenu en réalisant cette opération sur l'image continue puis en l'échantillonnant par la suite.

Une des illustrations les plus connues de cette propriété est l'estimation de gradient de luminance [Der87]. On montre ainsi que, en supposant que l'image a été échantillonnée en utilisant un noyau sommatif $n_{i,j}(x, y)$ donné, l'information échantillonnée $I_x(i, j)$ de gradient en x peut être obtenue par la convolution numérique de l'image échantillonnée avec la dérivée échantillonnée du noyau (au signe près). Des travaux sur les problèmes d'estimation de gradient ressort que l'influence du noyau sur le résultat est prépondérante [CP95].

Contrairement aux travaux ayant amené à la définition des opérateurs optimaux de dérivation des images (Canny/Derriche [Der87], Shen/Castan [SC92]), il nous semble difficile de spécifier mathématiquement les propriétés d'un noyau idéal permettant d'optimiser une mise en correspondance. Par contre, il nous semble acquis que l'ensemble des noyaux représentatifs de cette intégration doivent être recherchés dans la classe des noyaux centrés symétriques à support borné. La limite maximale du support du noyau est donnée par le pas d'échantillonnage (Δ_x, Δ_y) comme défini sur la figure 2.4.

En traitement d'image, on utilise aussi les noyaux sommatifs pour effectuer des opérations d'estimation de la luminance en dehors des valeurs échantillonnées [KK01], [LBL01]. Les noyaux utilisés doivent alors permettre d'estimer la luminance en n'importe quel pavé de l'image. Ainsi, soit P un pavé défini par $[x_1, x_2] \times [y_1, y_2]$ de l'espace de l'image (figure 2.5a), la luminance reçue sur le pavé P peut être estimé par :

$$\begin{aligned} \hat{I}(P) &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} \left(\sum_{i,j} I(i, j) m(x - x_i, y - y_i) \right) dx dy \\ \hat{I}(P) &= \sum_{i,j} \left(I(i, j) \int_{x_1}^{x_2} \int_{y_1}^{y_2} (m(x - x_i, y - y_i)) dx dy \right) \end{aligned} \quad (2.3)$$

$m(x, y)$ étant un noyau sommatif pour assurer l'indépendance et la conservation de l'énergie reçue. $\hat{I}(P)$ est la somme des intensités discrètes pondérées par le recouvrement du pavé sur les noyaux.

Ce noyau interpolatif peut avoir une autre interprétation. Si on considère que $I(i, j)$ est une réalisation d'un processus aléatoire de mesure de la luminance, alors l'utilisation du noyau $m(x, y)$ doit permettre d'utiliser l'ergodicité supposée de la mesure pour atténuer les effets du bruit.

Si $P(i, j)$ est la vraie surface d'intégration de la luminance (figure 2.5b), alors :

$$\hat{I}(i, j) = \hat{I}(P(i, j)) = \sum_{i,j} \left(I(i, j) \int_{x_i - \frac{dx}{2}}^{x_i + \frac{dx}{2}} \int_{y_i - \frac{dy}{2}}^{y_i + \frac{dy}{2}} (m(x - x_i, y - y_i)) dx dy \right)$$

avec $d_x = (x_2 - x_1) < \Delta_x$ et $d_y = (y_2 - y_1) < \Delta_y$.

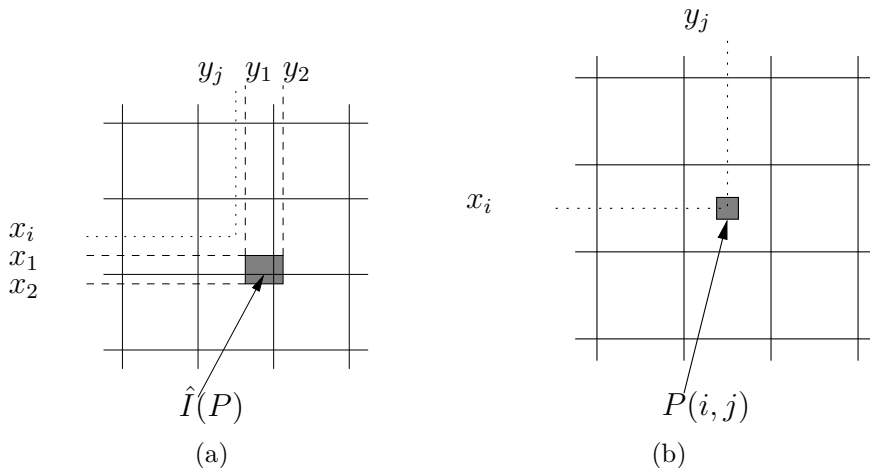


Figure 2.5 – Interpolation (a), ergodicité (b).

Si le processus est considéré comme non aléatoire, alors le noyau $m(x, y)$ doit être tel que :

$$\hat{I}(i, j) = I(i, j)$$

et donc :

$$\int_{x_i - \frac{d_x}{2}}^{x_i + \frac{d_x}{2}} \int_{y_i - \frac{d_y}{2}}^{y_i + \frac{d_y}{2}} (m(x - x_i, y - y_i)) dx dy = \int_{-\frac{d_x}{2}}^{\frac{d_x}{2}} \int_{-\frac{d_y}{2}}^{\frac{d_y}{2}} (m(x, y)) dx dy = 1 \quad (2.4)$$

$m(x, y)$ doit donc être aussi un noyau centré symétrique dont le support maximum est borné par (Δ_x, Δ_y) .

Si on combine ces deux approches, alors la densité lumineuse reconstruite $\hat{I}(i, j)$ en tout point de l'image est obtenue par la convolution de $\phi(x, y)$ avec un noyau sommatif centré $N(x, y)$ obtenu par convolution des deux noyaux $m(x, y)$ et $n(x, y)$. Les supports de ces deux noyaux étant bornés par (Δ_x, Δ_y) , le noyau sommatif obtenu a un support borné par $(2\Delta_x, 2\Delta_y)$.

On voit que l'approche sommative combine dans un même formalisme l'aspect intégratif de la mesure de luminance et l'aspect probabiliste de l'estimation de la luminosité en dehors des valeurs d'échantillonnage.

On voit aussi que, si les supports d'intégration étaient connus de façon certaine et si il était possible de formuler mathématiquement des contraintes sur le noyau interpolatif, il serait alors possible de définir un noyau $N(x, y)$ optimal pour traiter de la mise en correspondance. La méconnaissance des supports d'intégration de la luminance,

la difficulté d'imposer des contraintes d'optimalité ainsi que la volonté de trouver un modèle générique amène généralement à adopter soit un noyau gaussien, soit un noyau uniforme. Ce sont ces deux noyaux que nous proposons d'étudier maintenant.

2.3.1.1 Le noyau gaussien

Si on donne au noyau sommatif utilisé une connotation probabiliste, on peut alors s'appuyer sur le théorème de la limite centrale pour généraliser l'ensemble des noyaux utilisables pour modéliser aussi bien l'échantillonnage que la reconstruction par un noyau normal à deux dimensions [Bai90]. La même justification peut être donnée si on considère l'aspect intégratif de la mesure de luminance. La distribution de probabilité doit à la fois représenter le phénomène incertain qui a conduit à la mesure échantillonnée de l'information de luminance ainsi que l'estimation de cette information en dehors du maillage de l'image. Le noyau gaussien est le noyau le plus courant en vision par ordinateur [XZ96], [Zha98]. Ce modèle suppose que l'information est majoritairement contenue au centre du pixel et qu'elle suit une distribution normale centrée autour de ce point (Figure 2.6).

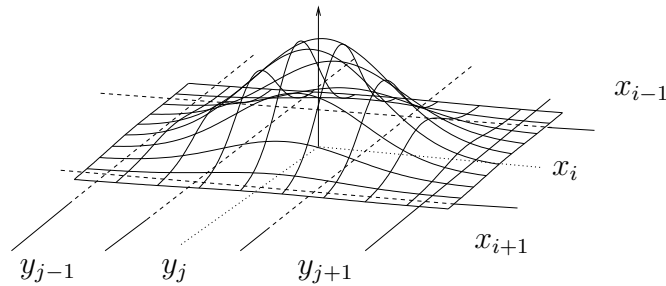


Figure 2.6 – Représentation de l'échantillonnage par une distribution de probabilité unimodale centrée.

Le noyau en (x_i, y_j) peut alors s'écrire :

$$N_{i,j}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} e^{\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-x_i)^2}{\sigma_1^2} - \frac{2\rho(x-x_i)(y-y_j)}{\sigma_1\sigma_2} + \frac{(y-y_j)^2}{\sigma_2^2} \right] \right\}}$$

avec σ_1 et σ_2 les écarts-type selon x et y et ρ un coefficient de corrélation entre x et y .

Le principal défaut que l'on peut voir à l'utilisation d'un noyau normal est de ne pas permettre de respecter la contrainte exprimée par la formule 2.4, la loi normale n'étant pas à support borné. En fait, la convergence vers la loi normale permet une justification d'un traitement de l'information en grand nombre. Lorsqu'il s'agit de traitements locaux, cet argument ne peut plus être invoqué. Le noyau gaussien devient un noyau sommatif comme un autre mais ne respectant pas la contrainte 2.4 et donc ne permettant pas de représenter uniquement la dégradation de l'information induite par l'échantillonnage.

2.3.1.2 Le noyau uniforme

Le choix d'une modélisation par loi uniforme permet de bien appréhender l'aspect intégratif de la mesure de luminance. C'est une loi bornée et qui semble minimiser l'a-priori de représentation puisque tous les points du domaine associé à ce noyau ont le même poids. Le choix du noyau de support (Δ_x, Δ_y) pour modéliser l'échantillonnage semble être celui qui minimise les risques d'erreurs consistant à rejeter l'hypothèse d'une information en dehors de ce support, connu aussi sous le nom d'erreurs de première espèce.

Le noyau en (x_i, y_j) modélisé par une loi uniforme comme représenté sur la figure 2.7 peut alors s'écrire :

$$N_{i,j}(x, y) = \begin{cases} \frac{1}{\Delta} & \text{si } x \in [x_i - \frac{\Delta_x}{2}; x_i + \frac{\Delta_x}{2}] \text{ et } y \in [y_j - \frac{\Delta_y}{2}; y_j + \frac{\Delta_y}{2}] \\ 0 & \text{sinon} \end{cases}$$

avec $\Delta = \Delta_x \Delta_y$.

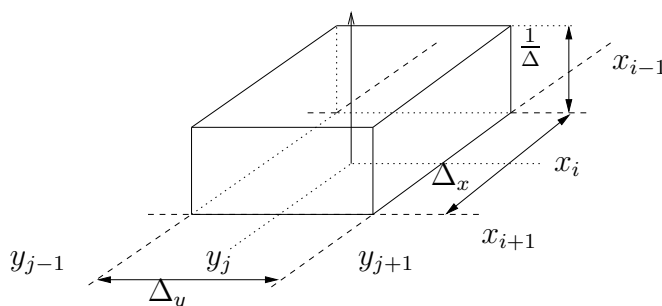


Figure 2.7 – Représentation de l'échantillonnage par une distribution uniforme.

Le choix du noyau uniforme pour représenter l'interpolation nécessite plus de précaution. En effet, si on choisit d'utiliser un noyau centré défini sur le pavé de support (Δ_x, Δ_y) , la propriété définie par l'équation 2.4 ne sera respectée que si :

$$(d_x, d_y) = (\Delta_x, \Delta_y)$$

Si on veut être sûr que la propriété 2.4 soit vérifiée pour toute valeur de (d_x, d_y) , il faut choisir une dimension la plus petite possible. Si on décide de représenter l'échantillonnage et l'interpolation par un noyau uniforme, la combinaison des deux noyaux est un noyau pyramidal dont le support est borné par $(2\Delta_x, 2\Delta_y)$ et non plus un noyau uniforme (sauf si le noyau d'interpolation est une impulsion de Dirac).

Cette représentation a deux inconvénients majeurs :

- ce n'est pas une représentation modale ce qui peut amener à des instabilités de représentation lorsque la zone d'intégration franchit une frontière entre deux noyaux,

- elle confond le fait que l'on ne connaisse pas le support d'intégration avec le fait que l'on suppose que ce support est maximal. En fait, cela revient à adopter, pour toute la famille des supports d'intégration, celui qui a la taille la plus grande (c'est à dire celle du capteur CCD).

2.3.2 Modélisation par un intervalle

Pour modéliser une méconnaissance complète du support d'intégration, il est possible de recourir à la modélisation par intervalle (ou pavé en 2D). Le calcul par intervalle [Moo66], aussi connu sous le nom de calcul d'erreur, modélise l'imprécision d'une mesure.

Soit $Q_{x,y}$ un pavé de \mathbb{R}^2 défini par $[a, b] \times [c, d]$ et P_{x_i, y_j} le pavé associé au pixel (i, j) , $I(i, j)$ est une valeur possible d'illumination du pavé $Q_{x,y}$ si :

$$Q_{x,y} \cap P_{x_i, y_j} \neq \emptyset$$

La principale difficulté de cette modélisation est le choix de la taille de l'intervalle puisqu'il doit contenir l'ensemble des valeurs possibles de la mesure. Ce que permet de représenter un pavé, dans le cas présent, c'est le support maximal d'intégration d'un pixel et donc :

- tout pavé $Q_{x,y}$ contenu dans un pavé P_{x_i, y_j} (figure 2.8a) peut avoir comme illumination la valeur $I(i, j)$,
- tout pavé $Q_{x,y}$ intersectant un ensemble φ de pixels $P_{x,y}$ a une illumination supérieure à la somme des illuminations correspondantes.

Notre problématique ici consiste à interpoler l'image, c'est à dire à associer une valeur d'illumination aux points qui sont en dehors du maillage de l'image. Si on choisit pour P_{x_i, y_j} un pavé centré en (x_i, y_j) et de taille maximale (Δ_x, Δ_y) et pour $Q_{x,y}$ un pavé de même taille mais centré en un point quelconque (x, y) , on voit que tester si $Q_{x,y}$ intersecte P_{x_i, y_j} est équivalent à la problématique visant à exprimer le fait que (x, y) appartient à P'_{x_i, y_j} , le pavé centré en (x_i, y_j) et de taille $(2\Delta_x, 2\Delta_y)$ (figure 2.8b).

Si $Q_{x,y}$ intersecte plusieurs pavés associés à des pixels, alors il est difficile de donner une valeur d'illumination associée à $Q_{x,y}$ à moins de mettre en oeuvre des procédures statistiques non sommatives comme le transfert pignistique [Sme00], [Sme94].

Si $Q_{x,y} = P_{x_i, y_j}$ alors :

$$\hat{I}(x, y) = I_{i,j}$$

sinon :

$$\hat{I}(x, y) = \text{statistique} \{ (x_i, y_j) / Q_{x,y} \cap P_{x_i, y_j} \neq \emptyset \}$$

en utilisant, par exemple, une statistique par intervalle tel que :

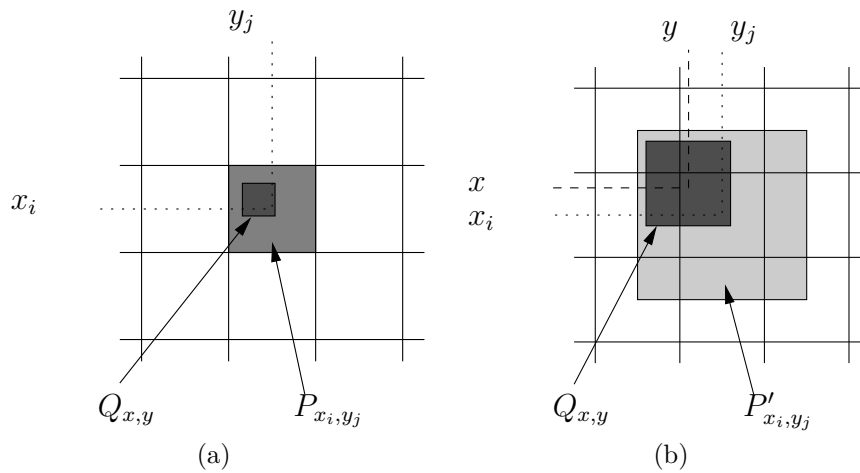


Figure 2.8 – Intervalle inclus dans le pixel (a), Intervalle contenant plusieurs pixels (b).

$$\hat{I}(x, y) = [I_{\min}, I_{\max}]$$

avec :

$$\begin{cases} I_{\min} = \mathop{\text{Inf}}_{I_{i,j} \subset P'} \{I_{i,j}\} \\ I_{\max} = \mathop{\text{Sup}}_{I_{i,j} \subset P'} \{I_{i,j}\} \end{cases}$$

Il est notoire que l'aspect tout ou rien de la représentation par intervalle ainsi que la difficulté de représentation de statistiques associées aux intervalles en font souvent un outil peu commode à utiliser et particulièrement dans le cadre que nous nous sommes fixé. B. Telle [Tel03] a montré néanmoins qu'il était possible de modéliser les problèmes d'échantillonnage mais uniquement dans le cadre de la reconstruction tridimensionnelle.

Il est possible de revenir à une signification probabiliste des intervalles en considérant un pavé comme intervalle de confiance à un niveau donné.

2.3.3 Modélisation mixte : les intervalles de confiance

L'intervalle de confiance de niveau β d'un noyau sommatif $n(x)$ est un intervalle de l'espace considéré tel que :

$$\int n(x) dx = \beta$$

avec $\beta \in [0, 1]$. La version bidimensionnelle d'un intervalle de confiance est un pavé Q tel que :

$$\int \int_Q n(x, y) dx dy = \beta$$

De cette définition, il vient qu'il existe, pour un niveau β donné, une infinité d'intervalles de confiance.

Nous avons exprimé le fait (§2.3.1) que les noyaux sommatifs que nous utilisons ont un support borné de façon à modéliser uniquement la dégradation de l'information induite par l'échantillonnage. Si on ne considère que les noyaux unimodaux symétriques, le mode, la moyenne et la médiane sont confondus et le plus petit intervalle de niveau β est celui qui est centré sur le mode du noyau [DPS93]. On peut supposer que cette propriété se retrouve en 2D. Si $n_{i,j}(x, y)$ est un noyau 2D unimodal symétrique centré sur (x_i, y_j) , ses projections suivant les axes x et y sont deux distributions unimodales symétriques centrées en x_i et y_j . Le produit cartésien des deux intervalles de confiance en x et y de niveau β forme Q^β , le pavé de confiance qui doit être le plus petit au sens de la cardinalité de niveau β centré sur (x_i, y_j) .

Dans le cas des noyaux uniformes, seules la moyenne et la médiane sont confondues. Ainsi, tous les intervalles de confiance pour une même valeur β ont la même surface puisque aucune valeur autour de laquelle construire les intervalles de confiance ne peut être privilégiée. V. Lasserre, dans sa thèse [Las99], propose de traiter ce cas de la même façon que lorsqu'il s'agit de noyaux modaux et choisit de centrer les intervalles sur le centre du mode (qui est imprécis) et qui correspond aussi à la valeur moyenne ainsi qu'à la médiane.

Il est intéressant de constater que l'ensemble des intervalles ainsi définis sont emboîtés dans le sens des β croissants (Figure 2.9).

Il pourrait être intéressant d'essayer de travailler sur cet ensemble emboîté d'intervalles de confiance. C'est l'approche que propose Dubois, Foulloy, Mauris et Prade dans [DFMP04] et sur laquelle nous nous appuyons dans ce travail.

2.3.4 Modélisation floue

La théorie des ensembles flous [Zad65] a été proposée par L.A. Zadeh en 1965 et elle est associée à la théorie des possibilités [Zad78] afin de représenter mathématiquement l'imprécision relative de certaines informations. Dans cette théorie interviennent les notions de mesures de possibilité et de nécessité popularisées par Dubois et Prade [DP88]. Il existe de nombreux ouvrages traitant du flou, nous conseillons aux lecteurs de se référer à ces ouvrages pour plus d'informations [DP88], [BM93], [Nov98].

2.3.4.1 Interprétation d'une distribution de possibilité

Il existe deux manières différentes de voir une distribution de possibilité [Las99] :

- vision verticale : La distribution de possibilité peut être vue comme une borne supérieure d'une famille de distribution de probabilité [DP92] :

$$\mathcal{P}(\Pi) = \{P, \forall A \subseteq \Omega, P(A) \leq \Pi(A)\}$$

- vision horizontale : L'ensemble flou normalisé que constitue une distribution de possibilité peut être vu par ses α -coupes comme un ensemble d'intervalle de

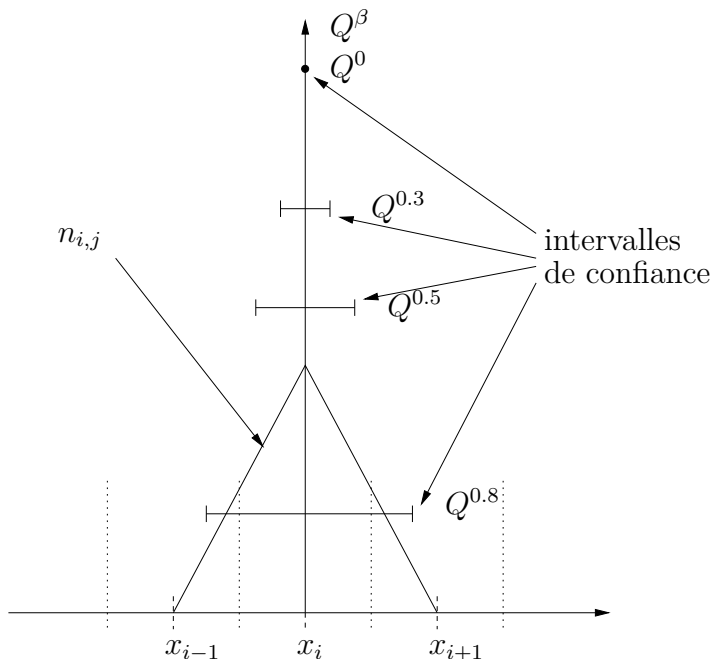


Figure 2.9 – Représentation d'intervalles de confiance associés à un noyau triangulaire.

confiance. Une distribution de possibilités peut donc être construite à partir de la connaissance d'un ou plusieurs de ces intervalles.

2.3.4.2 Modélisation d'une distribution de probabilité par une distribution de possibilité

Soit $n(x)$ un noyau sommatif unimodal et symétrique, on peut associer à $n(x)$ l'ensemble emboîté de ses intervalles de confiance β . Le niveau de confiance β d'un intervalle correspond à la limite du pourcentage de valeurs issues de la distribution de noyau $n(x)$ qui appartiennent à cet intervalle [DP92]. Cette conception de la confiance est l'opposée de celle représentée par le niveau α d'un sous-ensemble flou. Une coupe \mathcal{A}^α de niveau α d'un ensemble flou représente l'ensemble des valeurs correspondant au concept synthétisé par l'ensemble flou \mathcal{A} avec un niveau de confiance supérieur ou égal à α . On passe de la première à la seconde représentation par [Las99] :

$$\beta = 1 - \alpha$$

Il est alors possible d'associer à chaque distribution de probabilité une représentation par intervalles de confiances emboîtés. Cet ensemble d'intervalles de confiance emboîtés permet de définir une distribution de possibilité optimale [DPS93] respectant les propriétés suivantes :

- Propriété de cohérence : $\forall A \subseteq \Omega, P(A) \leq \Pi(A)$,
- Principe de préservation de la préférence : $\forall (a, b) \in \Omega^2 :$

$$\begin{cases} \text{si } p(a) > p(b) \Rightarrow \pi(a) > \pi(b) \\ \text{si } p(a) = p(b) \Rightarrow \pi(a) = \pi(b) \end{cases}$$

- Principe du maximum de spécificité : la surface en dessous de la distribution résultante doit être minimale (dans le sens de la cardinalité du sous-ensemble flou).

2.3.5 Modélisation du pixel par une distribution de possibilité pyramidale symétrique

Toute distribution de possibilité optimale associée à une distribution de probabilité unimodale, symétrique et bornée n'a aucun point d'inflexion sur \mathbb{R}^2 . Elle a une forme incurvée autour du mode [Las99]. Toutes les distributions de possibilité optimales définies autour d'un même mode et ayant le même support peuvent être majorées par une distribution de possibilité triangulaire de même mode et de même support [DFMP04].

Nous avons vu que le plus grand intervalle permettant de modéliser les effets de l'échantillonnage sur l'information d'illumination était le pavé centré en (x_i, y_j) et d'étalement $(2\Delta_x, 2\Delta_y)$. L'ensemble des noyaux sommatifs associés à cette représentation a pour mode (x_i, y_j) et un support inclus dans le pavé défini par les 8 voisins de chaque pixel (Figure 2.10). Chacun de ces noyaux se projette sur l'axe x (respectivement y) en une distribution symétrique centrée sur x_i (respectivement y_j) dont le support est inclus dans $]x_i - \Delta_x, x_i + \Delta_x[$ (respectivement $]y_j - \Delta_y, y_j + \Delta_y[$). Toutes ces distributions pouvant être représentées par le nombre flou \mathcal{X}_i (respectivement \mathcal{Y}_j), nous proposons d'associer à tout pixel $m_{i,j}$ un ensemble flou (pavé) $\mathcal{M}_{i,j}$ pyramidal obtenu par le produit cartésien de \mathcal{X}_i et \mathcal{Y}_j . Nous avons de bonne raison de penser que le produit cartésien des 2 ensembles triangulaires domine de la façon la plus spécifique toutes les distributions de possibilité 2D calculées à partir des noyaux sommatifs 2D.

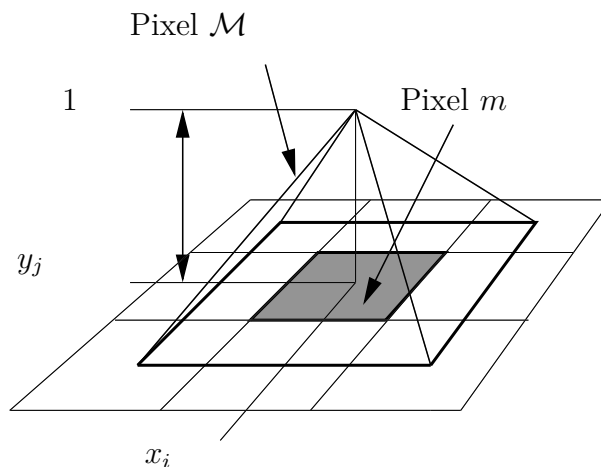
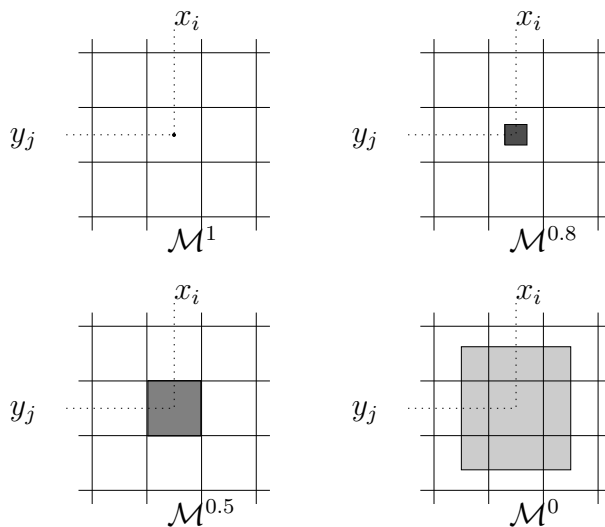


Figure 2.10 – Représentation floue d'un pixel.

Cette représentation permet de modéliser, entre autre, une information ponctuelle

Figure 2.11 – Localisation du pixel flou par α -coupes.

(cas du pixel \mathcal{M}^1 sur la figure 2.11) repéré par les coordonnées exactes lignes et colonnes, un intervalle de niveau de confiance 0.5 délimité par les coordonnées du pixel lui-même (cas du pixel $\mathcal{M}^{0.5}$ sur la figure 2.11), ou l'information de localisation dans le pire cas (cas du pixel \mathcal{M}^0 sur la figure 2.11, c'est à dire un intervalle de la taille du support centré sur le centre du pixel et de niveau de confiance 1).

Cette modélisation garantit de contenir tous les intervalles de confiance construit à partir de toute distribution de probabilité ayant même mode et même support.

2.4 Pixel, aspect chromatique

L'objectif du modèle que nous venons de définir est de minimiser l'impact de l'échantillonnage sur l'utilisation de la comparaison des luminances pour la réalisation d'appariements. Il existe cependant de nombreuses autres altérations de la variation de luminance entre deux pixels "correspondants". Ces différences de luminance peuvent être dues aux discontinuités de profondeur, aux occultations, au flou liée à la lentille de chaque caméra, aux bruits électroniques et surtout à la quantification de l'information de luminance et à la saturation des capteurs photosensibles.

La majeure partie des méthodes classiques de mise en correspondance utilisent une représentation statistique de l'erreur d'appariement [Sch94b], [ZW94], [GLY95], [CMPC96], [Bel96], [CHR96]. On est en droit de se poser des questions sur l'utilisation de distances statistiques lorsque le nombre d'informations à mettre en correspondance ne dépasse pas deux. Les techniques de régularisation s'appuient sur l'hypothèse d'ergodicité du bruit de mise en correspondance et utilisent le voisinage de chaque pixel pour évaluer une représentation de la statistique des variations des niveaux de gris. Plus le voisinage utilisé est étendu, plus la statistique est précise, mais moins elle est fiable, c'est à dire représentative de la variation locale. Ce manque de fiabilité est lié,

comme nous l'avons vu dans le chapitre précédent, aux régions ambiguës. Les mesures de corrélation de motifs ou de distances numériques sur les intensités sont particulièrement perturbées par des variations non ergodiques provoquées par des phénomènes qui peuvent difficilement être modélisés par de simples lois normales comme les occultations, etc.... Enfin, le phénomène de saturation, à notre connaissance, n'est jamais pris en compte.

2.4.1 Représentation logique de l'intensité lumineuse

F. Comby a proposé, au cours de travaux portant sur la détection du mouvement majoritaire dans une séquence d'images, d'utiliser une classification floue des niveaux de gris pour rendre plus robuste l'analyse des appariements. Dans [Com01], il propose une modélisation reposant sur une représentation par classe des phénomènes de saturation de l'information de luminance. F. Comby a montré que cette modélisation permet de minimiser l'a priori dans l'utilisation des informations de niveaux de gris pour la mise en correspondance et de représenter l'imprécision des valeurs de niveaux de gris.

Nous proposons de reprendre cette modélisation chromatique de l'information de luminance et de l'appliquer à notre modèle de pixel dans le but d'augmenter la robustesse du processus de mise en correspondance stéréoscopique.

On modélise le niveau de gris d'un pixel par sa compatibilité avec une des deux classes que sont le noir (niveau d'excitation du pixel inférieur au seuil de saturation bas) et le blanc (niveau d'excitation du pixel supérieur au seuil de saturation haut). Ces compatibilités sont représentées par les classes duales "pixels blancs" et "pixels noirs" comme il est proposé dans [Com01]. Les phénomènes d'instabilité au niveau de la frontière entre les deux classes sont minimisés par une représentation floue (figure 2.12).

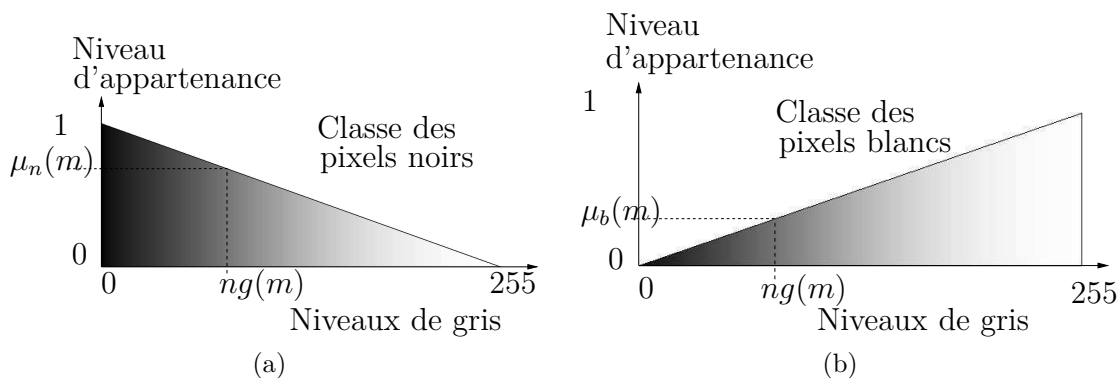


Figure 2.12 – Classe des pixels noirs (a), Classe des pixels blancs (b).

Chaque pixel m de niveau de gris $ng(m)$ est représenté par son degré d'appartenance à la classe des pixels blancs $\mu_b(m)$ et son degré d'appartenance à la classe des pixels noirs $\mu_n(m)$:

$$\begin{cases} \mu_b(m) = \frac{ng(m)}{255} \\ \mu_n(m) = 1 - \frac{ng(m)}{255} \end{cases}$$

Il existe une propriété très intéressante résultant de la modélisation floue du niveau de gris des pixels en conjonction avec le processus bipolaire. Cette propriété permet de s'affranchir des problèmes des effets de bord dans le processus de mise en correspondance. Puisque la valeur du niveau de gris du pixel extérieur nous est inconnue, elle peut être aussi bien blanche que noir. Ainsi, pour tout pixel extérieur à l'image :

$$\mu_b(m) = \mu_n(m) = 1$$

Cette classification peut prendre en compte la quantification de la valeur de luminosité en remplaçant les valeurs d'appartenance $\mu_b(m)$ et $\mu_n(m)$ par des mesures de possibilités $\Pi(\text{blanc}, \mathcal{NG}(m))$ et $\Pi(\text{noir}, \mathcal{NG}(m))$ et de nécessités $\mathcal{N}(\text{blanc}, \mathcal{NG}(m))$ et $\mathcal{N}(\text{noir}, \mathcal{NG}(m))$ avec :

$$\mathcal{NG}(m) = ng(m) \pm \Delta ng$$

Que vaut Δng ? Prenons un exemple simple pour le déterminer avec $ng(m) = 100$. Si le système de quantification effectue une troncature basse sur $\widetilde{ng}(m)$, la valeur non quantifiée ayant conduit à obtenir $ng(m)$, alors cette valeur est telle que :

$$100 \leq \widetilde{ng}(m) < 101$$

Si le système de quantification effectue une troncature haute alors $\widetilde{ng}(m)$ est telle que :

$$99 \leq \widetilde{ng}(m) < 100$$

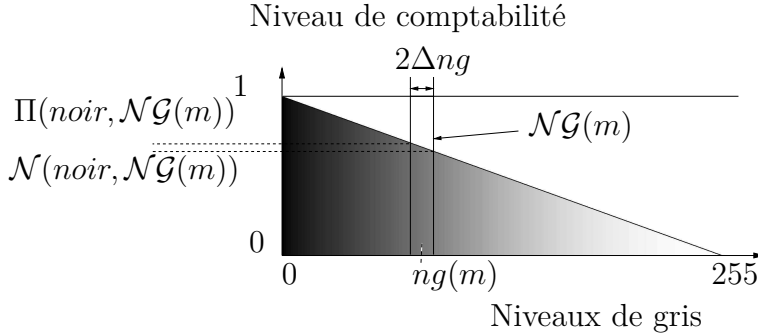
Enfin si le système arrondi la valeur alors $\widetilde{ng}(m)$ est telle que :

$$99.5 \leq \widetilde{ng}(m) < 100.5$$

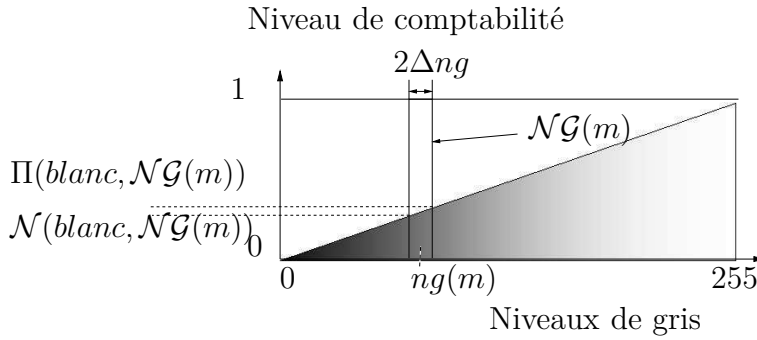
Le minimum d'apriorisme sur le type de quantification nous conduit à considérer $\widetilde{ng}(m) \in]99; 101[$. Δng est donc égal à 1. Les valeurs de possibilités et de nécessités sont obtenues par :

$$\begin{cases} \Pi(\text{blanc}, \mathcal{NG}(m)) = \underset{\text{Niveaux de gris}}{\text{Sup}} \quad \min(\text{blanc}, \mathcal{NG}(m)) \\ \mathcal{N}(\text{blanc}, \mathcal{NG}(m)) = \underset{\text{Niveaux de gris}}{\text{Inf}} \quad (\max(\text{blanc}, 1 - \mathcal{NG}(m))) \end{cases}$$

$$\begin{cases} \Pi(\text{noir}, \mathcal{NG}(m)) = \underset{\text{Niveaux de gris}}{\text{Sup}} (\min(\text{noir}, \mathcal{NG}(m))) \\ \mathcal{N}(\text{noir}, \mathcal{NG}(m)) = \underset{\text{Niveaux de gris}}{\text{Inf}} (\max(\text{noir}, 1 - \mathcal{NG}(m))) \end{cases}$$



(a)



(b)

Figure 2.13 – Compatibilité avec la classe des pixels noirs (a), Compatibilité avec la classe des pixels blancs (b).

2.4.2 Evaluation de l'information chromatique du pixel flou

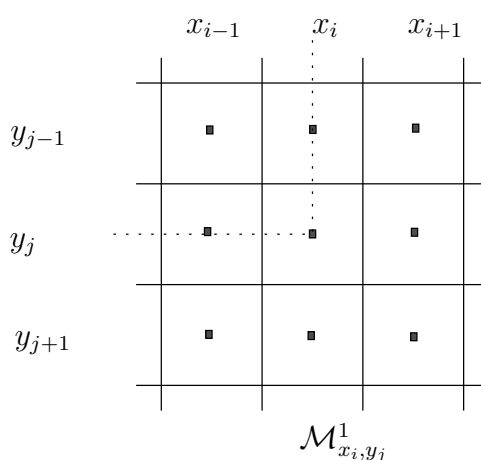
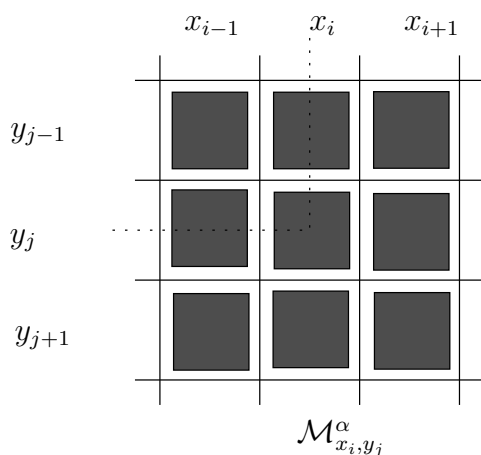
L'objet de ce paragraphe est de déterminer les valeurs de possibilité et nécessité de compatibilité chromatique avec les classes Blanc et Noir de notre pixel flou pour tous ses niveaux de coupe α . Nous allons montrer que, grâce à l'utilisation des α -coupes, nous pouvons prendre en compte les informations d'illumination du voisinage. Soit $\mathcal{M}_{x_i, y_j}^\alpha$ le pavé correspondant à la coupe α du pixel \mathcal{M}_{x_i, y_j} .

2.4.2.1 Cas du pixel \mathcal{M}^α avec $\alpha \in]0.5, 1]$

Si $\alpha \in]0.5, 1]$ alors $\mathcal{M}_{x_i, y_j}^\alpha$ la coupe de niveau α de \mathcal{M}_{x_i, y_j} n'intersecte aucune coupe de niveau α d'un autre pixel $\mathcal{M}_{x'_i, y'_j}^\alpha$ avec $x'_i \neq x_i$ et $y'_j \neq y_j$ (figure 2.14 et 2.15). La valeur de compatibilité de la luminance de la coupe de niveau α du pixel \mathcal{M}_{x_i, y_j} est déterminée par :

$$\left\{ \begin{array}{l} \Pi(\text{blanc}, \mathcal{NG}(\mathcal{M}_{x_i, y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Sup}} (\min(\text{blanc}, \mathcal{NG}(m))) \\ \mathcal{N}(\text{blanc}, \mathcal{NG}(\mathcal{M}_{x_i, y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Inf}} (\max(\text{blanc}, 1 - \mathcal{NG}(m))) \end{array} \right.$$

$$\left\{ \begin{array}{l} \Pi(\text{noir}, \mathcal{NG}(\mathcal{M}_{x_i, y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Sup}} (\min(\text{noir}, \mathcal{NG}(m))) \\ \mathcal{N}(\text{noir}, \mathcal{NG}(\mathcal{M}_{x_i, y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Inf}} (\max(\text{noir}, 1 - \mathcal{NG}(m))) \end{array} \right.$$

Figure 2.14 – Evaluation des informations chromatiques du pixel flou ($\alpha = 1$).Figure 2.15 – Evaluation des informations chromatiques du pixel flou ($\alpha \in]0.5, 1]$).

2.4.2.2 Cas du pixel \mathcal{M}^α avec $\alpha \in [0, 0.5]$

Si $\alpha \in [0, 0.5]$, alors $\mathcal{M}_{x_i, y_j}^\alpha$ intersecte les coupes de niveau α de ses huit voisins (figure 2.16). L'information de compatibilité de luminance est donc imprécise. Soit

$\mathcal{V}(\mathcal{M}_{x_i,y_j})$, l'union de l'ensemble des \mathcal{NG} des 9 pixels considérés, l'information de compatibilité de luminance peut être représentée par ses deux bornes :

$$\left\{ \begin{array}{l} \Pi(\text{blanc}, \mathcal{NG}(\mathcal{M}_{x_i,y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Sup}} (\min(\text{blanc}, \mathcal{V}(\mathcal{M}_{x_i,y_j}))) \\ \mathcal{N}(\text{blanc}, \mathcal{NG}(\mathcal{M}_{x_i,y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Inf}} (\max(\text{blanc}, 1 - \mathcal{V}(\mathcal{M}_{x_i,y_j}))) \end{array} \right.$$

$$\left\{ \begin{array}{l} \Pi(\text{noir}, \mathcal{NG}(\mathcal{M}_{x_i,y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Sup}} (\min(\text{noir}, \mathcal{V}(\mathcal{M}_{x_i,y_j}))) \\ \mathcal{N}(\text{noir}, \mathcal{NG}(\mathcal{M}_{x_i,y_j}^\alpha)) = \underset{\text{Niveaux de gris}}{\text{Inf}} (\max(\text{noir}, 1 - \mathcal{V}(\mathcal{M}_{x_i,y_j}))) \end{array} \right.$$

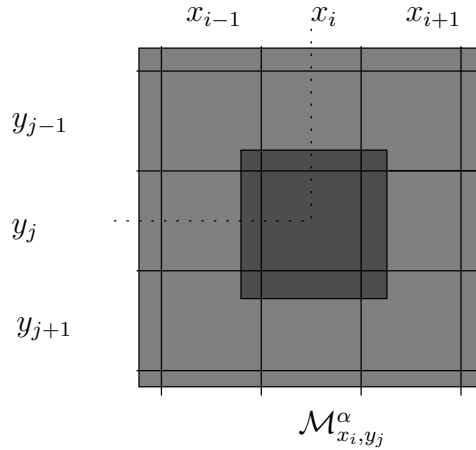


Figure 2.16 – Evaluation de la compatibilité de l'information de luminance ($\alpha \in [0, 0.5]$).

2.5 Mise en correspondance stéréoscopique

Nous proposons dans ce paragraphe de montrer comment utiliser notre modèle de pixel dans le calcul d'une mise en correspondance entre un pixel de l'image 1 et un pixel de l'image 2 en utilisant uniquement les informations chromatiques et épipolaires.

2.5.1 Concept de base

Plutôt que de modéliser la mise en correspondance d'un pixel \mathcal{M}_1 de l'image 1 avec un pixel \mathcal{M}_2 de l'image 2 par une mesure de probabilité (trop normative) ou un domaine ensembliste (pas assez nuancé), nous proposons ici d'utiliser la théorie des possibilités pour associer à chaque pixel \mathcal{M}_1 de l'image 1 un domaine flou $\mathcal{D}_2(\mathcal{M}_1)$ de l'image 2 des pixels pouvant être mis en correspondance avec \mathcal{M}_1 (figure 2.17). En remplaçant les règles classiques de restriction d'appariement par leur équivalent flou, on peut établir que :

Proposition 1

la mise en correspondance potentielle entre un pixel \mathcal{M}_1 de l'image 1 et un pixel \mathcal{M}_2 de l'image 2 est mesurée par la possibilité $\Pi(\mathcal{D}_2(\mathcal{M}_1), \mathcal{M}_2) = \Pi(\mathcal{D}_1(\mathcal{M}_2), \mathcal{M}_1)$.

Cette possibilité est définie de façon symétrique aussi bien par $\mathcal{D}_2(\mathcal{M}_1)$ que par $\mathcal{D}_1(\mathcal{M}_2)$.

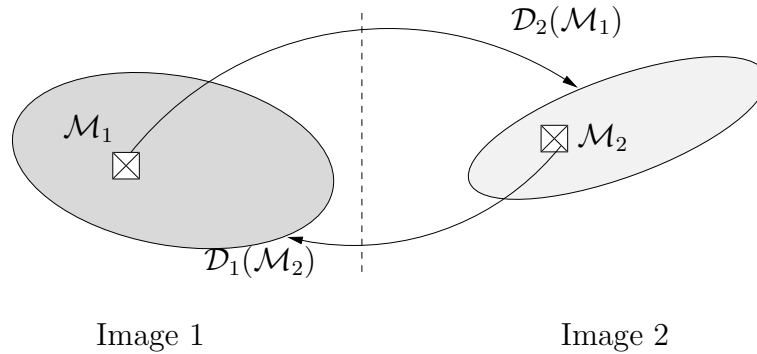


Figure 2.17 – Domaines de mise en correspondance

Chaque domaine flou peut être vu comme un emboîtement de domaines binaires affectés de leur mesure de confiance. Cette représentation des ensembles flous par α -coupe [DP94] permet de raisonner plus facilement avec des concepts ensemblistes ordinaires.

2.5.2 Processus bipolaire

Dans la représentation de l'information d'appariement, chaque énoncé de contrainte restreint le nombre possible de candidats. Si, après l'étude d'une contrainte, certains candidats sont déclarés comme impossibles, les autres restent non-impossibles, ce qui peut augmenter leur nécessité. On doit donc faire une distinction entre information positive (exemple : en accord avec la contrainte d'appariement \mathcal{C} le pixel \mathcal{M}_1 est correspondant du pixel \mathcal{M}_2) et information négative (exemple : connaissant l'appariement entre les pixels \mathcal{M}_1 et \mathcal{M}_2 , le pixel \mathcal{M}'_1 ne peut pas être le correspondant du pixel \mathcal{M}'_2). Ces deux états représentent des informations incomplètes aussi bien négatives (restrictions, impossibilités, rejets) que positives (confirmations, renforcement). Comme l'on présenté Dubois et Prade [DD03], ces deux informations ne sont pas de même nature et leur traitement relève de techniques différentes. Les informations négatives se combinent par conjonction alors que les informations positives se combinent par disjonction.

Enfin, les mesures de possibilités ne respectent ni le tiers exclus, ni la non contradiction, il est donc nécessaire de baser la mise en correspondance sur un processus bipolaire de décision [BDKP02]. Ainsi, un même critère peut, pour un niveau de confiance donné, à la fois accepter et refuser un appariement sans pour cela contredire l'axiomatique des possibilités. C'est la prise en compte de cette propriété que l'on nomme bipolarité.

Par la suite, nous utiliserons les notations suivantes :

- la possibilité de mise en correspondance entre \mathcal{M}_1 de l'image 1 et \mathcal{M}_2 de l'image 2 est mesurée par $\Pi(\mathcal{M}_1, \mathcal{M}_2)$,
- la possibilité de non mise en correspondance entre \mathcal{M}_1 et \mathcal{M}_2 est mesurée par $\overline{\Pi(\mathcal{M}_1, \mathcal{M}_2)}$,
- la non possibilité de ne pas mettre en correspondance \mathcal{M}_1 et \mathcal{M}_2 aussi appelée nécessité est mesurée par $\mathcal{N}(\mathcal{M}_1, \mathcal{M}_2) = 1 - \overline{\Pi(\mathcal{M}_1, \mathcal{M}_2)}$.

2.5.3 Mise en correspondance géométrique dans le cas d'un capteur stéréoscopique étalonné avec erreur

Nous nous abstrayons, dans un premier temps, du phénomène d'échantillonnage. Soient m_1 et m_2 les deux pixels comportant les deux projections dans les différents plans images d'un même point tridimensionnel P . Lorsque l'on dispose d'un système stéréoscopique étalonné, les positions de m_1 et m_2 sont contraintes par la géométrie épipolaire. Cette contrainte symétrique est explicitée par l'équation :

$$m_2^t F m_1 = 0 \quad (2.5)$$

Projetée sur l'image 2, cette contrainte exprime le fait que tout pixel de l'image 2 pouvant être mis en correspondance avec un pixel m_1 appartient à la droite $l_2(m_1)$ définie par :

$$l_2(m_1) = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = F m_1 = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} m_1 \quad (2.6)$$

Comme nous l'avons présenté au paragraphe 1.2.2, ce cas idéal n'est pas réaliste. En effet, l'échantillonnage des images et la mauvaise connaissance de la matrice fondamentale F nous conduit à considérer cette relation sous un aspect ensembliste. Pour être cohérent avec notre proposition initiale, nous représentons la distribution de chaque paramètre de la matrice F par un ensemble flou triangulaire $\mathcal{F}_{r,s}$ (Figure 2.18a) dont le mode est défini par la valeur la plus probable produite par l'étalonnage du capteur stéréoscopique (c'est à dire f_{rs}) et dont le support est calculé par une analyse "pire cas" des erreurs d'étalonnage [Tel03].

La relation (2.6) ne traduit plus une ligne épipolaire sur l'image 2 mais un bandeau épipolaire flou dont une représentation naïve est proposée sur la figure 2.18c. Ce bandeau est défini autour de la ligne épipolaire $|l_2(m_1)|$, ligne épipolaire la plus probable calculée à partir des informations du capteur et du pixel "ponctuel" $\mathcal{M}_{m_1}^0$ ou m_1 . Les bornes $\underline{l_2(m_1)}$ et $\overline{l_2(m_1)}$ sont calculées par l'analyse pire cas et définissent les limites du domaine de mise en correspondance $\mathcal{D}_2(m_1)$.

De même que la contrainte épipolaire met en relation un sous-ensemble de pixels de l'image 2 avec chaque pixel de l'image 1, la contrainte épipolaire floue ainsi définie permet de mettre en correspondance un sous-ensemble flou de l'image 2 avec chaque pixel de l'image 1. La possibilité $\Pi_G(\mathcal{M}_1, \mathcal{M}_2)$ de mettre en correspondance géométriquement un pixel flou \mathcal{M}_1 de l'image 1 avec un pixel flou \mathcal{M}_2 de l'image 2 s'obtient

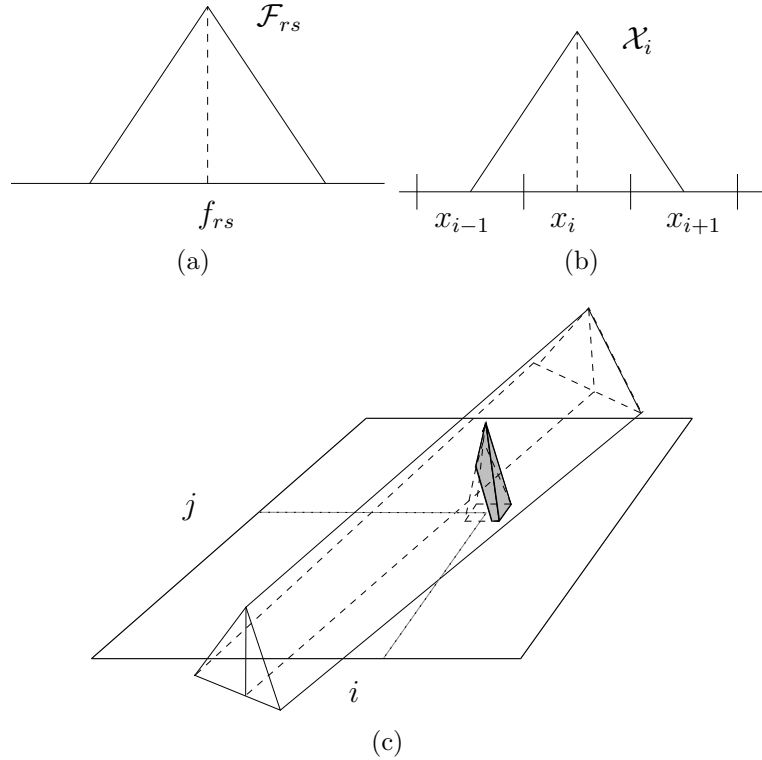


Figure 2.18 – Sous-ensemble flou de distribution triangulaire pour représenter chaque paramètre de la matrice fondamentale (a), Sous-ensemble flou de distribution triangulaire pour représenter les coordonnées ligne et colonne d'un pixel (b), Intersection entre le bandeau épipolaire flou et un pixel flou (c).

naturellement en utilisant la relation épipolaire (2.5) et le principe d'extension de la façon suivante :

$$\Pi_G(\mathcal{M}_1, \mathcal{M}_2) = \underset{\substack{f_{rs}, m_k \\ r, s=1..3 \\ k=1..2}}{\text{Sup}} \left\{ \min(\mu_{\mathcal{F}_{rs}}(f_{rs}), \mu_{\mathcal{M}_k}(m_k)) / m_2^t F m_1 = 0 \right\} \quad (2.7)$$

Cette écriture étant peu pratique à mettre en oeuvre, nous pouvons la modifier en nous plaçant toujours dans le "pire cas". En se servant des paramètres de la ligne épipolaire définis par l'équation (2.6) et en remplaçant le pixel m_1 par son modèle flou (les coordonnées ligne et colonne sont remplacées par une distribution triangulaire centrée sur le centre du pixel comme représenté sur la figure 2.18b) :

$$\mathcal{M}_1 = (\mathcal{X}_i, \mathcal{Y}_j, 1)$$

l'équation de la droite épipolaire peut être ré-écrite sous la forme ensembliste par :

$$\begin{pmatrix} \mathcal{A}_2 \\ \mathcal{B}_2 \\ \mathcal{C}_2 \end{pmatrix} = \mathcal{F} \odot \mathcal{M}_1$$

$$\begin{cases} \mathcal{A}_2 = \mathcal{F}_{11} \odot \mathcal{X}_i \oplus \mathcal{F}_{12} \odot \mathcal{Y}_j \oplus \mathcal{F}_{13} \\ \mathcal{B}_2 = \mathcal{F}_{21} \odot \mathcal{X}_i \oplus \mathcal{F}_{22} \odot \mathcal{Y}_j \oplus \mathcal{F}_{23} \\ \mathcal{C}_2 = \mathcal{F}_{31} \odot \mathcal{X}_i \oplus \mathcal{F}_{32} \odot \mathcal{Y}_j \oplus \mathcal{F}_{33} \end{cases} \quad (2.8)$$

Les notations \odot et \oplus représentent respectivement les opérateurs arithmétiques de la multiplication et de l'addition appliqués aux ensembles flous [DP80]. En se servant des paramètres de la ligne épipolaire définis par l'équation (2.8) et en remplaçant le pixel m_2 par son modèle flou \mathcal{M}_2 :

$$\mathcal{M}_2 = (\mathcal{X}_{i'}, \mathcal{Y}_{j'}, 1)^t$$

Soit le sous-ensemble flou \mathcal{ABC} défini par :

$$\mathcal{ABC} = \mathcal{A} \odot \mathcal{X}_{i'} \oplus \mathcal{B} \odot \mathcal{Y}_{j'} \oplus \mathcal{C}$$

La contrainte épipolaire revient tout simplement à chercher à quel point 0 appartient à \mathcal{ABC} . Nous pouvons alors réécrire la possibilité de mise en correspondance géométrique (2.7) :

$$\Pi_G(\mathcal{M}_1, \mathcal{M}_2) = \mu_{\mathcal{ABC}}(0) \quad (2.9)$$

Notre processus étant basé sur une représentation bipolaire des informations, nous pouvons établir une mesure de non compatibilité géométrique entre deux pixels notée $\overline{\Pi_G(\mathcal{M}_1, \mathcal{M}_2)}$. Cela revient à chercher à quel point 0 n'appartient pas à \mathcal{ABC} . On peut alors calculer à quel point il est possible de ne pas mettre en correspondance géométriquement nos deux pixels par :

$$\overline{\Pi_G(\mathcal{M}_1, \mathcal{M}_2)} = 1 - \mu_{\mathcal{ABC}}(0) \quad (2.10)$$

2.5.4 Mise en correspondance chromatique

La mise en correspondance de deux pixels à partir de leur niveau de gris s'appuie sur l'hypothèse que les projections d'un même point P de l'espace tridimensionnel ont des intensités lumineuses comparables. On peut formuler la proposition suivante :

dans deux images en niveaux de gris acquises par un capteur stéréoscopique, un point tridimensionnel dont la projection dans la première image est un pixel blanc a peu de chance d'avoir comme projection dans la deuxième image un pixel noir et vice-versa.

Partant de cette hypothèse "faible", nous proposons d'utiliser la classification floue des niveaux de gris pour l'analyse des appariements. Au cours de sa thèse [Com01], F.

Comby a développé des outils (histogrammes quasi-continus [SC02]) qui s'appuient sur une version robuste des statistiques de modes. Cette approche ne peut être utilisée ici en raison du peu d'informations disponibles (voisinage considéré restreint).

Nous proposons de reprendre cette modélisation en remplaçant les outils d'analyse statistiques par une analyse par règles. La possibilité de mettre en correspondance deux pixels en s'appuyant sur les informations chromatiques peut être mesurée par :

Définition 1 *La mise en correspondance de deux pixels m_1 et m_2 , projections du même point P , est envisageable si ces deux pixels appartiennent à la même classe. C'est à dire (" m_1 est blanc" ET " m_2 est blanc") OU (" m_1 est noir" ET " m_2 est noir").*

L'utilisation de cette proposition en logique floue s'exprime simplement par la formule :

$$\Pi_C(\mathcal{M}_1, \mathcal{M}_2) = \max \left\{ \begin{array}{l} \min [\Pi(\text{blanc}, \mathcal{NG}(\mathcal{M}_1)), \Pi(\text{blanc}, \mathcal{NG}(\mathcal{M}_2))] , \\ \min [\Pi(\text{noir}, \mathcal{NG}(\mathcal{M}_1)), \Pi(\text{noir}, \mathcal{NG}(\mathcal{M}_2))] \end{array} \right\} \quad (2.11)$$

ou $\Pi_C(\mathcal{M}_1, \mathcal{M}_2)$ est une mesure de possibilité et $\Pi(\text{blanc}, \mathcal{NG}(\mathcal{M}))$ (respectivement $\Pi(\text{noir}, \mathcal{NG}(\mathcal{M}))$) représente le degré de compatibilité de luminance du pixel flou \mathcal{M} à la classe des pixels blancs (respectivement noirs). Ce n'est pas une mesure de similarité mais une mesure de "co-appartenance" à une même classe. Elle est beaucoup moins sensible à une mauvaise modélisation et à des bruits non-ergodiques [Com01]. Par contre, elle nécessite l'emploi d'une représentation bipolaire. La possibilité de ne pas mettre en correspondance deux pixels en s'appuyant sur les informations chromatiques peut être mesurée par :

Définition 2 *La mise en correspondance de deux pixels m_1 et m_2 , projections du même point P , est impossible si ces deux pixels appartiennent à des classes différentes. C'est à dire (" m_1 est blanc" ET " m_2 est noir") OU (" m_1 est noir" ET " m_2 est blanc").*

La "non co-appartenance" à la même classe est mesurée par :

$$\overline{\Pi_C(\mathcal{M}_1, \mathcal{M}_2)} = \max \left\{ \begin{array}{l} \min [\Pi(\text{blanc}, \mathcal{NG}(\mathcal{M}_1)), \Pi(\text{noir}, \mathcal{NG}(\mathcal{M}_2))] , \\ \min [\Pi(\text{noir}, \mathcal{NG}(\mathcal{M}_1)), \Pi(\text{blanc}, \mathcal{NG}(\mathcal{M}_2))] \end{array} \right\} \quad (2.12)$$

avec :

$$\overline{\Pi_C(\mathcal{M}_1, \mathcal{M}_2)} \geq 1 - \Pi_C(\mathcal{M}_1, \mathcal{M}_2)$$

c'est à dire :

$$\mathcal{N}_C(\mathcal{M}_1, \mathcal{M}_2) \leq \Pi_C(\mathcal{M}_1, \mathcal{M}_2)$$

2.5.5 Fusion des contraintes de mise en correspondance

En considérant uniquement la mise en correspondance des pixels \mathcal{M}_1 et \mathcal{M}_2 , alors seules sont disponibles les informations portant sur les contraintes chromatiques et géométriques. Si on veut faire le minimum d'apriorisme sur la mise en correspondance,

on établit que, pour que ces deux pixels puissent être en correspondance, il faut qu'ils soient compatibles aussi bien chromatiquement que géométriquement. En utilisant la t -norme conjonctive min, cela se traduit par :

$$\Pi(\mathcal{M}_1, \mathcal{M}_2) = \min(\Pi_C(\mathcal{M}_1, \mathcal{M}_2), \Pi_G(\mathcal{M}_1, \mathcal{M}_2)) \quad (2.13)$$

avec $\Pi(\mathcal{M}_1, \mathcal{M}_2)$ la possibilité de mise en correspondance du pixel \mathcal{M}_1 avec le pixel \mathcal{M}_2 . L'utilisation de l'opérateur min est justifiée par le fait que c'est la plus grande des opérations d'intersection [DP88].

De la même façon, nous pouvons établir une mesure de possibilité de ne pas mettre en correspondance les pixels \mathcal{M}_1 et \mathcal{M}_2 , en choisissant comme t -norme disjonctive max :

$$\overline{\Pi(\mathcal{M}_1, \mathcal{M}_2)} = \max(\overline{\Pi_C(\mathcal{M}_1, \mathcal{M}_2)}, \overline{\Pi_G(\mathcal{M}_1, \mathcal{M}_2)}) \quad (2.14)$$

avec l'utilisation de l'opérateur max comme la plus grande des opérations d'union [DP88].

2.5.6 Evaluation des domaines de mise en correspondance

Dans ce paragraphe, nous présentons un certain nombre de résultats sur une configuration quelconque du capteur stéréoscopique.

Soit la paire d'images stéréoscopiques représentée sur la figure 2.19. Nous avons choisi d'évaluer le domaine de mise en correspondance du pixel \mathcal{M}_1 tel que défini sur l'image 1.

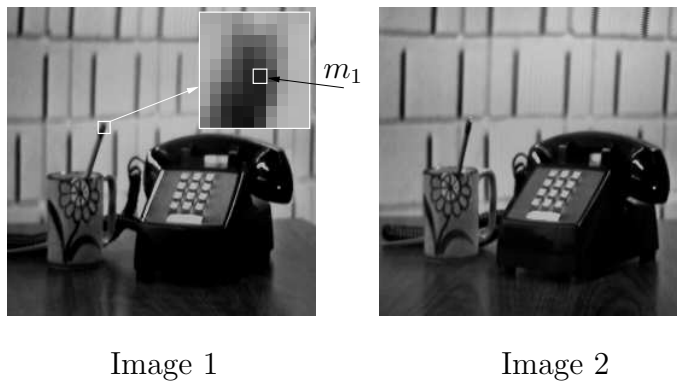
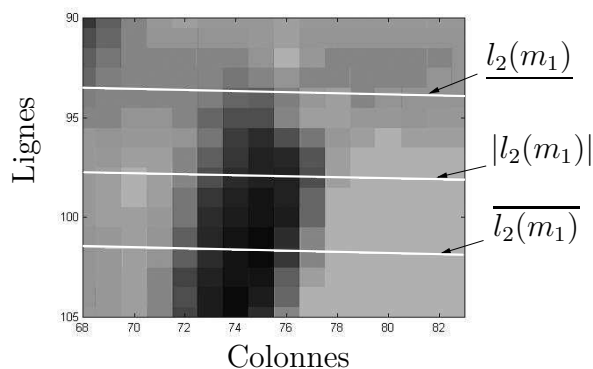


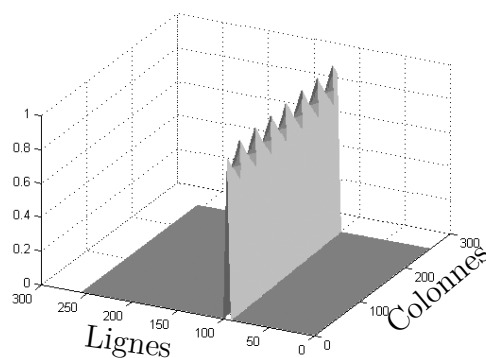
Figure 2.19 – Couple d'images stéréoscopique.

Sur la figure 2.20a sont représentées les trois lignes épipolaires définissant le bandeau épipolaire 2.20b. On peut constater sur la figure 2.20a que les lignes épipolaires étant définies dans l'espace continu ne passent pas par le centre de chaque pixel, ceci explique le phénomène de dentelure au sommet du bandeau épipolaire.

Nous avons calculé les distributions de possibilités de compatibilité chromatique et de non compatibilité chromatique, distributions représentées sur la figure 2.21. Les deux images correspondent aux valeurs prise par $\Pi_C(\mathcal{M}_1, \mathcal{M}_2)$ et $\overline{\Pi_C(\mathcal{M}_1, \mathcal{M}_2)}$, le



(a)



(b)

Figure 2.20 – Lignes épipolaires Inf, Moy et Sup (a), bandeau épipolaire (b),

blanc correspondant à $\Pi = 1$ et le noir à $\Pi = 0$. Le pixel \mathcal{M}_1 est plutôt noir, ainsi les images représentant les compatibilités chromatiques apparaissent en niveau inversé.

Sur la figure 2.22 sont représentées les distributions de possibilité de mise en correspondance $\Pi(\mathcal{M}_1, \mathcal{M}_2)$ et de nécessité $\mathcal{N}(\mathcal{M}_1, \mathcal{M}_2)$ (pour rappel $\mathcal{N}(\mathcal{M}_1, \mathcal{M}_2) = 1 - \Pi(\mathcal{M}_1, \mathcal{M}_2)$). Le pixel \mathcal{M}_2 est défini comme étant un des pixels de l'image 2.

Sur la figure 2.23 est représenté une "coupe" selon une ligne de l'image des distributions de possibilité et de nécessité de mise en correspondance. On constate un pic sur les deux distributions localisant la zone la plus compatible avec le pixel \mathcal{M}_1 .

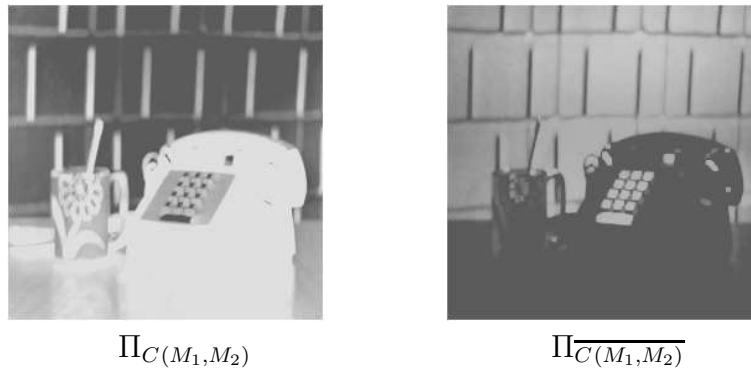


Figure 2.21 – Possibilité et impossibilité chromatique.

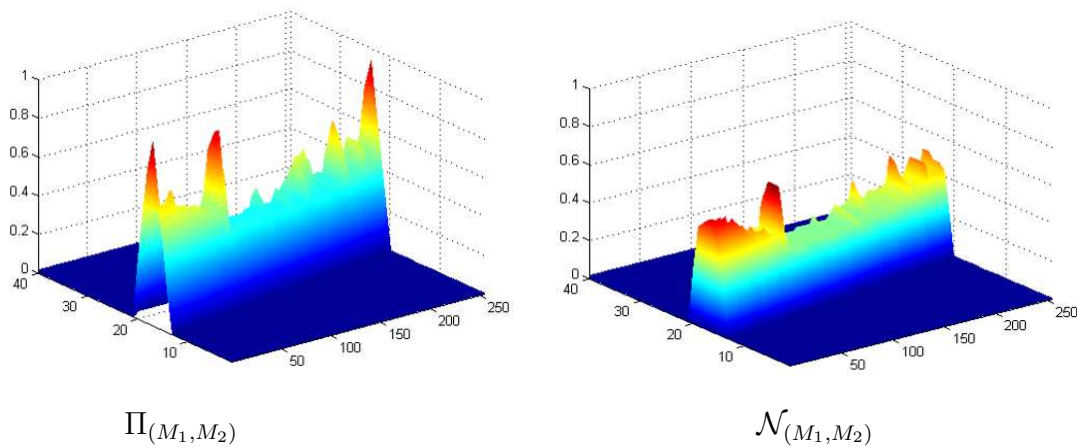


Figure 2.22 – Distributions de possibilité et de nécessité.

2.6 Conclusion

Dans ce chapitre nous avons proposé un nouveau modèle de représentation du pixel dont les deux objectifs sont de prendre en compte l'imprécision due à l'échantillonnage ainsi que l'imprécision de la valeur de la luminance associée au pixel. Nous avons présenté comment se servir de ce modèle pour établir des mises en correspondance de pixels dans deux images stéréoscopiques. Ce modèle, ainsi que les outils développés, s'appuie sur la théorie des sous-ensembles flous ainsi que sur les processus bipolaires.

Dans le premier paragraphe, nous avons étudié la dégradation de l'hypothèse de conservation de la distribution de luminance par l'échantillonnage. Nous sommes arrivés à la conclusion que deux pixels, projections échantillonnées d'un même point tridimensionnel, avaient des valeurs de luminance différentes indépendamment de tout bruit. La représentation du phénomène d'échantillonnage est alors apparue indispensable dans le cadre de la mise en correspondance. Dans notre démarche, nous souhaitons que la modélisation des hypothèses de travail ajoute un apriorisme minimal sur la procédure d'appariement tout en restant le plus général possible pour que les traitements effec-

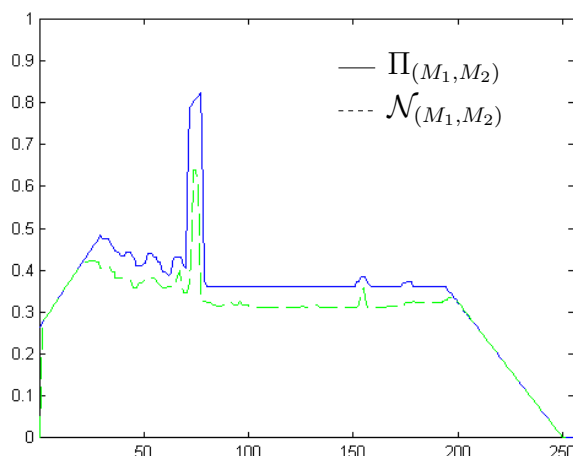


Figure 2.23 – Coupe des distributions de possibilité et de nécessité.

tués puissent s'abstraire de la technologie d'obtention des images. Nous avons défini que le modèle minimal représentant le phénomène d'échantillonnage devait modéliser aussi bien le support minimal représenté par un pixel ponctuel que le support maximal représenté par les huit voisins du pixel considéré.

Dans le deuxième paragraphe, nous avons dressé un panorama des méthodes qui sont généralement utilisées pour représenter la dégradation de l'information due à l'échantillonnage.

La première modélisation utilise des noyaux sommatifs aussi bien pour représenter l'échantillonnage que pour estimer la valeur de l'intensité d'illumination en dehors des valeurs échantillonnées. Les noyaux les plus généraux sont les noyaux gaussiens et les noyaux uniformes. Le principal défaut du noyau gaussien est que la loi de probabilité associée n'est pas à support borné. L'inconvénient du noyau uniforme est qu'il n'a pas de mode et qu'il génère des instabilités de représentation lorsque la zone d'intégration franchit une frontière entre deux pixels. De façon générale, l'utilisation des distributions de probabilités reste complexe du fait des formes mathématiques des lois associées. Enfin, la propagation des incertitudes modélisées par distributions de probabilités n'est pas des plus simples.

La deuxième modélisation utilise un intervalle (ou plutôt un pavé) pour représenter notre méconnaissance du support d'intégration. Le principal inconvénient est son aspect non nuancé ainsi que le choix difficile d'une statistique pour représenter de la façon la plus précise l'information liée à l'interpolation.

La troisième représentation permet de faire le lien entre la modélisation probabiliste et la modélisation par intervalle. Les intervalles de confiance permettent de donner un ensemble d'intervalles de niveaux de confiance différents associés à une loi de probabilité. La construction des intervalles de confiance centrés sur le mode permet d'obtenir les intervalles de confiance les plus spécifiques. L'intérêt est donc de travailler sur l'ensemble de ces intervalles. C'est ce que permet la modélisation floue.

La distribution de possibilité optimale telle que définie par Dubois et Prade permet de représenter l'ensemble des intervalles de confiance emboîtés associés à une loi de probabilité affectés de leur niveau de confiance. Par contre, l'utilisation de la possibilité optimale nécessite le choix d'une loi de probabilité et des calculs importants pour être construite. La distribution de possibilité triangulaire permet de majorer toutes les distributions de possibilité optimales ayant le même mode et le même support. Nous avons donc décidé d'associer, à chaque pixel, un ensemble flou pyramidal centré sur les coordonnées du pixel ponctuel et de support inclus dans le pavé défini par les 8 voisins de chaque pixel. Cette modélisation garantit de contenir tous les intervalles de confiance les plus spécifiques construits à partir d'une distribution de probabilité ayant le même mode et le même support.

Dans le troisième paragraphe, nous avons proposé une modélisation de l'information de luminance visant à être moins sensible aux bruits ainsi qu'à prendre en compte le phénomène de quantification et de saturation. Cette approche se base sur une modélisation des phénomènes de saturation de l'information de luminance en deux classes floues duales. La prise en compte de la quantification est représentée par des mesures de possibilité et de nécessité. Nous avons aussi montré comment le modèle de pixel flou est inclus dans la détermination de ces distributions.

Dans le quatrième paragraphe, nous avons montré comment utiliser notre modèle de pixel flou dans le calcul d'une mise en correspondance entre deux pixels dans le cadre d'images stéréoscopiques. Puisque la bijection entre deux pixels de deux images ne peut être obtenue, nous avons associé à chaque pixel de chaque image un domaine flou de mise en correspondance. Ces domaines sont construits à partir des compatibilités chromatiques entre les pixels mais aussi à partir du respect de la contrainte épipolaire floue. Nous avons illustré notre démarche en fin de ce paragraphe par un exemple de mise en correspondance.

Notre objectif dans ce chapitre était d'élaborer un modèle de pixel pouvant représenter de façon simple aussi bien le phénomène d'échantillonnage de l'information de luminance que celui de la quantification et de saturation en étant le moins possible sensible aux bruits secondaires. Il reste cependant un certain nombre d'améliorations à apporter. Tout d'abord, il serait nécessaire de reprendre les travaux de Dubois, Mauris et al. [DFMP04] afin d'étendre la propriété du plus petit intervalle le plus spécifique au pavé 2D. Enfin, il serait intéressant de quantifier l'apport de la prise en compte des informations chromatiques de voisinage dans notre pixel flou (pour toutes les valeurs de $\alpha \in [0, 0.5]$) sur les distributions de possibilité de mise en correspondance.

Dans le chapitre suivant, nous proposons d'appliquer ce modèle à la mise en correspondance stéréoscopique dense. C'est à dire que nous allons utiliser les contraintes géométriques d'appariements pour essayer de diminuer les domaines de mise en correspondance. Nous proposons de baser notre processus sur un algorithme itératif et coopératif.

Chapitre 3

Mise en correspondance dense de pixels dans une paire d'images stéréoscopique

Dans ce chapitre nous proposons d'utiliser le modèle de pixel flou que nous avons présenté pour réaliser une mise en correspondance dense pour un couple d'images stéréoscopiques calibrées. Notre démarche consiste à utiliser une version floue des contraintes géométriques imposées par le calibrage pour réduire le domaine de correspondance associé à chaque pixel. Nous proposons d'augmenter la spécificité des domaines par un processus coopératif et itératif.

Le premier paragraphe de ce chapitre présente le principe général de notre méthode. Le deuxième paragraphe présente la modélisation des contraintes de mise en correspondance stéréoscopiques qui vont être utilisées pour restreindre les domaines d'appariements possibles. Le troisième paragraphe présente le choix final d'appariement dans le cas restreint des images redressées. Finalement, une conclusion fait une synthèse sur les apports et les limites de cette nouvelle approche. Nous proposons un certain nombre de perspectives pour en améliorer le principe.

3.1 Principe de base

Les appariements chromatiques et géométriques définis dans le chapitre précédent nous ont permis de définir, pour chaque pixel \mathcal{M}_1 de l'image 1 (resp. chaque pixel \mathcal{M}_2 de l'image 2) un domaine flou $\mathcal{D}_2(\mathcal{M}_1)$ (resp. $\mathcal{D}_1(\mathcal{M}_2)$) des pixels de l'image 2 (resp. de l'image 1) pouvant lui correspondre. Ces domaines sont définis sous la forme d'une distribution de possibilité d'appariement $\Pi(\mathcal{M}_1, \mathcal{M}_2)$. Cette distribution de possibilité ne prend pas en considération des contraintes annexes telles que l'unicité, l'ordre, ... Nous proposons de la modifier pour obtenir une nouvelle distribution respectant mieux ces contraintes. A l'instar des méthodes proposées par Zitnick et Kanade [ZK00], Tomasi

et Manduchi [TM96] et Gouet et Montesinos [GMP98], nous utilisons les points dont les mesures de confiance d'appariement sont plus élevées pour lever les ambiguïtés de mise en correspondance sur les points dont les mesures de confiance sont plus faibles. L'aspect itératif nous est donné par l'utilisation des α -coupes.

Pour un niveau de coupe α donné, la distribution de possibilité $\Pi(\mathcal{M}_1, \mathcal{M}_2)$ permet de définir, pour chaque pixel \mathcal{M}_1 de l'image 1 (resp. pour chaque pixel \mathcal{M}_2 de l'image 2) un domaine binaire $\mathcal{D}_2^\alpha(\mathcal{M}_1)$ de l'image 2 (resp. $\mathcal{D}_1^\alpha(\mathcal{M}_2)$ de l'image 1) par :

$$\begin{cases} \mathcal{D}_2^\alpha(\mathcal{M}_1) = \{\mathcal{M}_2 \in I_2 / \Pi(\mathcal{M}_1, \mathcal{M}_2) \geq \alpha\} \\ \mathcal{D}_1^\alpha(\mathcal{M}_2) = \{\mathcal{M}_1 \in I_1 / \Pi(\mathcal{M}_1, \mathcal{M}_2) \geq \alpha\} \end{cases}$$

La démarche générale du processus itératif que nous proposons consiste à scruter ces domaines de mise en correspondance pour l'ensemble des valeurs possibles de α et de modifier, pour chaque valeur de α , les valeurs des appariements pour des niveaux de coupe inférieure. A l'initialisation du processus de mise en correspondance, $\alpha = 1$ et décroît jusqu'à 0.

De façon à utiliser au mieux la particularité de la représentation possibiliste des appariements, le processus de modification de la distribution de possibilité est bipolaire en ce sens qu'une distribution de mise en correspondance permet de modifier des impossibilités de mise en correspondance et vice-versa. Cette approche bipolaire permet de maintenir la cohérence du modèle et d'en assurer la robustesse.

3.2 Modélisation des contraintes stéréoscopiques

Nous proposons d'utiliser les trois contraintes de mise en correspondance stéréoscopiques classiques que sont les contraintes d'unicité, d'ordre et de point de visée.

Les modification des mesures de possibilité que nous proposons sont basées sur le modèle "si alors" proposé par Dubois et Prade dans [DP88]. On peut trouver des variantes et des utilisations de ce modèle dans [DP96a], [DP96b], [BDKP01], [BDKP02], [DD03].

Nous utilisons les distributions de possibilité de mise en correspondance $\Pi(\mathcal{M}_1, \mathcal{M}_2)$ pour établir de nouvelles distributions de nécessité de mise en correspondance notées $\tilde{\mathcal{N}}(\mathcal{M}_1, \mathcal{M}_2)$ et donc de nouvelles distributions de possibilité de non mise en correspondance notées $\overline{\Pi}(\mathcal{M}_1, \mathcal{M}_2)$. De la même façon, nous utilisons les distributions de nécessité de mise en correspondance $\mathcal{N}(\mathcal{M}_1, \mathcal{M}_2)$ pour établir de nouvelles distributions de possibilité de mise en correspondance $\tilde{\Pi}(\mathcal{M}_1, \mathcal{M}_2)$.

3.2.1 Modélisation de la contrainte d'unicité

Si on se limite à observer des objets opaques dont la disparité n'est pas très forte alors

Définition 3 *un objet dont la projection est un pixel sur la première image a une projection qui est aussi un pixel sur la seconde image.*

L'expression d'une telle règle permet de réduire le domaine de mise en correspondance de chaque pixel de chaque image en prenant en compte les domaines de mise en correspondance des autres pixels.

Soit $\mathcal{D}_2(\mathcal{M}_1)$, le domaine de la seconde image mis en correspondance a priori avec un pixel \mathcal{M}_1 de la première image. Une distribution qui viole la contrainte d'unicité est une distribution telle qu'il existe un pixel \mathcal{M}'_1 dont le domaine $\mathcal{D}_2(\mathcal{M}'_1)$ vérifie (Figure 3.1) :

$$\mathcal{D}_2(\mathcal{M}_1) \cap \mathcal{D}_2(\mathcal{M}'_1) \neq \phi$$

Si on considère des domaines binaires, alors cette situation est indécidable. En effet, on ne peut effectuer le choix d'apparier les pixels contenus dans l'intersection des deux domaines à l'un ou l'autre des pixels de l'image 1. Par contre, l'utilisation des α -coupes sur les mesures de possibilité et de nécessité de mise en correspondance nous permettent de lever certaines ambiguïtés d'appariement.

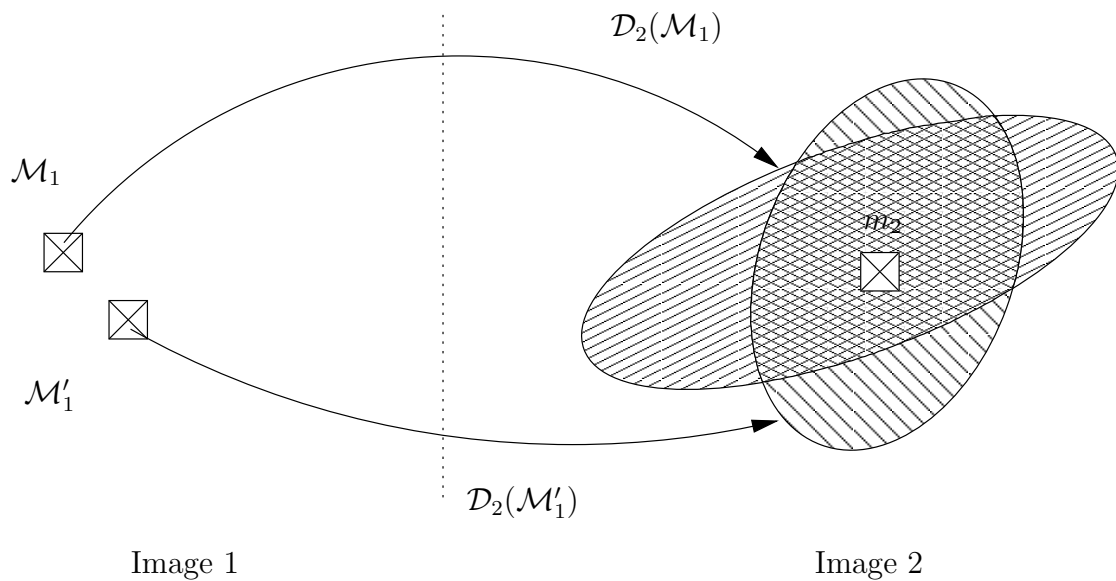


Figure 3.1 – Violation de la contrainte d'unicité.

3.2.1.1 Violation de la contrainte d'unicité - modification à partir des possibilités de mise en correspondance

Soit $\Pi(\mathcal{M}_1, \mathcal{M}_2)$ la mesure de possibilité qui modélise que "il est possible que les pixels \mathcal{M}_1 et \mathcal{M}_2 soient correspondants". Si il existe un pixel $\mathcal{M}'_1 \neq \mathcal{M}_1$ dans l'image 1 tel que :

- $\mathcal{D}_2^0(\mathcal{M}_1) \cap \mathcal{D}_2^0(\mathcal{M}'_1) \neq \phi$
- $\mathcal{M}_2 \subseteq \mathcal{D}_2^\alpha(\mathcal{M}_1)$ avec $\alpha \neq 0$
- $\mathcal{M}_2 \not\subseteq \mathcal{D}_2^\alpha(\mathcal{M}'_1)$

alors, dans ce cas, \mathcal{M}_2 et \mathcal{M}_1 sont possiblement correspondants au niveau α et \mathcal{M}_2 et \mathcal{M}'_1 ne le sont pas.

Soit un univers de discours associé à la contrainte d'unicité avec comme prédicats $\{A, B\} = \{\text{"correspondants"}, \text{"non-correspondants"}\}$. Soit $x = \text{"}\mathcal{M}_1 \text{ et } \mathcal{M}_2\text{"}$ et $y = \text{"}\mathcal{M}'_1 \text{ et } \mathcal{M}_2\text{"}$. On peut énoncer que :

"si x est A alors y est B "

c'est à dire :

"si \mathcal{M}_1 et \mathcal{M}_2 sont correspondants alors \mathcal{M}'_1 et \mathcal{M}_2 sont non-correspondants"

Du fait de l'aspect nuancé de notre processus nous devons reformuler cette règle sous la forme :

"(SI il est possible que \mathcal{M}_1 et \mathcal{M}_2 soient correspondants au niveau α) ET
(SI il n'est pas possible que \mathcal{M}'_1 et \mathcal{M}_2 soient correspondants au niveau α)
ALORS (il n'est pas nécessaire que \mathcal{M}'_1 et \mathcal{M}_2 soient correspondants au niveau $1 - \alpha$)"

Nous utilisons cette règle pour diminuer la nécessité de mettre en correspondance \mathcal{M}'_1 avec \mathcal{M}_2 du point de vue de la contrainte d'unicité sachant \mathcal{M}_1 par :

$$\tilde{\mathcal{N}}(\mathcal{M}'_1, \mathcal{M}_2)_{|\mathcal{M}_1} = \underset{\mathcal{M}_1 \neq \mathcal{M}'_1}{Sup} \{ \min(\mathcal{N}(\mathcal{M}'_1, \mathcal{M}_2), 1 - \alpha) \}$$

ou encore pour augmenter la possibilité de non-mise en correspondance de \mathcal{M}'_1 avec \mathcal{M}_2 du point de vue de la contrainte d'unicité sachant \mathcal{M}_1 par :

$$\tilde{\Pi}(\overline{\mathcal{M}'_1}, \overline{\mathcal{M}_2})_{|\mathcal{M}_1} = \underset{\mathcal{M}_1 \neq \mathcal{M}'_1}{Inf} \left\{ \max(\Pi(\overline{\mathcal{M}'_1}, \overline{\mathcal{M}_2}), \alpha) \right\} \quad (3.1)$$

Nous transférons la confiance que nous avons dans la possibilité de mettre en correspondance \mathcal{M}_1 et \mathcal{M}_2 sur la possibilité de non mise en correspondance de \mathcal{M}'_1 et \mathcal{M}_2 .

De la même façon, soit $\mathcal{D}_1(\mathcal{M}_2)$, le domaine de la première image mis en correspondance a priori avec un pixel \mathcal{M}_2 de la deuxième image. Une distribution viole la contrainte d'unicité si il existe un pixel \mathcal{M}'_2 dont le domaine $\mathcal{D}_1(\mathcal{M}'_2)$ vérifie :

$$\mathcal{D}_1(\mathcal{M}_2) \cap \mathcal{D}_1(\mathcal{M}'_2) \neq \phi$$

Dans ce cas, nous modifions la nécessité de mettre en correspondance \mathcal{M}_1 avec \mathcal{M}'_2 du point de vue de la contrainte d'unicité sachant \mathcal{M}_2 par :

$$\tilde{\mathcal{N}}(\mathcal{M}_1, \mathcal{M}'_2)_{|\mathcal{M}_2} = \underset{\mathcal{M}_2 \neq \mathcal{M}'_2}{Sup} \{ \min(\mathcal{N}(\mathcal{M}_1, \mathcal{M}'_2), 1 - \alpha) \} \quad (3.2)$$

ou encore :

$$\tilde{\Pi}(\overline{\mathcal{M}_1, \mathcal{M}'_2})_{|\mathcal{M}_2} = \underset{\mathcal{M}_2 \neq \mathcal{M}'_2}{Inf} \left\{ \max(\Pi(\overline{\mathcal{M}_1, \mathcal{M}'_2}), \alpha) \right\} \quad (3.3)$$

3.2.1.2 Violation de la contrainte d'unicité - modification à partir des nécessités de mise en correspondance

Pour un niveau de coupe donné α , la distribution de nécessité $\mathcal{N}(\mathcal{M}_1, \mathcal{M}_2)$ permet de définir, pour chaque pixel \mathcal{M}_1 de l'image 1 (resp. pour chaque pixel \mathcal{M}_2 de l'image 2) un autre domaine binaire $\mathcal{D}_2^\alpha(\mathcal{M}_1)$ de l'image 2 (resp. $\mathcal{D}_1^\alpha(\mathcal{M}_2)$ de l'image 1) par :

$$\begin{cases} \mathcal{D}_2^\alpha(\mathcal{M}_1) = \{\mathcal{M}_2 \in I_2 / \mathcal{N}(\mathcal{M}_1, \mathcal{M}_2) \geq \alpha\} \\ \mathcal{D}_1^\alpha(\mathcal{M}_2) = \{\mathcal{M}_1 \in I_1 / \mathcal{N}(\mathcal{M}_1, \mathcal{M}_2) \geq \alpha\} \end{cases}$$

Soit $\mathcal{N}(\mathcal{M}_1, \mathcal{M}_2)$ la mesure de nécessité qui modélise que "les pixels \mathcal{M}_1 et \mathcal{M}_2 sont nécessairement correspondants" avec $\mathcal{N}(\mathcal{M}_1, \mathcal{M}_2) = 1 - \Pi(\overline{\mathcal{M}_1, \mathcal{M}_2})$. S'il existe un pixel $\mathcal{M}'_1 \neq \mathcal{M}_1$ dans l'image 1 tel que :

- $\mathcal{D}_2^0(\mathcal{M}_1) \cap \mathcal{D}_2^0(\mathcal{M}'_1) \neq \phi$
- $\mathcal{M}_2 \subseteq \mathcal{D}_2^\alpha(\mathcal{M}_1)$ avec $\alpha \neq 0$
- $\mathcal{M}_2 \not\subseteq \mathcal{D}_2^\alpha(\mathcal{M}'_1)$

alors, dans ce cas, \mathcal{M}_2 et \mathcal{M}_1 sont nécessairement correspondants au niveau α et \mathcal{M}_2 et \mathcal{M}'_1 ne le sont pas.

Nous pouvons établir la règle de la contrainte d'unicité sous la forme :

"(SI il est nécessaire que \mathcal{M}_1 et \mathcal{M}_2 soient correspondants au niveau α) ET
(SI il n'est pas nécessaire que \mathcal{M}'_1 et \mathcal{M}_2 soient correspondants au niveau α)
ALORS (il n'est pas possible que \mathcal{M}'_1 et \mathcal{M}_2 soient correspondants au niveau $1 - \alpha$)"

Nous utilisons cette règle pour diminuer la possibilité de mettre en correspondance \mathcal{M}'_1 avec \mathcal{M}_2 du point de vue de la contrainte d'unicité sachant \mathcal{M}_1 en écrivant :

$$\tilde{\Pi}(\mathcal{M}'_1, \mathcal{M}_2)_{|\mathcal{M}_1} = \underset{\mathcal{M}_1 \neq \mathcal{M}'_1}{Sup} \left\{ \min(\Pi(\mathcal{M}'_1, \mathcal{M}_2), 1 - \alpha) \right\}$$

Pour l'autre image, la modification de la distribution de possibilité est exprimée par :

$$\tilde{\Pi}(\mathcal{M}_1, \mathcal{M}'_2)_{|\mathcal{M}_2} = \underset{\mathcal{M}_2 \neq \mathcal{M}'_2}{Sup} \left\{ \min(\Pi(\mathcal{M}_1, \mathcal{M}'_2), 1 - \alpha) \right\}$$

3.2.1.3 Respect de la contrainte d'unicité

Par cette formulation de la contrainte d'unicité, on voit qu'il peut exister des appariements tels que :

$$\mathcal{D}_1^\alpha(\mathcal{M}_2) \cap \mathcal{D}_1^\alpha(\mathcal{M}'_2) \neq \phi$$

ou :

$$\mathcal{D}_2^\alpha(\mathcal{M}_1) \cap \mathcal{D}_2^\alpha(\mathcal{M}'_1) \neq \phi$$

Cette propriété est intéressante dans le cadre d'une modélisation prudente des mises en correspondance. En effet, nous avons souligné dans le paragraphe 1.3.1.2 qu'un même objet tridimensionnel pouvait avoir des projections de taille différente sur les deux images stéréoscopiques. L'utilisation classique de la contrainte d'unicité conduirait à exclure cette éventualité. Notre méthode permet de ne pas prendre en compte la contrainte d'unicité lorsque la situation est indécidable. Un même point de l'image 1 peut donc avoir plusieurs correspondants dans l'image 2 tout en respectant la contrainte d'unicité.

3.2.2 Modélisation de la contrainte d'ordre

Pour simplifier l'explication de cette modélisation, nous proposons de nous abstraire, dans un premier temps, du phénomène d'échantillonnage. Si m_1 et m_2 sont correspondants, alors tous les pixels appartenant à $l_2(m_1)$ ont leurs correspondants qui appartiennent à $l_1(m_2)$ [Fau93].

Démonstration :

soit la contrainte épipolaire entre m_1 et m_2 :

$$m_2^t F m_1 = 0 \tag{3.4}$$

Les lignes épipolaires s'écrivent alors :

$$\begin{cases} l_2(m_1) = F m_1 \\ l_1(m_2) = F^t m_2 \end{cases} \tag{3.5}$$

Soit m'_1 un pixel de l'image 1 appartenant à $l_1(m_2)$ c'est à dire :

$$l_1(m_2) m'_1 = 0$$

qui peut s'écrire aussi :

$$m_2^t F m'_1 = 0 \tag{3.6}$$

Par comparaison entre les équations (3.4) et (3.6) on peut écrire l'égalité suivante :

$$F m'_1 = F m_1$$

ou encore :

$$l_2(m'_1) = l_2(m_1)$$

Prenons maintenant en compte l'échantillonnage et l'imprécision des paramètres de la matrice fondamentale. On peut dire que si \mathcal{M}_1 et \mathcal{M}_2 sont considérés comme correspondants, tous les pixels appartenant à $\mathcal{L}_2(\mathcal{M}_1)$ (le bandeau épipolaire de l'image 2 des correspondants potentiels du pixel flou \mathcal{M}_1), ont leur correspondants qui appartiennent à $\mathcal{L}_1(\mathcal{M}_2)$.

3.2.2.1 Principe

Sous certaines conditions, qui ont été présentées dans le chapitre 1, l'ordre relatif des projections le long des deux lignes épipolaires duales est le même dans les deux images. Nous pouvons alors établir la proposition suivante :

Proposition 2

si \mathcal{M}_1 et \mathcal{M}_2 sont considérés comme projections d'un même point alors tous les pixels à droite (resp. gauche) de \mathcal{M}_1 le long de son bandeau épipolaire sont à droite (resp. gauche) de \mathcal{M}_2 .

C'est à dire que, pour n'importe quel pixel \mathcal{M}'_1 de l'image 1 choisi, par exemple, à droite du pixel \mathcal{M}_1 le long de $\mathcal{L}_1(\mathcal{M}_2)$, une partie de son domaine de mise en correspondance $\mathcal{D}_2(\mathcal{M}'_1)$ respecte la contrainte d'ordre, partie représentée hachurée sur la figure 3.2.

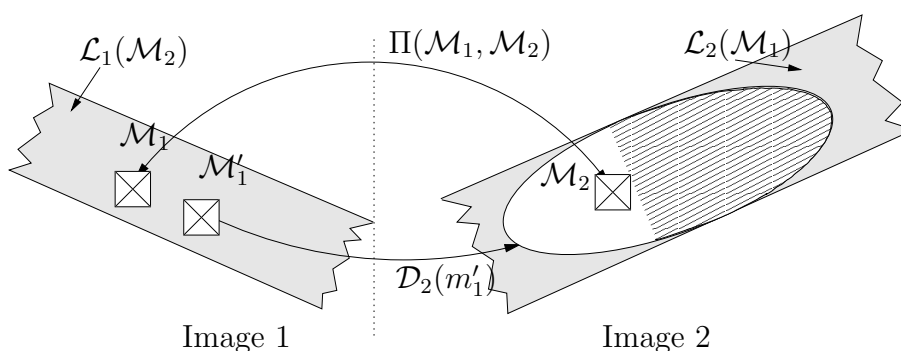


Figure 3.2 – Respect de la contrainte d'ordre.

Le même raisonnement peut être tenu à la fois pour tout point à gauche de \mathcal{M}_1 mais aussi pour n'importe quel pixel de l'image 2 le long du bandeau épipolaire $\mathcal{L}_2(\mathcal{M}_1)$.

Par contre, il est nécessaire de définir précisément les relations spatiales que sont "à droite de" et "à gauche de".

3.2.2.2 Relations spatiales

- Cas des images redressées

Dans le cas des images redressées, les lignes épipolaires sont parfaitement horizontales et passent par le centre de chaque pixel (figure 3.3a). Un pixel \mathcal{M}'_1 dont le noyau a comme coordonnées (x'_i, y'_j) est à droite (resp. gauche) du pixel \mathcal{M}_1 dont le noyau a comme coordonnées (x_i, y_j) s'il appartient au domaine $D^D_{\mathcal{M}_1, \mathcal{L}_1(\mathcal{M}_2)}$ (resp. $D^G_{\mathcal{M}_1, \mathcal{L}_1(\mathcal{M}_2)}$) représentant le domaine des pixels étant "à gauche" (resp. "à droite") du pixel \mathcal{M}_1 . Le domaine "à droite" (resp. "à gauche") est défini par le bandeau épipolaire et l'ensemble des pixels dont les noyaux ont pour coordonnées (x'_i, y'_j) avec $y'_j > y_j + 1$ (resp. $y'_j < y_j - 1$).

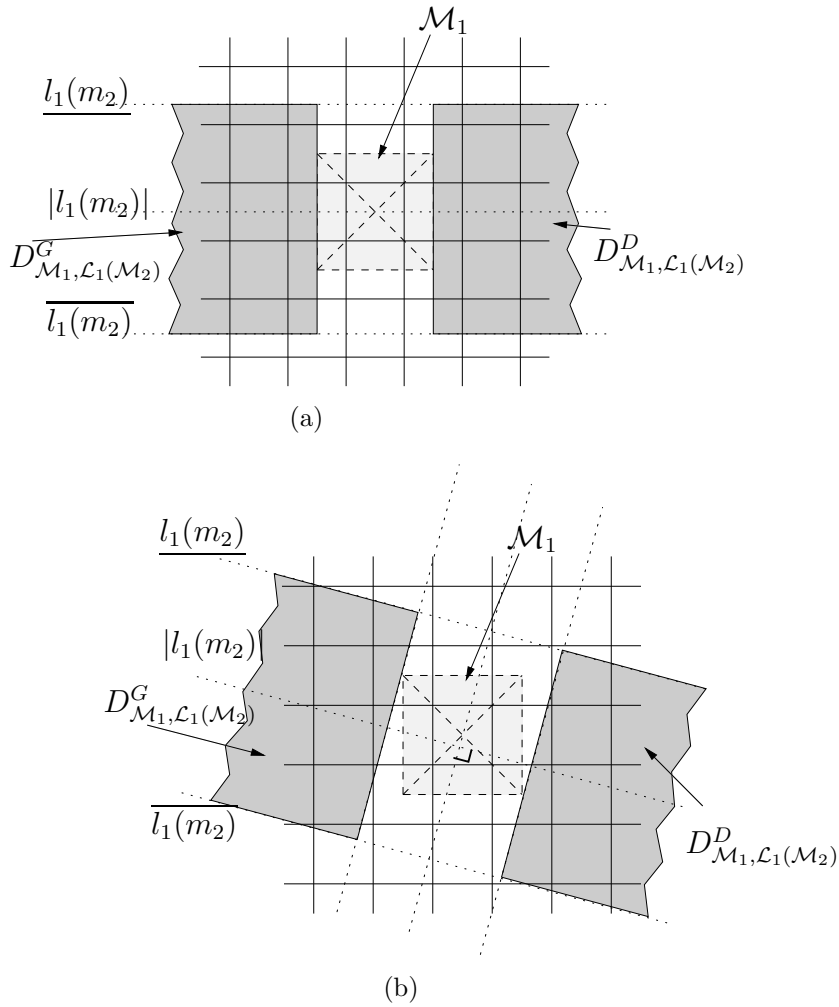


Figure 3.3 – Définition des relations spatiales dans le cas des images redressées (a), dans le cas général (b).

– Cas général

Le cas général est plus difficile à définir car les lignes épipolaires ne passent pas par les centres de tous les pixels (figure 3.3b). En s'inspirant des travaux de Bloch [Blo99], [BS02] portant sur la définition floue des relations spatiales, nous avons défini, en choisissant comme élément structurant le pixel flou, les deux ensembles binaires

droite et gauche de la façon suivante :

si le pixel \mathcal{M}_1 de l'image 1 est considéré comme correspondant du pixel \mathcal{M}_2 de l'image 2, alors \mathcal{M}_1 appartient au bandeau épipolaire $\mathcal{L}_1(\mathcal{M}_2)$. On calcule la droite qui passe par le noyau de \mathcal{M}_1 , c'est à dire le pixel m_1 , et qui est perpendiculaire à $l_1(m_2)$. Si la pente de $l_1(m_2)$ est positive, on trace deux droites parallèles à la droite perpendiculaire passant par le coin inférieur gauche et supérieur droit du pixel flou \mathcal{M}_1 , si la pente est négative (ce qui est le cas de la figure 3.3b), les deux droites passent par le coin supérieur gauche et inférieur droit du pixel flou. Ces deux droites permettent de séparer le bandeau épipolaire en trois ensembles binaires :

- la partie "à gauche" $D_{\mathcal{M}_1, \mathcal{L}_1(\mathcal{M}_2)}^G$,
- la partie "à droite" $D_{\mathcal{M}_1, \mathcal{L}_1(\mathcal{M}_2)}^D$,
- et la partie centrale.

Un pixel \mathcal{M}'_1 de l'image 1 est défini "à gauche" (resp. "à droite") de \mathcal{M}_1 si $\mathcal{M}'_1 \subset D_{\mathcal{M}_1, l_1(m_2)}^G$ (resp. $\mathcal{M}'_1 \subset D_{\mathcal{M}_1, l_1(m_2)}^D$).

3.2.2.3 Proposition duale de la contrainte d'ordre

Nous pouvons exprimer la proposition négative duale de la contrainte d'ordre :

Proposition 3

si \mathcal{M}_1 et \mathcal{M}_2 sont considérés comme projections d'un même point alors tous les pixels à droite (resp. gauche) de \mathcal{M}_1 ne peuvent pas être à gauche (resp. droite) de \mathcal{M}_2 .

Pour expliquer cette démarche, prenons le cas d'un pixel \mathcal{M}_2 de l'image 2 (Figure 3.4). On peut associer à \mathcal{M}_2 le domaine épipolaire $\mathcal{L}_1(\mathcal{M}_2)$. Soient \mathcal{M}_1 et \mathcal{M}'_1 deux pixels de $\mathcal{L}_1(\mathcal{M}_2)$ avec \mathcal{M}'_1 à droite de \mathcal{M}_1 et $\Pi(\mathcal{M}_1, \mathcal{M}_2) > \Pi(\mathcal{M}'_1, \mathcal{M}_2)$. La contrainte d'ordre permet de modifier une partie du domaine de mise en correspondance de \mathcal{M}'_1 ($\mathcal{D}_2(\mathcal{M}'_1)$) en remarquant que tout point à gauche de \mathcal{M}_2 ne peut être mis en correspondance avec \mathcal{M}'_1 si \mathcal{M}_1 est en correspondance avec \mathcal{M}_2 , et ce avec un niveau de coupe minimal de $(1 - \Pi(\mathcal{M}_1, \mathcal{M}_2))$.

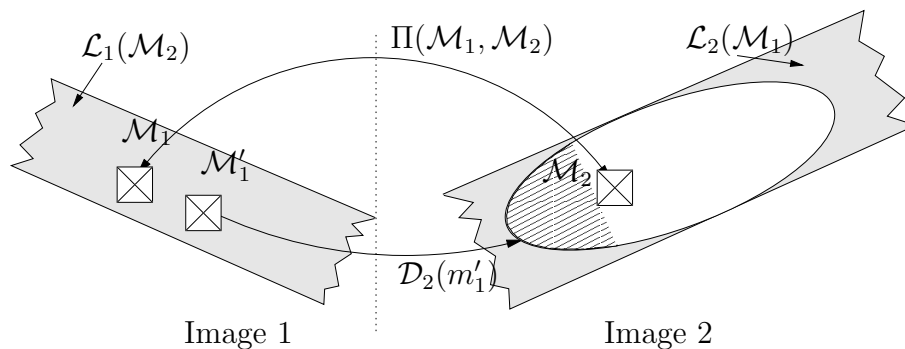


Figure 3.4 – Violation de la contrainte d'ordre.

Le même raisonnement peut être tenu pour tout point à gauche de \mathcal{M}_1 mais aussi pour n'importe quel pixel de l'image 2.

3.2.2.4 Violation de la contrainte d'ordre - modification à partir des possibilités de mise en correspondance

L'appariement de \mathcal{M}'_1 avec \mathcal{M}'_2 viole la contrainte d'ordre au niveau α si :

$$\begin{aligned} & [(\mathcal{M}'_1 \text{ est "à droite" de } \mathcal{M}_1) \text{ ET } (\mathcal{M}'_2 \text{ est "à gauche" de } \mathcal{M}_2)] \text{ OU} \\ & [(\mathcal{M}'_1 \text{ est "à gauche" de } \mathcal{M}_1) \text{ ET } (m'_2 \text{ est "à droite" de } \mathcal{M}_2)] \text{ ET} \\ & [(\mathcal{M}'_1 \in \mathcal{L}_1(\mathcal{M}_2)) \text{ ET } (\mathcal{M}'_2 \in \mathcal{L}_2(\mathcal{M}_1))] \text{ ET} \\ & [\mathcal{M}_2 \subseteq \mathcal{D}_2^\alpha(\mathcal{M}_1), \mathcal{M}_1 \subseteq \mathcal{D}_1^\alpha(\mathcal{M}_2)] \text{ ET } [\mathcal{M}'_2 \not\subseteq \mathcal{D}_2^\alpha(\mathcal{M}'_1), \mathcal{M}'_1 \not\subseteq \mathcal{D}_1^\alpha(\mathcal{M}'_2)]. \end{aligned}$$

Dans ce cas, on peut modifier la nécessité de mettre en correspondance \mathcal{M}'_1 avec \mathcal{M}'_2 du point de vue de la contrainte d'ordre, connaissant l'appariement de \mathcal{M}_1 avec \mathcal{M}_2 par :

$$\tilde{\mathcal{N}}(\mathcal{M}'_1, \mathcal{M}'_2)_{|\mathcal{M}_1, \mathcal{M}_2} = \underset{\substack{\mathcal{M}_1 \neq \mathcal{M}'_1 \\ \mathcal{M}_2 \neq \mathcal{M}'_2}}{\text{Sup}} \{ \min(\mathcal{N}(\mathcal{M}'_1, \mathcal{M}'_2), 1 - \alpha) \} \quad (3.7)$$

3.2.2.5 Violation de la contrainte d'ordre - modification à partir des nécessités de mise en correspondance

L'appariement de \mathcal{M}'_1 avec \mathcal{M}'_2 viole la contrainte d'ordre au niveau α si :

$$\begin{aligned} & [(\mathcal{M}'_1 \text{ est "à droite" de } \mathcal{M}_1) \text{ ET } (\mathcal{M}'_2 \text{ est à gauche de } \mathcal{M}_2)] \text{ OU} \\ & [(\mathcal{M}'_1 \text{ est à gauche de } \mathcal{M}_1) \text{ ET } (\mathcal{M}'_2 \text{ est à droite de } \mathcal{M}_2)] \text{ ET} \\ & [(\mathcal{M}'_1 \in \mathcal{L}_1(\mathcal{M}_2)) \text{ ET } (\mathcal{M}'_2 \in \mathcal{L}_2(\mathcal{M}_1))] \text{ ET} \\ & [\mathcal{M}_2 \subseteq \mathcal{D}_2^\alpha(\mathcal{M}_1), \mathcal{M}_1 \subseteq \mathcal{D}_1^\alpha(\mathcal{M}_2)] \text{ ET } [\mathcal{M}'_2 \not\subseteq \mathcal{D}_2^\alpha(\mathcal{M}'_1), \mathcal{M}'_1 \not\subseteq \mathcal{D}_1^\alpha(\mathcal{M}'_2)]. \end{aligned}$$

Dans ce cas, on peut modifier la possibilité de mettre en correspondance \mathcal{M}'_1 avec \mathcal{M}'_2 du point de vue de la contrainte d'ordre, connaissant l'appariement de \mathcal{M}_1 avec \mathcal{M}_2 , par :

$$\tilde{\Pi}(\mathcal{M}'_1, \mathcal{M}'_2)_{|\mathcal{M}_1, \mathcal{M}_2} = \underset{\substack{\mathcal{M}_1 \neq \mathcal{M}'_1 \\ \mathcal{M}_2 \neq \mathcal{M}'_2}}{\text{Sup}} \{ \min(\Pi(\mathcal{M}'_1, \mathcal{M}'_2), \alpha) \} \quad (3.8)$$

Par contre, il faut prendre la précaution de respecter le fait que la possibilité doit être supérieure à la nécessité :

$$\tilde{\Pi}(\mathcal{M}'_1, \mathcal{M}'_2) \geq \mathcal{N}(\mathcal{M}'_1, \mathcal{M}'_2)$$

La diminution de $\tilde{\Pi}(\mathcal{M}'_1, \mathcal{M}'_2)$ est limitée par la borne inférieure $\mathcal{N}(\mathcal{M}'_1, \mathcal{M}'_2)$.

3.2.3 Cas des images redressées - contrainte du point de visée

Dans ce paragraphe, nous proposons de restreindre notre algorithme au cas particulier des images redressées (cf §1.2.6).

Si \mathcal{M}_1 est un pixel de l'image 1 dont le noyau m_1 a comme coordonnées (x_i, y_j) , son correspondant dans l'image 2 \mathcal{M}_2 appartient forcément à un segment $[a_2, \mathcal{M}_{21}]$ du bandeau épipolaire dont la ligne épipolaire moyenne $l_2(m_1)$ a comme coordonnée x'_j . a_2 est l'intersection de ce bandeau avec le bord gauche de l'image et \mathcal{M}_{21} est le pixel dont le noyau m_{21} a comme coordonnées (x'_i, y_j) (figure 3.5). Cette contrainte, que

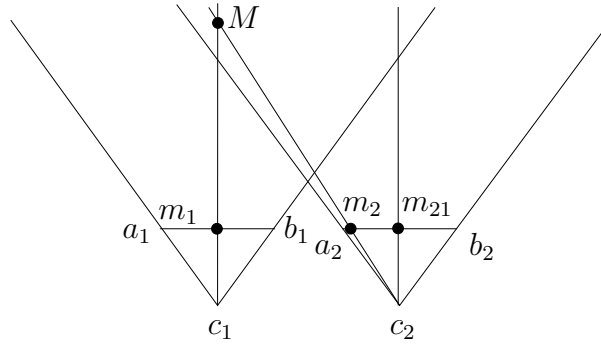


Figure 3.5 – Un cas particulier de stéréovision.

nous avons appelé contrainte du point de visée, nous permet d'établir des mesures de possibilité de mise en correspondance et de possibilité de non mise en correspondance forte par :

$$\forall \mathcal{M}_2 \text{ n'appartenant pas à l'ensemble flou formé par } a_2 \text{ et } \mathcal{M}_{21} \Rightarrow \begin{cases} \Pi(\mathcal{M}_1, \mathcal{M}_2) = 0 \\ \mathcal{N}(\mathcal{M}_1, \mathcal{M}_2) = 0 \end{cases}$$

Ce même raisonnement doit être effectué en considérant cette fois-ci l'image 2, c'est à dire :

$$\forall \mathcal{M}_1 \text{ n'appartenant pas à l'ensemble flou formé par } a_1 \text{ et } \mathcal{M}_{12} \Rightarrow \begin{cases} \Pi(\mathcal{M}_1, \mathcal{M}_2) = 0 \\ \mathcal{N}(\mathcal{M}_1, \mathcal{M}_2) = 0 \end{cases}$$

avec \mathcal{M}_{12} le pixel ayant pour noyau m_{12} le point de coordonnées (x'_i, y_j) dans l'image 1. Cette contrainte géométrique est une contrainte forte puisqu'elle permet, dans le cas des images redressées, de diminuer par deux le nombre total de candidats potentiels.

3.3 Création d'une carte de disparité spécifique, comparaison avec les méthodes classiques

Les cartes d'appariements que nous obtenons ne permettent pas de reconstruire des volumes tridimensionnels au sens classique du terme, c'est à dire un ensemble de

points de l'espace tridimensionnel. Il serait par contre envisageable de projeter en 3D les domaines d'appariements possibilistes pour créer des cartes d'occupation pour la navigation robotique par exemple [ML00], [Ous03].

Afin de pouvoir tester la pertinence des appariements que nous obtenons, il est nécessaire de comparer notre approche aux approches classiques. C'est pourquoi nous proposons d'utiliser des méthodes classiques de programmation dynamique et de coupe de graphe pour construire la carte de disparité dense. L'adaptation de ces techniques au cas des mesures imprécises d'appariements sort du cadre de ce travail. Nous sommes cependant conscients que cette généralisation est nécessaire. Nous proposons de modéliser le degré de vérité de la proposition "le pixel \mathcal{M}_1 de l'image 1 et le pixel \mathcal{M}_2 de l'image 2 sont correspondants". Il est défini à partir des mesures de possibilités de mise en correspondance et de non mise en correspondance modifiées telles que proposé par Dubois et Prade [DP88] par :

$$V(\mathcal{M}_1, \mathcal{M}_2) = \frac{\tilde{\Pi}(\mathcal{M}_1, \mathcal{M}_2) + \tilde{\mathcal{N}}(\mathcal{M}_1, \mathcal{M}_2)}{2}$$

Nous utilisons cette mesure de la même façon qu'un score de mise en correspondance (SAD, SSD, NCC). Du fait du degré de complexité algorithmique que cela entraînerait, cette construction s'applique uniquement sur des images redressées dont on ne garde que les mesures modifiées de mise en correspondance ligne à ligne. De toutes les mises en correspondances que nous venons de calculer, nous ne gardons que les lignes de même ordonnée.

Les algorithmes utilisés sont ceux proposés par Scharstein et Szeliski [SS98] disponibles sur leurs site internet (<http://www.middlebury.edu/stereo>). Les images de test sont des images de référence aussi bien synthétiques que naturelles. Pour chacune de ces images, une carte de disparité de référence est disponible : soit obtenue directement pour les images synthétiques soit obtenue par des méthodes utilisant des sources de lumière structurée [SS03] pour les images naturelles.

3.3.1 Cas des images redressées

Soit la paire d'images stéréoscopiques redressées de la figure 3.6. Nous nous intéressons à la mise en correspondance de l'ensemble des lignes x des deux images.

Nous avons représenté la répartition des mesures de possibilité de mise en correspondance (P.M.E.C) et de possibilité de non mise en correspondance (P.N.M.E.C.) initiales sur la figure 3.7. Sont représentés, sur l'axe des ordonnées, les pixels de la ligne x de l'image 1 et, sur l'axe des abscisses, les pixels de la ligne x de l'image 2. Le domaine de répartition des mesures d'appariement étant $[0, 1]$, nous proposons de les représenter par la couleur blanche pour une mesure égale à 1 et par la couleur noire pour une mesure égale à 0.

Sur la figure 3.8, on peut voir la répartition des possibilités modifiées de mise en correspondance et de non mise en correspondance des pixels de nos deux lignes.

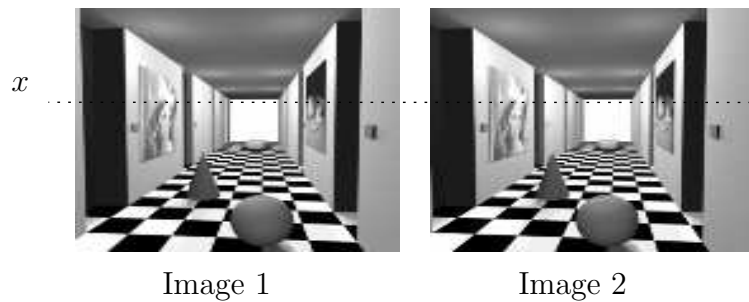


Figure 3.6 – Images stéréoscopiques.

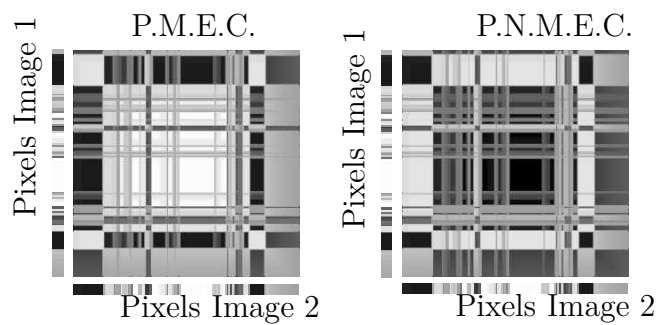


Figure 3.7 – Répartition des possibilités de mise en correspondance (P.M.E.C.) et des possibilités de non mise en correspondance (P.N.M.E.C.) initiales.

Ces distributions sont obtenues après utilisation de notre processus de modification coopératif et itératif. On constate tout d'abord les effets de la contrainte de point de vue. Cette contrainte définit, dans la représentation que nous avons choisi, une zone triangulaire d'impossibilité stricte de mise en correspondance (possibilité=0 ; non possibilité=1).

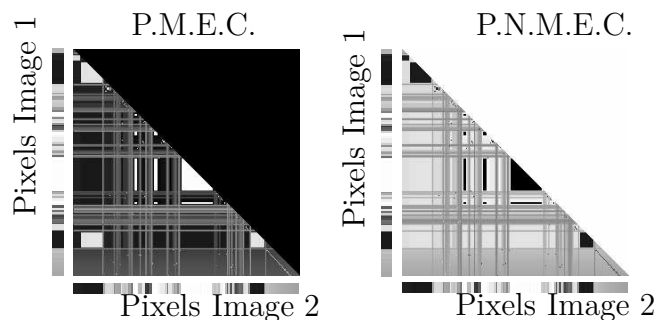


Figure 3.8 – Répartition des possibilités de mise en correspondance (P.M.E.C.) et des possibilités de non mise en correspondance (P.N.M.E.C.) modifiées.

On constate les effets de la contrainte d'unicité : aux pixels blancs (resp. noir) de l'image 1 on a uniquement attribué les pixels blancs (resp. noir) de l'image 2 en rendant non possible les autres appariements. La contrainte d'ordre, un peu de la même manière que la contrainte de point de vue mais de façon plus nuancée, définit des zones de

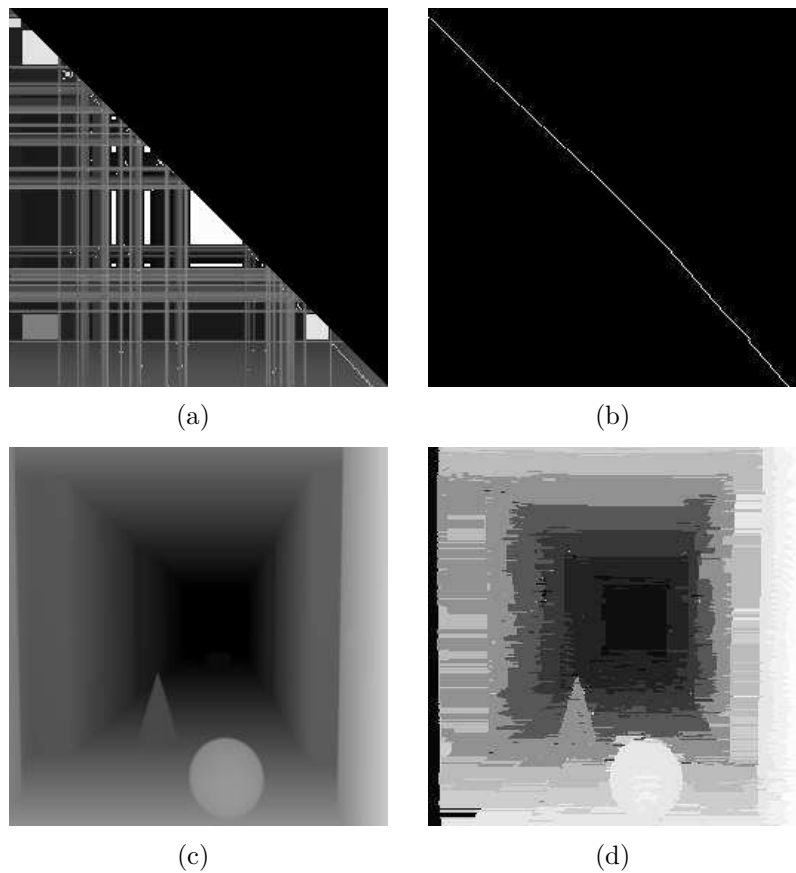


Figure 3.9 – Distribution de la mesure de vérité d'appariement de la ligne x (a), distribution de la vraie mesure de d'appariement de la ligne x (b), image de disparité vraie (c), image de disparité calculée à partir de notre mesure (d).

non possibilité de mise en correspondance.

Sur la figure 3.9a, on peut voir la différence entre la mesure que nous venons de calculer et la vraie distribution de disparité (figure 3.9b) construite à partir de la vraie image de disparité représentée sur la figure 3.9c.

En utilisant un algorithme de recherche d'appariement par maximisation de fonction de ressemblance [CHR96] (ou programmation dynamique) basée sur notre mesure, on peut mettre en correspondance l'ensemble des pixels des deux images. On peut voir sur la figure 3.9d la carte des disparités. Un pixel blanc représente un point proche, un pixel noir représente un point éloigné. On peut constater l'apparition d'un phénomène de traînage propre aux méthodes basées sur la programmation dynamique. On peut remarquer aussi que certains points qui ont normalement le même niveau de disparité (par exemple l'ensemble des points appartenant aux deux pans de mur noirs parallèles aux plans images) ont des niveaux de disparité différents entre les différentes ligne. Ce phénomène est due au fait que les méthodes par programmation dynamique n'apparient que les points de même ordonnée et qu'aucune contrainte inter-lignes n'est utilisé.

3.3.2 Comparaison

Nous proposons, dans ce paragraphe, de comparer les résultats obtenus par notre approche avec les résultats obtenus en utilisant les méthodes classiques d'appariement. Nous avons décidé d'évaluer les appariements en construisant la carte de disparité par des méthodes de mise en correspondance de motif, de programmation dynamique et de coupure de graphe. Pour ces deux dernières méthodes, nous avons tout d'abord utilisé les scores de mise en correspondance classique puis nous avons utilisé notre mesure d'appariement. Nous présentons dans le tableau (3.1), pour chaque méthode, les différents paramètres de simulations utilisés. Dans la suite de ce paragraphe nous rappelons brièvement les méthodes que nous allons utiliser avec leur principaux défauts, nous présentons les images de test ainsi que leur principales difficultés. Enfin, nous montrons les résultats que nous avons obtenus.

3.3.2.1 Les méthodes

Nous rappelons que les méthodes de mise en correspondance de motif cherche la position de la fenêtre (dont la taille est paramétrable) qui minimise un score le long d'une ligne épipolaire et sur une longueur limitée par un paramètre d'excursion maximale. Les scores sont calculés par des distances statistiques entre un motif de référence et un motif candidat. Nous avons retenu, comme distances statistiques, la SAD (Somme des écarts absolus), ZSAD (Somme des écarts absolus centrés), SSD (Somme des écarts quadratiques), ZSSD (Somme des écarts quadratiques centrés), NCC (Intercorrélation normalisée), ZNCC (Intercorrélation normalisée centrée) et CENSUS (Recensement). La principale difficulté à laquelle est confronté l'utilisateur de ces méthodes est le choix de la taille du motif et de la taille de la zone d'excursion : une fenêtre trop petite sera trop sensible aux variations d'illumination et les statistiques ne seront pas robustes, une fenêtre trop grande sera imprécise car, en raison de la différence de point de vue entre les deux images, beaucoup de pixels appartenant aux deux motifs ne seront pas correspondants. Concernant la zone d'excursion maximale, si la longueur est trop petite, la zone de recherche risque de ne pas couvrir le motif correspondant. Si la longueur est trop grande, on risque, dans le cas de motifs répétitifs, d'effectuer un appariement incorrect. Dans les deux cas, on obtient un mauvais appariement. Les deux principaux inconvénients de ces méthodes sont de délivrer un résultat d'appariement même si la zone correspondante est partiellement ou totalement occultée et de délivrer une carte de disparité dont les contours des objets sont floutés ceci étant dû à l'utilisation de la fenêtre glissante.

L'approche basée sur la programmation dynamique cherche, le long d'un ligne épipolaire, à mettre en correspondance l'ensemble des pixels de façon à optimiser une fonction de coût globale. Ce coût global est calculé par agrégation de coûts locaux utilisant un score de mise en correspondance de motif comme, par exemple, la somme des écarts absolus ou SAD. Dans nos expérimentations, nous avons tout d'abord utilisé un score de mise en correspondance locale basée sur une mesure SAD (méthode DP1) puis nous l'avons remplacé dans le processus de programmation dynamique par

Nom de la méthode	Type de méthode	Taille de la fenêtre	Coût d'occultation	Coefficient de lissage	Seuil	Facteur de pénalité du lissage si le gradient est trop petit
SAD	Motif	5×5	–	–	–	–
ZSAD	Motif	5×5	–	–	–	–
SSD	Motif	5×5	–	–	–	–
ZSSD	Motif	5×5	--	–	–	–
NCC	Motif	5×5	–	–	–	–
ZNCC	Motif	5×5	–	–	–	–
CENSUS	Motif	5×5	–	–	–	–
DP1	Programmation dynamique et SAD	–	20	20	–	–
DP2-Fuzzy	Programmation dynamique et notre mesure	–	20	20	–	–
GC1	Coupure de graphe et SAD	–	–	20	4	2
GC2	Coupure de graphe et SAD	–	–	20	8	2
GC3	Coupure de graphe et SAD	–	–	50	8	4
GC4-Fuzzy	Coupure de graphe et notre mesure	–	–	20	4	2
GC5-Fuzzy	Coupure de graphe et notre mesure	–	–	20	8	2
GC6-Fuzzy	Coupure de graphe et notre mesure	–	–	50	8	4

Tableau 3.1 – Paramètres de simulation.

notre mesure d'appariement (DP2-Fuzzy). L'avantage de cette approche par rapport aux méthodes de mise en correspondance de motif est qu'elle est moins sensible aux régions ambiguës (occultations, manque de texture, etc.) et ce grâce à l'utilisation de contraintes globales. Les deux principaux inconvénients sont qu'une erreur locale est propagée tout le long de la ligne de recherche corrompant ainsi le coût global et, comme la mise en correspondance ne s'effectue uniquement ligne à ligne, on constate généralement un effet de traînage dû au manque de cohérence bidimensionnelle. Dans le cas de la méthode DP1, le score de mise en correspondance étant un score de mise en correspondance de motif, on retrouve les effets de la fenêtre glissante sur le résultat de l'appariement global : les détails des objets dans la carte de disparité disparaissent au profit de zones homogènes.

Pour pallier l'inconvénient de manque de cohérence, les méthodes basées sur les coupures de graphe reprennent l'approche par programmation dynamique (c'est à dire l'utilisation d'un ensemble de score de mise en correspondance locale représenté par l'image de l'espace des disparités ou DSI) et tentent d'estimer la fonction de disparité en minimisant une fonction de coût tridimensionnelle (ligne \times colonne \times disparité). Dans nos expérimentations, nous avons tout d'abord utilisé un score de mise en correspondance locale basé sur une mesure de type SAD avec trois jeux de paramètres différents (méthodes GC1, GC2 et GC3) puis nous avons utilisé dans le processus de coupure de graphe notre mesure d'appariement (méthodes GC4-Fuzzy, GC5-Fuzzy et GC6-Fuzzy). Le principal inconvénient est, en plus d'un temps de calcul incroyablement grand en comparaison des deux précédentes approches, que les cartes de disparité obtenues sont généralement aplaties, c'est à dire qu'elles sont constituées de zones de même profondeur dont le nombre est relativement faible. A l'instar de l'approche par programmation dynamique, l'utilisation de fenêtres pour le calcul du score de mise en correspondance locale a tendance à créer des zones homogènes.

3.3.2.2 Les images

Le premier couple d'images stéréoscopiques que nous allons appairer est présenté sur la figure 3.10. Il s'agit d'images de synthèse : c'est à dire que les variations d'illumination pour deux pixels correspondant sont faibles et proviennent uniquement de la variation de point de vue. La différence de position entre les deux vues est très faible ce qui nous a conduit à choisir un coefficient d'excursion maximal de 15 pixels pour les méthodes de mise en correspondance de motif. La scène contient peu d'occultations et le sol a, comme texture, un motif à damier répétitif, ce qui posera des problèmes pour les scores de mise en correspondance de motifs.

La deuxième paire stéréoscopique représentée sur la figure 3.11 est relativement difficile à mettre en correspondance du fait des textures complexes et répétitives et surtout de la grande variation des disparités : en effet, la disparité maximale pour l'objet le plus proche est d'environ 50 pixels. On est très loin ici de l'hypothèse de faible déplacement qui est généralement utilisée pour des mises en correspondance basées sur les valeurs d'illuminations (cf §2.2.1). C'est pourquoi nous avons choisi un coefficient d'excursion maximal de 60 pixels pour les méthodes de mise en correspondance de

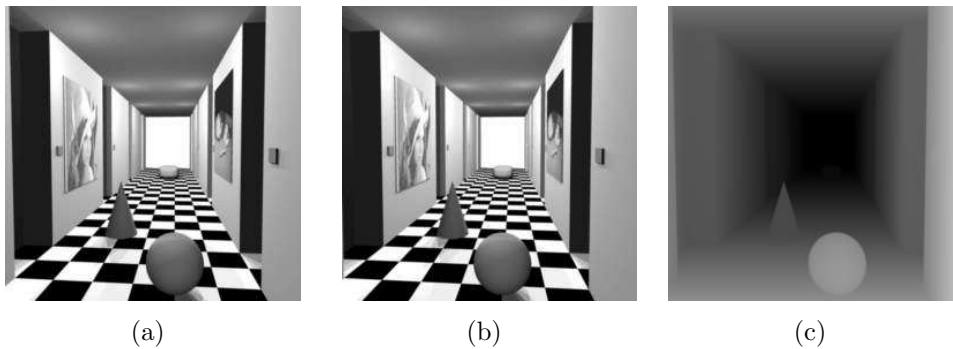


Figure 3.10 – Images stéréoscopiques 1 : Corridor de Lena, image gauche (a), image droite (b), image de disparité vraie (c).

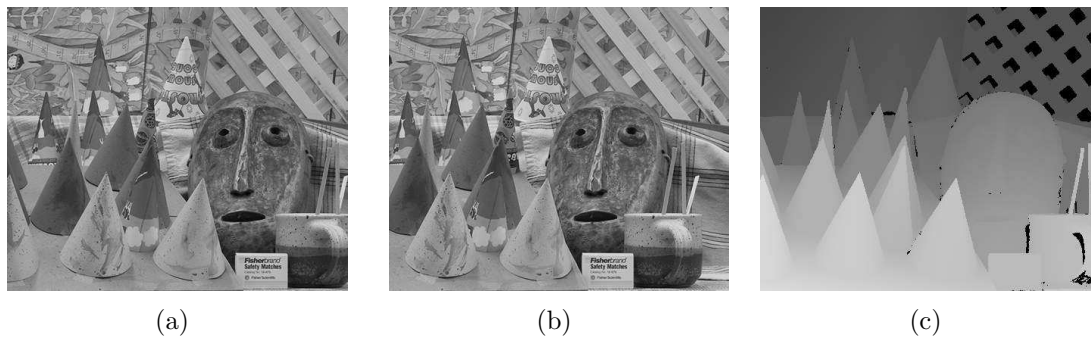


Figure 3.11 – Images stéréoscopiques 2 : Cônes de Scharstein et Szeliski, image gauche (a), image droite (b), image disparité vraie (c).

motif. La différence entre les deux vues étant grande, il existe dans cette paire beaucoup de pixel occultés, ce qui va engendrer dans les méthodes de mise en correspondance de motifs, de nombreux faux appariements. De plus, comme c'est une scène réelle, deux pixels correspondants ont des niveaux de gris relativement différents en raison des variations d'illumination, de réflectance et de réglages de caméras différents. De ce fait, les méthodes de mise en correspondance globale comme la programmation dynamique, auront tendance à propager des faux appariements le long des lignes de recherche. Ce phénomène sera amoindri par l'ajout de la cohérence bidimensionnelle dans le cas des méthodes par coupure de graphe.

Enfin, pour la troisième paire d'image stéréoscopique de la figure 3.12, il s'agit encore d'images naturelles. Les différents plans de disparité sont en nombre limité et la différence de point de vue est moyenne. Nous avons choisi un coefficient d'excursion maximal de 40 pixels pour les méthodes de mise en correspondance de motif. Les principales difficultés de mise en correspondance pour ce couple d'images sont un arrière plan fortement texturé et un grand nombre d'occultations ce qui engendrera de nombreux faux appariements pour les méthodes de mise en correspondance de motif. Concernant la zone du bras de la lampe, elle engendre localement des fortes discontinuités dans la fonction de disparité. Cette discontinuité aura tendance à disparaître dans les mé-

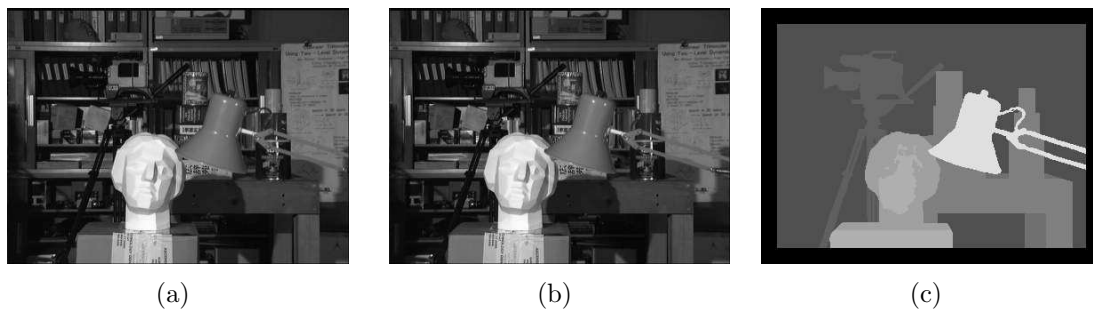


Figure 3.12 – Images stéréoscopiques 3 : Statue de Tsukuba, image gauche (a), image droite (b), image disparité vraie (c).

thodes de mise en correspondance globale au profit de plus grandes régions de disparité uniforme.

3.3.2.3 Résultats des méthodes de mise en correspondance de motif

Tout d'abord, concernant les résultats de la figure 3.13, on peut remarquer de nombreux faux appariements (damier du sol, le mur à droite, la sphère). Ceci est dû aux motifs répétitifs du sol et au manque de texture du mur. De plus, un effet de flou apparaît sur le contour des objets dû à l'utilisation d'une fenêtre glissante pour le calcul du score de mise en correspondance.

On retrouve ces mêmes défauts sur la figure 3.14. Du fait de la différence importante entre les deux images, il existe de nombreux pixels occultés. Ainsi, ces zones occultées sont mis en correspondance de façon incorrecte (frontière entre les cônes principalement). C'est le principal défaut des méthodes utilisant des scores de mise en correspondance de motifs : celui de délivrer un score, même si le motif correspondant est partiellement occulté. Dans ce genre d'image le compromis entre une fenêtre de petite taille qui permet d'obtenir des résultats précis mais peu fiables et une fenêtre de grande taille dont l'utilisation permet d'obtenir des résultats plus fiables mais peu précis est difficile à trouver. On constate alors que, dans ce cas, la méthode qui obtient le meilleur résultat est la méthode basée sur le recensement (CENSUS c.f. §1.3.4.2) qui est plus robuste aux variations de luminosité que les statistiques de moments.

Concernant les résultats de la figure 3.15, on peut encore remarquer un certain nombre de faux appariements (les pixels fortement blancs). Ceci est dû à un arrière-plan fortement texturé associé à une zone de recherche trop grande. C'est l'autre principal défaut des méthodes de mise en correspondance de motifs, c'est à dire le choix judicieux de l'excursion maximale de la fenêtre glissante. On constate aussi que les statistiques de moments normalisés génèrent plus de résultats erronés que leur version non-normalisée. Ceci est dû à l'arrière plan fortement texturé.

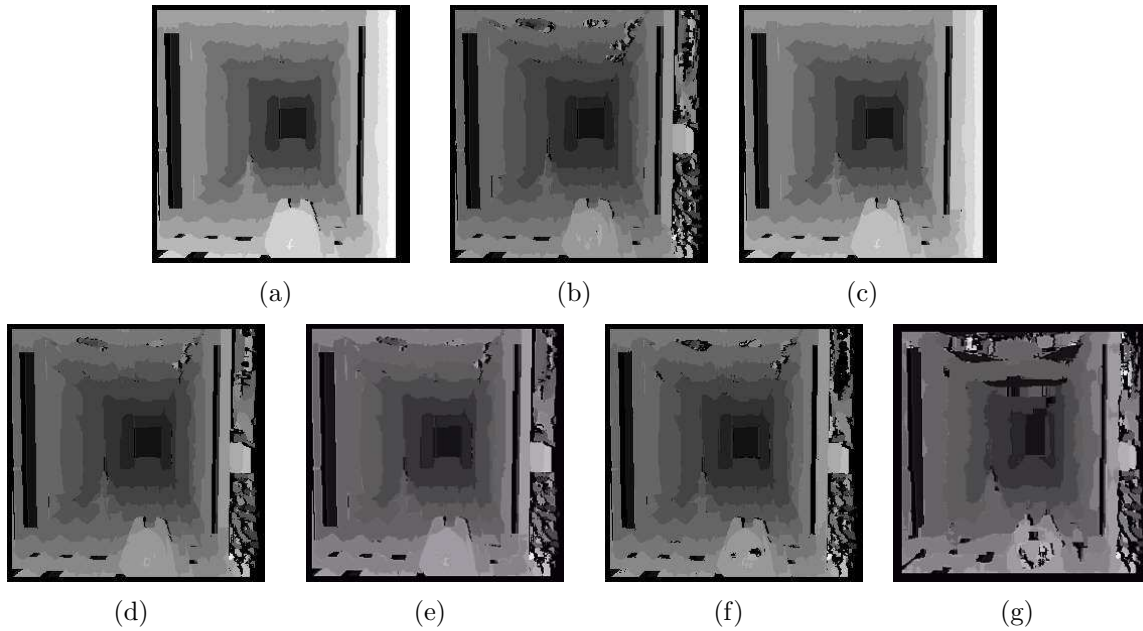


Figure 3.13 – Résultats des méthodes de mise en correspondance de motifs, images stéréoscopique1 : SAD(a), ZSAD(b), SSD(c), ZSSD(d), NCC(e), ZNCC(f), CENSUS(g).

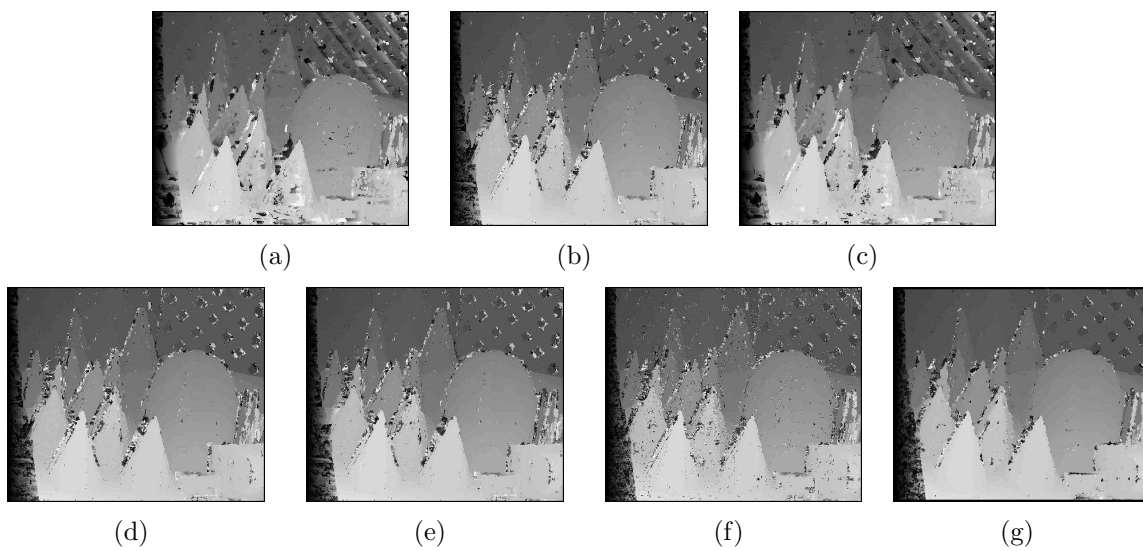


Figure 3.14 – Résultats des méthodes de mise en correspondance de motifs, images stéréoscopiques2 : SAD(a), ZSAD(b), SSD(c), ZSSD(d), NCC(e), ZNCC(f), CENSUS(g).

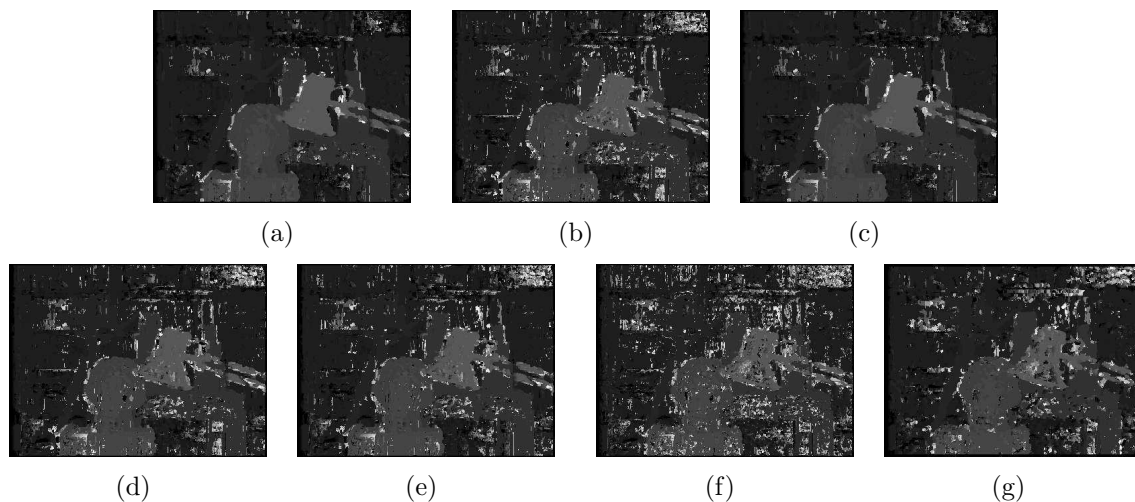


Figure 3.15 – Résultats des méthodes de mise en correspondance de motifs, images stéréoscopiques³ : SAD(a), ZSAD(b), SSD(c), ZSSD(d), NCC(e), ZNCC(f), CENSUS(g).

3.3.2.4 Résultats des méthodes par programmation dynamique

Concernant les méthodes basées sur la programmation dynamique, il y a beaucoup moins de points erronés du fait de la minimisation du score sur la totalité de la ligne (figure 3.16a, figure 3.17a et figure 3.18a). Pour la paire stéréoscopique de la figure 3.16a, les scores de mise en correspondance obtenus par la mesure SAD (figure 3.13a) étaient peu bruités. L'utilisation de la contrainte globale de mise en correspondance a permis de lever les ambiguïtés d'appariement des motifs répétitifs. Par contre, l'utilisation d'une fenêtre glissante pour le calcul du score de mise en correspondance de motif tend à lisser la fonction de disparité. On constate que l'utilisation de notre mesure de mise en correspondance (figure 3.16b) permet de préserver un grand nombre de détails du fait que nous n'utilisons pas de fenêtre pour le calcul de notre mesure.

Pour les résultats des figures 3.17a et 3.18a, comme il y a beaucoup de points erronés obtenus par le score de mise en correspondance locale, ces erreurs locales se propagent sur la totalité de la ligne. On constate que l'utilisation de notre mesure de mise en correspondance (figure 3.17b et figure 3.18b) permet d'obtenir une carte de disparité utilisable du fait que les points appariés l'ont été de façon nuancée et que les données utilisées ont été modifiées par notre processus de mise en correspondance. De plus, l'utilisation de notre mise en correspondance chromatique permet de diminuer l'impact des variations de luminosité pour des points correspondants. Les résultats obtenus sur la figure 3.18a nous amène à constater que le coefficient de lissage ainsi que le coût d'une occultation sont trop importants. Néanmoins l'utilisation de notre mesure de mise en correspondance (figure 3.18b) permet d'obtenir une carte de disparité plus précise pour les objets au premier plan du fait de l'orientation du choix d'appariement du chemin à coût minimal par notre processus de mise en correspondance. Cependant, concernant l'arrière plan, les résultats obtenus par notre méthode sont moins bons que ceux obtenus par un score classique. Ceci est dû à l'adaptation de notre mesure à la

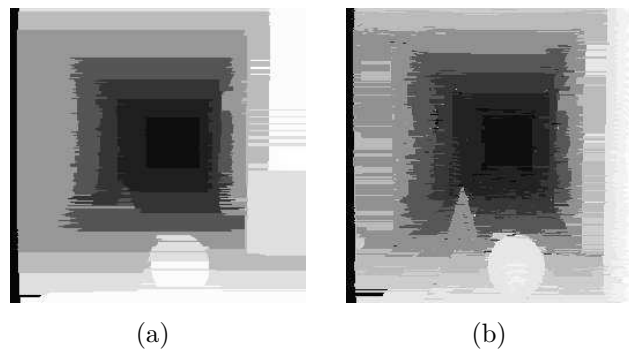


Figure 3.16 – Résultats des méthodes de mise en correspondance par programmation dynamique, images stéréoscopiques 1 : DP1 (a), DP2-Fuzzy (b).

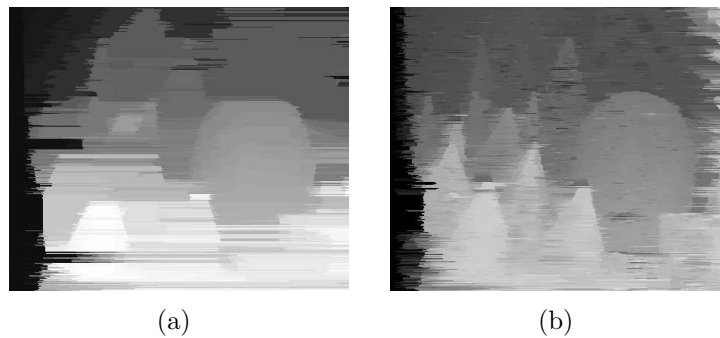


Figure 3.17 – Résultats des méthodes de mise en correspondance par programmation dynamique, images stéréoscopiques 2 : DP1 (a), DP2-Fuzzy (b).

programmation dynamique. Il n'y a pas suffisamment de points fortement blancs ou fortement noirs pour pouvoir différencier la caméra de l'arrière-plan.

3.3.2.5 Résultats des méthodes par coupure de graphe

Enfin, concernant les méthodes basées sur la coupure de graphe utilisant les scores de mise en correspondance de motifs (figure 3.19a-c, figure 3.20a-c et figure 3.21a-c), on constate une rapide disparition des détails quand le coefficient de lissage augmente et ce au profit des zones homogènes. On retrouve aussi le défaut principal de cette méthode : une carte de disparité aplatie, c'est à dire avec peu de niveaux de disparité différents. Ces défauts ont déjà été relevés dans le chapitre 1. On peut constater que l'utilisation de notre mesure de mise en correspondance (figure 3.19d-f, figure 3.20d-f et figure 3.21d-f) permet de garder un plus grand nombre de régions de disparité différentes. Ainsi les cartes de disparité de la figure 3.21d-e sont beaucoup plus bruitées que la carte de la figure 3.21f. Ceci provient du fait qu'il existe, dans nos scores de mise en correspondance, une trop grande diversité de niveaux de disparité que seul un fort coefficient de lissage permet de niveler.

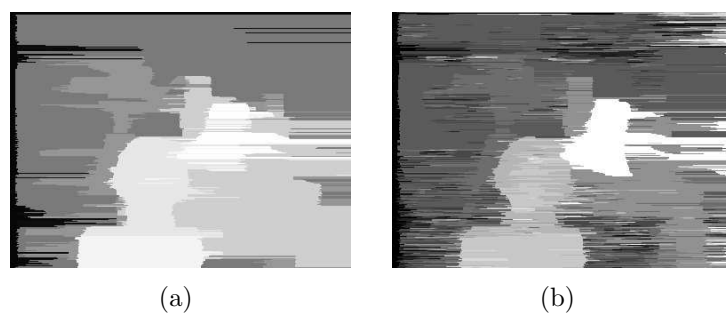


Figure 3.18 – Résultats des méthodes de mise en correspondance par programmation dynamique, images stéréoscopiques 3 : DP1 (a), DP2-Fuzzy (b).

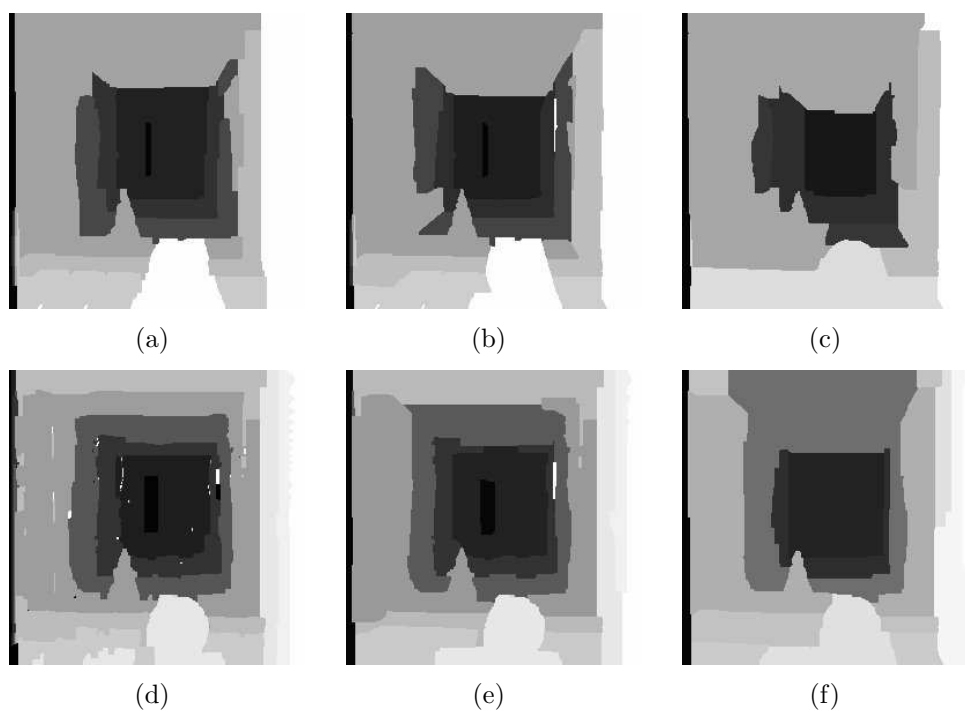


Figure 3.19 – Résultats des méthodes de mise en correspondance par coupure de graphe, images stéréoscopiques 1 : GC1 (a), GC2 (b), GC3 (c), GC4-Fuzzy (d), GC5-Fuzzy (e), GC6-Fuzzy (f).

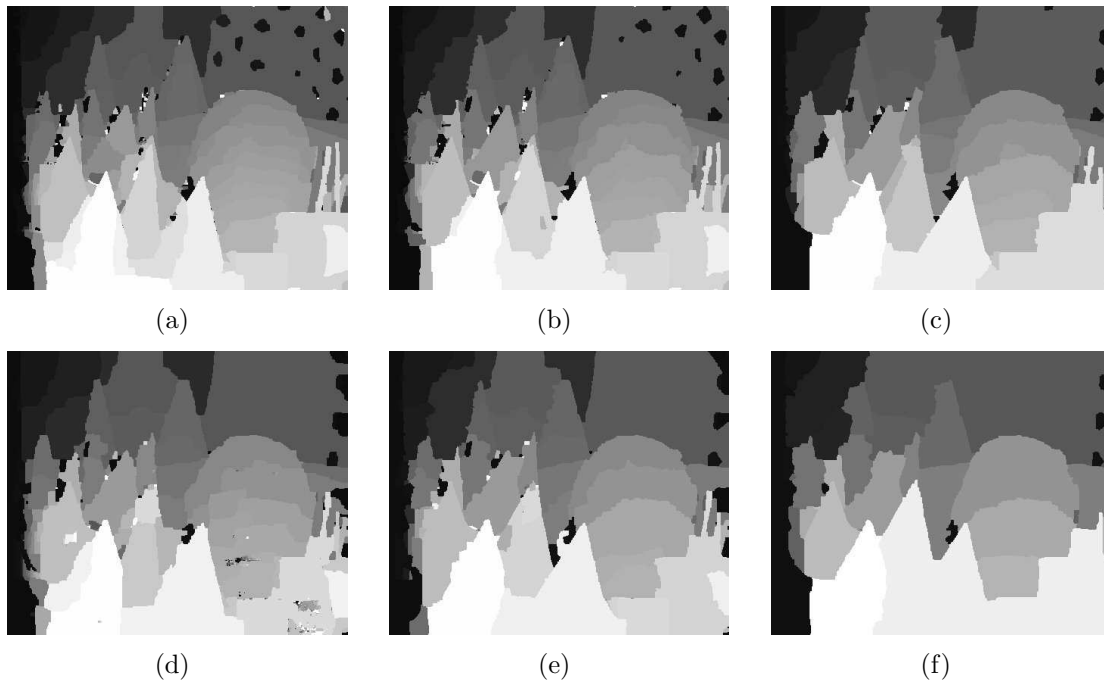


Figure 3.20 – Résultats des méthodes de mise en correspondance par coupure de graphe, images stéréoscopiques 2 : GC1 (a), GC2 (b), GC3 (c), GC4-Fuzzy (d), GC5-Fuzzy (e), GC6-Fuzzy (f).

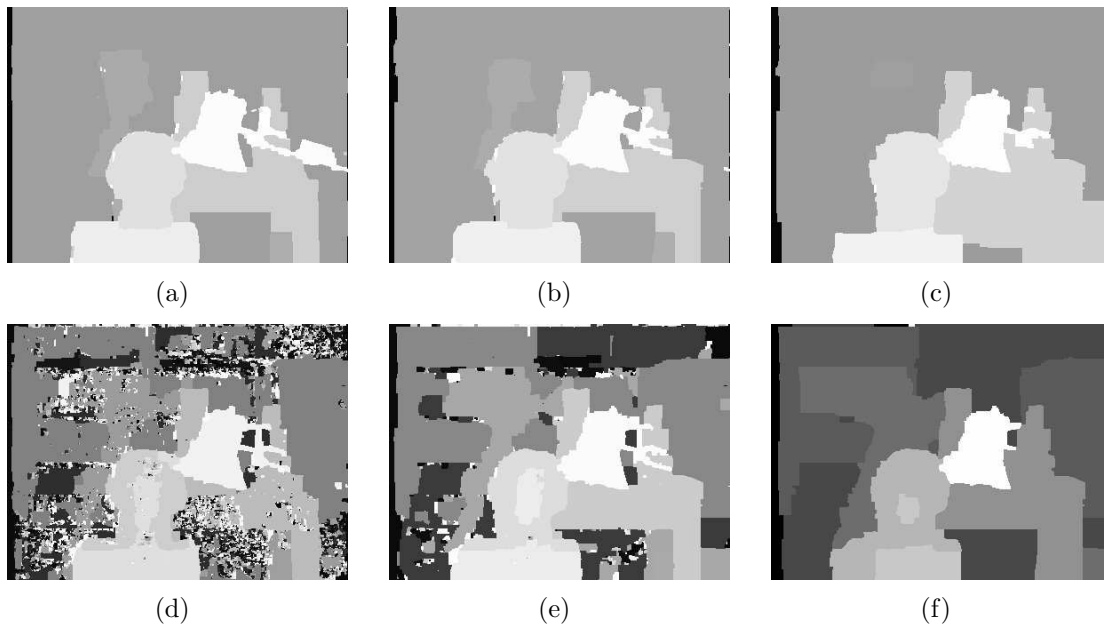


Figure 3.21 – Résultats des méthodes de mise en correspondance par coupure de graphe, images stéréoscopiques 3 : GC1 (a), GC2 (b), GC3 (c), GC4-Fuzzy (d), GC5-Fuzzy (e), GC6-Fuzzy (f).

3.4 Conclusion

Les mesures de possibilités de mise en correspondance présentées dans le chapitre 2 ne suffisent pas à définir des domaines de mise en correspondances dans les images stéréoscopiques suffisamment spécifiques pour en permettre une utilisation simple. Pour augmenter la spécificité des domaines, nous proposons une généralisation des contraintes géométriques utilisées habituellement en stéréovision calibrée. L'approche proposée se différencie des approches classiques par le fait qu'elle se base sur des hypothèses beaucoup moins contraignantes que les méthodes usuelles vis-à-vis des contraintes d'appariements. Au cours de ce chapitre, nous avons proposé une méthode permettant d'inclure des contraintes telles que : la contrainte d'ordre, d'unicité et de point de visée. Ces contraintes améliorent la spécificité des domaines d'appariement en gardant l'aspect "prudent" de la démarche possibiliste.

L'approche présentée consiste à exprimer ces contraintes sous forme de règles sûres (règle dont le niveau de coupe est 1) et à utiliser ces règles pour faire une modification bipolaire de la distribution des mesures de confiance d'appariement. Pour chaque modification des domaines de mise en correspondance, nous avons utilisé les appariements dont les mesures de confiance étaient les plus élevées pour modifier les appariements de niveau de confiance plus faibles.

Les domaines d'appariement modifiés que nous obtenons ne nous permettent pas de déterminer facilement des cartes de disparité. Néanmoins, nous avons pu comparer nos résultats avec les résultats obtenus par les approches classiques en établissant une mesure de mise en correspondance sur des images redressées.

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

- Les performances de ce type de score de mise en correspondance sont très satisfaisantes. En effet, les résultats sont quasiment toujours meilleurs que ceux obtenus par les méthodes traditionnelles.
- Notre méthode possède un bon comportement vis-à-vis des scènes fortement texturées et ayant de fortes variations de disparité. L'utilisation de la contrainte d'unicité sous sa forme floue permet, dans le cas de fortes disparités, de ne pas apparier un pixel d'une image avec un seul autre dans l'autre image mais de mettre en correspondance des blocs de pixels ayant le même niveau de confiance d'appariement.
- Notre méthode permet, dans le cas de l'utilisation de la programmation dynamique, d'orienter la recherche du chemin à coût minimum grâce aux contraintes d'unicité et d'ordre.

Notre méthode possède cependant quelques inconvénients et nécessite des améliorations. Actuellement, le temps de calcul est beaucoup trop élevé (15 secondes pour l'image du corridor contre 2 secondes pour la méthode de programmation dynamique classique). Ceci est dû à l'étude du respect ou de la violation des contraintes pour chaque niveau de coupe. Actuellement, seule la contrainte d'unicité peut être calculée

sans tenir compte des différents niveaux de coupes (sauf pour $\alpha = 0.5$ car les possibilités de compatibilité chromatique changent à ce niveau de coupe). La contrainte d'ordre est la contrainte qui nécessite le plus de temps de calcul.

Le deuxième point à améliorer est le calcul des relations spatiales "à droite" et "à gauche". Nous pensons utiliser pleinement les travaux de Bloch [Blo99], [BS02] afin de définir cette fois des domaines flous en lieu et place des domaines binaires actuels.

Un troisième point à améliorer concerne les modifications des distributions de possibilité de mise en correspondance et de non-mise en correspondance. Plutôt que de transférer la confiance que l'on a dans un appariement sur un autre appariement de confiance inférieur, il serait peut être intéressant d'utiliser la théorie de l'évidence de Dempster-Shafer [Dem67], [Sha76], [YFK94] pour réévaluer la confiance que l'on a dans un appariement en considérant les autres appariements.

Il reste à trouver une méthode de choix d'appariement final qui puisse respecter l'aspect garanti de l'information et qui puisse traiter des images autres que celles qui sont redressées. Nous pensons qu'il faudra chercher du côté des méthodes de coupures de graphes à 5 dimensions (ligne et colonne de chaque image plus la disparité) au lieu des 3 actuellement (ligne de chaque image et la disparité). Tout cela permettra de rendre l'estimation de la carte de disparité encore plus robuste et ce, sur tout type d'images stéréoscopiques. Finalement, il serait peut-être plus judicieux de chercher une méthode qui calcule directement, à partir des mesures de mise en correspondance, des volumes tridimensionnels.

Conclusion générale et perspectives

Le principe de la stéréovision consiste à reconstruire une information de relief apparent à partir d'au moins deux caméras. Cette reconstruction nécessite d'une part de connaître un modèle de projection de la luminance des caméras ainsi que leur position relative lors de la prise d'images et, d'autre part, d'être capable de mettre en correspondance les projections, sur chaque rétine, des points d'un même objet tridimensionnel.

Le travail de thèse, que nous présentons dans ce manuscrit, s'inscrit dans la problématique de la mise en correspondance dense de pixels dans un couple d'images stéréoscopiques à niveau de gris. Nous proposons une nouvelle approche se basant sur une représentation floue de l'information spatiale de luminance délivrée par les capteurs d'images numériques. Pour bien expliquer notre démarche, il nous a semblé nécessaire de bien reprendre les bases de la formation des images numériques en insistant sur les étapes d'échantillonnage et de numérisation. L'impact de la représentation de l'information de luminance est en effet prépondérant sur la mise en correspondance puisque celle-ci s'appuie sur l'hypothèse que la luminance émise par tout point de l'environnement tridimensionnel peut être considérée comme uniforme si la variation de point de vue est faible.

Nous avons dressé un panorama des méthodes de mise en correspondance proposées par la littérature (mise en correspondance de motifs, d'indices, mise en correspondance par programmation dynamique, par coupure de graphe, mise en correspondance coopératives). Nous avons remarqué que la plupart de ces méthodes utilisent une représentation statistique de l'erreur d'appariement. Cependant, on est en droit de se poser des questions sur l'utilisation d'outils de statistique lorsque le nombre d'informations à mettre en correspondance ne dépasse pas deux. Pour pallier le biais de cette approche, il est d'usage d'utiliser des techniques de régularisation pour estimer les moments statistiques de la distribution de luminance locale. Ces techniques s'appuient sur l'hypothèse d'ergodicité du bruit de mesure en utilisant le voisinage de chaque pixel. Plus le voisinage utilisé est étendu, plus la statistique sera précise, mais moins elle sera fiable. Le manque de fiabilité est lié aux régions ambiguës (zones peu ou fortement texturées, scènes ayant beaucoup de disparités). La violation fréquente des hypothèses d'ergodicité et de luminance similaires pour deux pixels correspondants compromet le bon fonctionnement de la mise en correspondance et réduit donc la robustesse de ce

type de méthode. Plutôt que de tenter d'augmenter la fiabilité des résultats obtenus lors de la mise en correspondance par une complexification du modèle des mesures de similarité, nous avons préféré opter pour une approche visant à minimiser l'influence des a-priori sur le processus d'appariement. Nous avons basé notre raisonnement sur une approche faisant appel à la théorie des possibilités. Cette démarche nous a permis de remplacer la notion de probabilité de mise en correspondance par la notion de domaine pondéré des correspondants possibles. La grande différence entre ces deux notions est que les domaines obtenus sont imprécis mais que leur fiabilité est ordonnée.

Notre objectif principal, dans ce travail sur les appariements, est de réaliser une méthode de mise en correspondance qui soit fiable et robuste vis-à-vis de perturbations telles que le processus d'intégration spatial et temporel de la luminance ainsi que par des phénomènes tels que l'occultation partielle, le changement de point de vue ou la mise en défaut du modèle de réflexion Lambertien.

Nous avons donc orienté nos travaux sur une reformulation de l'hypothèse d'appariement de deux pixels à partir de la mesure de luminance.

Nous avons défini une nouvelle représentation de l'information échantillonnée de la luminance permettant de nous abstraire de la technologie d'obtention des images. Cette représentation s'appuie sur une modélisation par une partition floue de l'imprécision de la mesure de luminance induite par l'échantillonnage spatial et sur une classification duale de l'information de luminance associée à chaque pixel. Nous avons associé, à cette représentation, une modélisation de la contrainte épipolaire propre aux capteurs stéréoscopiques étalonnés permettant de prendre en compte l'imprécision des paramètres d'étalonnage. Cette représentation permet d'associer, à chaque pixel d'une des images, un domaine pondéré discret des pixels de l'autre image pouvant lui correspondre géométriquement et chromatiquement.

Cette distribution n'étant pas assez spécifique (les domaines de mise en correspondance pour un niveau de confiance donné sont trop grands), nous proposons de les modifier en utilisant des contraintes supplémentaires, telles que l'ordre ou l'unicité. Nous utilisons, pour ce faire, un processus itératif et coopératif bipolaire. Les distributions de possibilités que nous obtenons sont plus spécifiques. C'est l'ensemble de cette démarche qui permet de minimiser l'influence de la représentation du problème sur les mesures d'appariements.

La validité de notre approche a été établie par des expérimentations sur des images stéréoscopiques de synthèses et réelles. Ces dernières nous ont permis de tester le comportement de notre méthode vis-à-vis des variations d'illumination et de disparité. Nous avons pu constater que les résultats obtenus équivalaient ceux des méthodes globales classiques utilisant des scores de mise en correspondance de motif. Nous avons montré que la recherche des correspondants, pour les méthodes basées sur la programmation dynamique, étaient orientée par notre processus de spécification des domaines d'appariements. L'utilisation de notre processus nous a permis d'obtenir, dans certain cas, des cartes de disparité plus exactes que les méthodes traditionnelles.

Les perspectives de ce travail sont multiples. Tout d'abord, vis à vis du modèle de

pixel, il serait intéressant de regarder l'extension en 2D de la proposition Dubois et al. [DFMP04] concernant le lien existant entre noyau sommatif et noyau flou. Il serait nécessaire de quantifier l'apport de la prise en compte du voisinage dans l'information chromatique du pixel flou.

Concernant le processus d'augmentation de la spécificité des domaines de mise en correspondance, il serait intéressant de tester une méthode utilisant des modifications de masses de croyance plutôt que d'utiliser des modifications de possibilité et de nécessité conditionnelles. Il faudrait aussi, en lieu et place des domaines "à droite" et "à gauche" binaire, s'inspirer des travaux d'I. Bloch [Blo99] pour mieux modéliser la contrainte d'ordre.

Nous avons vu dans l'état de l'art qu'il existait certains points pour lesquels la proposition de la contrainte d'ordre était inversée : il s'agit des points contenus, comme l'a appelé Faugeras, dans le cône interdit. La prise en compte de cette particularité permettrait de mieux apparier les points provenant de scènes complexes ayant de fortes variations de disparités.

Il serait peut être intéressant de voir comment on pourrait modifier notre processus de mise en correspondance de domaine avec l'adjonction d'une troisième vue pour lever plus d'ambiguïtés d'appariement.

Enfin, le développement d'un algorithme de mise en correspondance globale utilisant la totalité des domaines de mise en correspondance modifiés serait la première amélioration à effectuer pour augmenter la précision des résultats. Reste que la complexité d'un tel processus est loin d'être négligeable. Puisque l'objectif final est de reconstruire des volumes tridimensionnels, il serait plus judicieux d'éviter le calcul de la carte des disparités et de calculer des volumes d'occupations directement à partir des données imprécises. Une application directe serait un processus de calcul itératif d'occupation de l'espace environnant le système stéréoscopique. Chaque nouvelle vue stéréoscopique supplémentaire permettrait de les modifier et d'en augmenter la précision et la fiabilité.

Bibliographie

- [AEH98] N. Andreff, B. Espiau, and R. Horaud. Une methode d'auto-etalonnage pince-camera. Technical report, INRIA, Octobre 1998.
- [AG93] P. Aschwenden and W. Guggenbuhl. Experimental results from a comparative study on correlation-type registration algorithms. *Robust Computer Vision*, pages 268–289, 1993.
- [Aya89] N. Ayache. *Vision Stereoscopique et Percepction Multisensorielle, Application a la robotique mobile*. Science Informatique, 1989.
- [Bai90] G. Baillargeon. *Methodes statistiques de l'ingenieur*, volume 1. Editions SMG, 1990.
- [BB81] H. H. Baker and T. O. Binford. Depth from edge and intensity based stereo. *7th International Joint Conference on Artificial Intelligence*, pages 631–636, 1981.
- [BB96] S. Betge-Brezetz. *Modelisation Incrementale et Localisation par Amers pour la Navigation d'un Robot Mobile Autonome en Environnement Naturel*. PhD thesis, LAAS, 1996.
- [BBP67] H. B. Barlow, C. Blakemore, and J. D. Pettigrew. The neural mechanism of binocular depth discrimination. *J. Physiol. Lond.*, 193 :327–342, 1967.
- [BC01] J. Banks and P. Corke. Quantitative evaluation of matching methods and validity measures for stereo vision. *Int'l J. Robotics Research*, 20(7) :512–532, 2001.
- [BDKP01] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Modification possibiliste de connaissances et de preferences. *Journee nationales sur les modeles de raisonnement JNMR*, pages 93–108, 2001.
- [BDKP02] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bipolar possibilistic representations. *Uncertainty in Artificial Intelligence*, 2002.
- [Bel96] P. N. Belhumeur. A bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19(3) :237–262, 1996.
- [BFZ93] M. Buffa, O. Faugeras, and Z. Zhengyou. A stereovision-based navigation system for a mobile robot. *Rapport de recherche INRIA*, May 1993.
- [BGCM02] C. Buehler, S. Gortler, M. Cohen, and L. McMillan. Minimal surfaces for stereo. *Proceedings of European Conference on Computer Vision*, pages 885–899, 2002.

- [BHF96] F. Bigone, O. Henricsson, and P. Fua. Automatic extraction of generic house roofs from high resolution aerial imagery. *Proc. European Conf. Computer Vision*, pages 85–96, 1996.
- [BI99] A. F. Bobick and S. S. Intille. Large occlusion stereo. *International Journal on Computer Vision (IJCV)*, 33(3) :181–200, 1999.
- [Blo99] I. Bloch. Fuzzy relative position between objects in image processing : A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7) :657–664, July 1999.
- [BM93] B. Bouchon-Meunier. *La logique floue*. P.U.F. (Que-sais-je?), 1993.
- [BN98] D. N. Bhat and S. K. Nayar. Ordinal measures for visual correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 :415–423, 1998.
- [BS02] I. Bloch and A. Saffioti. Les robots et la morphologie mathématique floue. *Rencontres Francophone sur la Logique Floue et ses Application (LFA)*, pages 145–152, Octobre 2002.
- [BT98a] S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. *IEEE Int'l Conf. Computer Vision*, pages 1073–1080, 1998.
- [BT98b] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4) :401–406, 1998.
- [BT99] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. *Int'l Joint Conf. Artificial Intelligence*, 1 :489–495, 1999.
- [BVZ98] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pages 648–655, 1998.
- [BVZ99] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Proceedings of International Conference on Computer Vision*, 1 :377–384, 1999.
- [Can86] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence*, 8(6) :679–698, 1986.
- [CDW76] P. G. H. Clarke, I. M. L. Donaldson, and D. Whitteridge. Binocular visual mechanisms in cortical area. *J. Physiol. Lond.*, 256 :509–526, 1976.
- [CHMR92] I. J. Cox, S. L. Hingorani, B. M. Maggs, and S. B. Rao. Stereo without disparity gradient smoothing : a bayesian sensor fusion solution. *Proceedings of British Machine Vision Conference*, pages 337–346, 1992.
- [CHR96] I. J. Cox, S. L. Hingorani, and S. B. Rao. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3) :542–567, 1996.
- [Cle02] M. Clerc. Wavelet-based correlation for stereopsis. *Proceedings 7th European Conference on Computer Vision*, 2 :495–509, 2002.

- [CM99] Q. Chen and G. Medioni. A volumetric stereo matching method : Application to image-based modeling. *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pages 29–34, 1999.
- [CMPC96] A. Crouzil, L. Massip-Pailhes, and S. Castan. A new correlation criterion based on gradient fields similarity. *Proceedings of 13th IAPR International Conference on Pattern Recognition*, pages 632–636, August 1996.
- [Com01] F. Comby. *Estimation du Mouvement Apparent Majoritaire dans une Sequence d'Images Video par Accumulation de Votes Bimodaux sur un Histogramme Approche*. PhD thesis, Universite de Montpellier II, 2001.
- [Cox94] I. J. Cox. A maximum likelihood n-camera stereo algorithm. *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pages 733–739, 1994.
- [CP95] J.P. Coquerez and S. Philipp. *Analyses d'images : Filtrage et Segmentation*. "Enseignement de ". Masson, 1995.
- [CW81] M. Courtois and G. Weill. The spot satellite remote sensing mission. *Photogrammetric Engineering and Remote Sensing*, pages 1163–1171, 1981.
- [DA89] U. R. Dhond and J. K. Aggarwal. Structure from stereo - a review. *IEEE Trans ; Systems, Man and Cybernetics*, 19 :1489–1510, 1989.
- [Day00] C. R. Dayer. *Volumetric Scene Reconstruction from Multiple View*, volume Foundations of Image Analysis. Kluwer, 2000.
- [DD03] H. Prade D. Dubois. Informations bipolaires : Une introduction. *Information-Interaction-Intelligence*, 3(1) :89–106, 2003.
- [Dem67] A. P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38, 1967.
- [Der87] R. Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *International journal of Computer Vision*, 1(2) :167–187, 1987.
- [DF90] R. Deriche and O. Faugeras. 2d curve matching using high curvature points. *Proceedings of International Conference on Pattern Recognition*, pages 240–242, June 1990.
- [DFMP04] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10 :273–297, 2004.
- [DP80] D. Dubois and H. Prade. *Fuzzy Sets and Systems*. Academic Press, 1980.
- [DP88] D. Dubois and H. Prade. *Possibility Theory : An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.
- [DP92] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49 :65–74, 1992.
- [DP94] D. Dubois and H. Prade. Fuzzy sets, a convenient fiction for modeling vagueness and possibility. *IEEE Transactions on Fuzzy Systems*, pages 16–21, 1994.

- [DP96a] D. Dubois and H. Prade. Fuzzy sets in approximate reasoning : A personal view. *Implementations and Applications for Fuzzy Logic*, M.J. Patyra, D.M. Mlynek (Eds.), Wiley and B.G. Teubner, New York and Stuttgart, pages 3–35, 1996.
- [DP96b] D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84 :169–185, 1996.
- [DPS93] D. Dubois, H. Prade, and S. Sandri. *On Possibility/Probability Transformations*, pages 103–112. Kluwer Academic Publishers, r. lowen and m. roubens edition, 1993.
- [Fau93] O. Faugeras. *Three-Dimensional Computer Vision, a Geometric Viewpoint*. MIT Press, 1993.
- [FB03] J-S. Franco and E. Boyer. Exact polyhedral visual hulls. *British Machine Vision Conference BMVC*, 1 :329–338, 2003.
- [FF62] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.
- [FHZa93] O. Faugeras, B. Hotz, Z.Zhan, and al. Real time correlation-based stereo : Algorithm, implementations and applications. Technical Report 2013, INRIA, 1993.
- [FL95] P. Fua and Y. Leclerc. Object-centered surface reconstruction : Combining multi-image stereo and shading. *International Journal on Computer Vision*, 16 :35–56, 1995.
- [FM90] O. Faugeras and S. Maybank. Motion from point matches : multiplicity of solutions. *International journal of Computer Vision*, pages 225–246, 1990.
- [Gar97] C. Garcia. Matis 3d : An adaptative stereo vision machine. *10th Scandinavian Conference on Image Analysis*, 1997.
- [GKR92] V. S. Gluth, G. W. Kunkel, and U. A. Rauhala. Global least squares matching. *Proceedings of International Geoscience and Remote Sensing Symposium*, 2 :1615–1618, 1992.
- [GLY95] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. *International Journal on Computer Vision (IJCV)*, 14 :211–226, 1995.
- [GMP98] V. Gouet, P. Montesinos, and D. Pele. Stereo matching of color images using differential invariants. *IEEE Int'l Conf. Image Processing*, pages 152–156, 1998.
- [Gri81] W. E. L. Grimson. A computational study of the human early visual system : From images to surfaces. *MIT Press*, 1981.
- [HA89] W. A. Hoff and N. Ahuja. Surfaces from stereo : Integrating feature matching, disparity estimation, and contour detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2) :121–136, 1989.
- [Han84] M. J. Hannah. Description of sri's baseline stereo system. *Technical Report SRI International*, 342, October 1984.

- [Han89] M. J. Hannah. A system for digital stereo image matching. *Photogrammetric Engineering and Remote Sensing*, 55 :1765–1770, 1989.
- [Hir01] H. Hirschmuller. Improvements in real-time correlation-based stereovision. *Proceedings of Workshop on Stereo and Multi-Baseline Vision*, pages 141–148, 2001.
- [HM95] R. Horaud and O. Monga. *Vision par ordinateur : outils fondamentaux*. Hermes, 1995.
- [Hor86] B. K. P. Horn. *Robot Vision*. MIT Press, 1986.
- [Hou93] S. Houzelle. *Extraction Automatique d’Objets Cartographiques par Fusion d’Informations Extraites d’Images Satelites*. PhD thesis, Ecole Nationale Supérieure des Telecommunications, 1993.
- [HS81] B. K. P. Horn and B. G. Schunk. Determining optical flow. *Artificial Intelligence*, 17 :185–204, 1981.
- [HS88] C. Harris and M. Stephens. A combined corner and edge detector. *Proceedings of the 4th ALVEY vision conference*, pages 147–151, 1988.
- [HZ00] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [IG98] H. Ishikawa and D. Geiger. Occlusions, discontinuities and epipolar lines in stereo. *Proceedings of 5th European Conference on Computer Vision*, pages 232–246, 1998.
- [II86] M. Ito and A. Ishii. Three-view stereo analysis. *Pattern Analysis and Machine Intelligence*, pages 524–532, 1986.
- [JKDW01] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*. Springer-Verlag, London, 2001.
- [JR70] E. G. Johnston and A. Rosenfeld. Geometrical operations on digitalized pictures. In *PPP70*, 1970.
- [Jul60] B. Julesz. Binocular depth perception of computer-generated pattern. *Bell System Tech. J.*, 39 :1125–1162, 1960.
- [Jul62] B. Julesz. Towards the automation of binocular depth perception. *IFIPS Congress*, 1962.
- [KK01] S. H. Kim and B. K. Kim. A noise model of postures measured by off-the-shelf vision. *Proceedings of International Conference on Computational Intelligence, Robotics and Autonomous System*, 2001.
- [KKCP96] D. H. Kim, K. B. Koo, W. Y. Choi, and R. H. Park. Stereo matching using hierarchical features for robotic applications. *Advanced Robotics*, 10 :1–14, 1996.
- [KO94] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window : Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9) :920–932, 1994.

- [Kol84] I. Kollar. Statistical theory of quantization : Results and limits. *Periodica Polytechnica Ser. Electrical Engineering*, 28(2-3) :173–189, 1984.
- [KOY00] M. Kumano, A. Ohya, and S. Yuta. Obstacle avoidance of autonomus mobile robot using stereo sensor. *Second International Symposium on Robotic and Automation ISRA*, pages 497–502, 2000.
- [KPKP89] B. Kamar-Parsi and B. Kamgar-Parsi. Evaluation of quantization error in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(9) :929–940, 1989.
- [KZ01] V. Kolmogorov and R. Zabih. Computing visual correspondance with occlusions using graph cuts. *Proceedings of the 8th International Conference on Computer Vision*, pages 508–515, 2001.
- [KZ02a] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. *Proceedings of European Conference on Computer Vision*, pages 82–96, 2002.
- [KZ02b] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *Proceedings of the European Conference on Computer Vision*, pages 65–81, 2002.
- [Las96] P. Lasserre. *Vision pour la Robotique Mobile en Environnement Naturel*. PhD thesis, LAAS, 1996.
- [Las99] V. Lasserre. *Modelisation Floue des Incertitudes de Mesures de Capteurs*. PhD thesis, Universite de Savoie, 1999.
- [LBL01] W.-Y. V. Leung, P. J. Bones, and R. G. Lane. Statistical interpolation of sampled images. *Optical Engineering (SPIE)*, 40(4) :547–553, April 2001.
- [LF93] Q. T. Luong and O. Faugeras. Self-calibration of a stereo rig from unknown camera motions and point correspondances. Technical Report 2014, INRIA, Juillet 1993.
- [LF97] Q. T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondances and fundamental matrices. *International Journal of Computer Vision*, 22(3) :261–289, 1997.
- [LH81] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293 :133–135, 1981.
- [LK81] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereovision. *Int'l Joint Conf. Artificial Intelligence*, pages 674–679, 1981.
- [LQ02] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8) :1140–1146, 2002.
- [Lux85] A. Lux. *Algorithme et Controle en Vision par Ordinateur*. PhD thesis, INPG Grenoble, 1985.
- [LZ99] P. Lepinay and R. Zapata. Realisation d'une carte de correlation visuelle en temp reel. Technical Report 6578, LIRMM, 1999.

- [MDG03] M.Z.Brown, D.Burschka, and G.D.Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8) :993–1008, 2003.
- [MH80] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings Royal Society London*, 207 :187–217, 1980.
- [ML00] D. Murray and J. Little. Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 8(2) :161–171, 2000.
- [ML04] D. Murray and J. Little. Environment modeling with stereo vision. *Intelligent Robots and Systems*, pages 3116–3122, 2004.
- [MM98] A. R. Mansouri and A. Mitiche. Selective image diffusion : Application to disparity estimation. *IEEE Int'l Conf. Image Processing*, 3 :284–288, 1998.
- [Moo66] R. Moore. *Interval Analysis*. Prentice-Hall, 1966.
- [MP77] D. Marr and T. Poggio. A theory of human stereo vision. *A. I. Memo Artificial Intelligence Lab, M. I. T.*, 451, November 1977.
- [MP79] D. Marr and T. Poggio. A computational theory of human stereo vision. *Royal Society of London*, B204, 1979.
- [Nov98] V. Novak. Fuzzy logic. in *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 1 :75–109, 1998.
- [OK85] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7 :139–154, 1985.
- [OKO01] M. Okutomi, Y. Katayama, and S. Oka. A simple stereo algorithm to recover precise object boundaries and smooth surfaces. *Proceedings of Workshop on Stereo and Multi-Baseline Vision*, pages 158–165, 2001.
- [Oli00] Musse Olivier. *Contribution a la mise en correspondance non rigide d'images medicales : une approche parametrique hierarchique sous contraintes topologique*. PhD thesis, ISIS Strasbourg, 2000.
- [Ous03] M. Oussalah. Introduction to fusion based systems - technical contributions of software computing and application to robotics. *Autonomous Robotic Systems : Software Computing and Hardware Computing Methodologies and Applications*, pages 35–71, 2003.
- [PS02] S. Paris and F. Sillion. Optimisation a base de flot de graphe pour l'acquisition d'informations 3d a partir de sequences d'images. *Actes des 15emes journées de l'AFIG*, 2002.
- [RC98] S. Roy and I. J. Cox. A maximun-flow formulation of the n-camera stereo correspondance probleme. *Proceedings of 6th International Conference on Computer Vision*, pages 492–499, 1998.
- [RG91] S. Randriamasy and Gagalowicz. Region based stereo matching oriented image processing. *Proceedings of Computer Vision and Pattern Recognition*, pages 736–737, 1991.

- [Rob65] L. G. Roberts. Machine perception for three-dimensional solids. *Optical and Electro-optical Information*, pages 159–197, 1965.
- [Roy99] S. Roy. Stereo without epipolar lines : A maximum-flow formulation. *International Journal of Computer Vision*, 34(2) :147–161, 1999.
- [Sar02] R. Sara. Finding the largest unambiguous component of stereo matching. *Proceedings 7th European Conference on Computer Vision*, 3 :900–914, 2002.
- [SB97] R. Sara and R. Bajcsy. On occluding contour artifacts in stereo vision. *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pages 852–857, 1997.
- [SC92] J. Shen and S. Castan. An optimal linear operator for step edge detection. *Computer Vision, Graphics and Image Processing*, 54(2) :112–133, March 1992.
- [SC02] O. Strauss and F. Comby. Modal estimation using rough histograms. *LFA*, pages 35–42, 2002.
- [Sch94a] D. Scharstein. A gradient-based evidence measure for image matching. Technical Report 94-1439, Cornell CS, 1994.
- [Sch94b] D. Scharstein. Matching images by comparing their gradient fields. *Proceedings of 12th IAPR International Conference on Pattern Recognition*, pages 630–639, November 1994.
- [Sch96] C. Schmid. *Appariement d'images par invariants locaux de niveau de gris*. PhD thesis, INPG, 1996.
- [Sch01] A. G. Schilling. Sampling problems in computer graphics. Habilitationsschrift zur Vorlage bei der Fakultät für Informatik der Eberhard-Karls-Universität Tübingen, October 2001.
- [SCM⁺04] G. G. Slabaugh, W. B. Culbertson, T. Malzbender, M. R. Stevens, and R. W. Schafer. Methods for volumetric reconstruction of visual scenes. *International Journal of Computer Vision*, 57(3) :179–199, 2004.
- [SDJ96] P. Sayd, M. Dhome, and J.M.Lavest. Axial stereovision for recovering straight cylinders. *IEEE Conference on Robotics and Cybernetics*, pages 560–564, 1996.
- [Sha76] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [She04] S. Shen. Image registration by hierarchical matching of local intensity histograms. *MICCAI*, September 2004.
- [Sil86] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*, volume Monographs on Statistics and Applied Probability 26. Chapman and Hall/crc, 1986.
- [Sme94] P. Smets. The transferable belief model. *Artificial Intelligence*, 66 :191–243, 1994.

- [Sme00] P. Smets. Constructing the pignistic probability function in a context of uncertainty. *in M. Henrion, R.D. Schacter, L.N. Kanal and J.F. Lemmers, Uncertainty in Artificial Intelligence*, 5 :29–39, 2000.
- [SS98] D. Scharstein and R. Szeliski. Stereo matching with non-linear diffusion. *Int'l Joint Conf. Artificial Intelligence*, 28 :155–174, 1998.
- [SS02] R. Szeliski and D. Scharstein. Symmetric sub-pixel stereo matching. *Proceedings 7th European Conference on Computer Vision*, 2 :525–540, 2002.
- [SS03] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. *IEEE Conference on Computer Vision and Pattern Recognition*, 1 :195–202, 2003.
- [SSZ02] J. Sun, H.Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. *European Conf. Computer Vision*, pages 510–524, 2002.
- [ST94] J. Shi and C. Tomasi. Good features to track. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.
- [Ste97] C. V. Stewart. Bias in robust estimation caused by discontinuous and multiple structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8) :818–833, 1997.
- [Sun98] C. Sun. Multi-resolution rectangular subregioning stereo matching using fast correlation and dynamic programming techniques. *CMIS Report*, 98(246), 1998.
- [SZ98] C. Schmid and A. Zisserman. The geometry and matching of curves in multiple views. *Proc. European Conf. Computer Vision*, pages 104–118, 1998.
- [Tel03] B. Telle. *Methode Ensembliste pour une Reconstruction Tridimensionnelle Garantie par Stereovision*. PhD thesis, Université de Montpellier II, 2003.
- [TG00] T. Tuytelaars and L. V. Gool. Wide baseline stereo matching based on local, affinely invariant regions. *Proceedings of British Machine Vision Conference*, 2 :412–425, 2000.
- [Thi92] E. Thiel. Distances discrettes. *Discret Geometry for Computer Image*, pages 69–78, 1992.
- [TM96] C. Tomasi and R. Manduchi. Stereo without search. In *ECCV*, volume 1, pages 452–465, 1996.
- [TSK01] H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. *Proceedings of the 8th International Conference on Computer Vision*, pages 532–539, 2001.
- [VC95] V. Venkateswar and R. Chellappa. Hierarchical stereo and motion correspondance. *International Journey of Computer Vision*, 15 :245–269, 1995.
- [Vek99] O. Veksler. *Efficient Graph-Based Energy Minimization Methods in Computer Vision*. PhD thesis, Cornell University, 1999.

- [VL01] E. Vincent and R. Laganier. Matching feature points en stereo pairs : A comparative study of some strategies. *Machine Graphics and Vision*, 10(3) :237–259, 2001.
- [WBH98] G-Q Wei, W. Brauer, and G. Hirzinger. Intensity- and gradient-based stereo matching using hierarchical gaussian basis functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11) :1143–1160, 1998.
- [WH99] J. H. Wang and C. P. Hsiao. On disparity matching in stereo vision via a neural network framework. *Proc. Natl. Sci. Counc. ROC*, 23(5) :665–678, 1999.
- [Wil91] R. P. Wildes. Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8) :761–774, 1991.
- [WK96] B. Widrow and I. Kollar. Statistical theory of quantization. *IEEE Transactions on Instrumentation and Measurement*, 45(2) :353–361, 1996.
- [Xie89] M. Xie. *Contribution a la visio dynamique : Reconstruction d'objets 3D polyedriques par une camera mobile*. PhD thesis, Rennes, 1989.
- [XZ96] G. Xu and Z. Zhang. *Epipolar Geometry in Stereo, Motion and Object Recognition*, volume 6. Kluwer Academic, 1996.
- [YFK94] R. R. Yager, M. Fedrizzi, and J. Kacprzyk. *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, 1994.
- [YT96] Y. Yvinec and B. Thiesse. Estimation de disparites 2d par une approche en frequences dans le cadre d'une geometrie epipolaire inconnue. *10eme congres Reconnaissance des Formes et Intelligence Artificielle (AFCET)*, 1996.
- [Zad65] L. A. Zadeh. Fuzzy sets. *Informations and Control*, 8 :338–353, 1965.
- [Zad78] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 :3–28, 1978.
- [Zha93] Z. Zhang. Le probleme de la mise en correspondance : L'etat de l'art. *Raport de recherche INRIA*, 2146, 1993.
- [Zha98] Z. Zhang. Determining the epipolar geometry and its uncertainty : A review. *International Journal of Computer Vision*, 27(2) :161–198, 1998.
- [ZK00] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7) :675–684, 2000.
- [ZK02] Y. Zhang and C. Kambhamettu. Stereo matching with segmentation-based cooperation. *Proceedings 7th European Conference on Computer Vision*, 2 :556–571, 2002.
- [ZW94] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance. *Third European Computer Vision*, pages 150–158, 1994.

Une nouvelle approche de mise en correspondance stéréoscopique dense par méthodes possibilistes

Résumé : Ce manuscrit présente une approche novatrice du problème de la mise en correspondance dense de pixel dans les images stéréoscopiques. L'originalité de ce travail repose sur l'utilisation d'un ensemble de techniques de représentation de l'imprécis et de l'incertain permettant de minimiser l'a-priori dans le processus de mise en correspondance. Ces techniques s'appuient sur une représentation par une partition floue de l'imprécision de la mesure de luminance induite par l'échantillonnage spatial et sur une classification duale de l'information de luminance associée à chaque pixel. Nous avons associé, à cette représentation, une modélisation de la contrainte épipolaire propre aux capteurs stéréoscopiques étalonnés permettant de prendre en compte l'imprécision des paramètres d'étalonnage. Cette représentation permet d'associer, à chaque pixel d'une des images, un domaine pondéré discret des pixels de l'autre image pouvant lui correspondre géométriquement et chromatiquement.

Nous proposons ensuite de modifier ces distributions de possibilité en utilisant des contraintes supplémentaires, telles que l'ordre ou l'unicité, pour en augmenter la spécificité. Nous utilisons, pour ce faire, un processus itératif et coopératif bipolaire. C'est l'ensemble de cette démarche qui permet de minimiser l'influence de la représentation du problème sur les mesures d'appariements.

La validité de notre approche est établie par des expérimentations sur des images stéréoscopiques de synthèses et réelles dont on connaît la vérité terrain. Celles-ci permettent de tester le comportement de notre méthode vis-à-vis des variations d'illumination et de disparité.

Mots clefs : Mise en correspondance stéréoscopique dense, théorie des possibilités, processus bipolaire, imprécision.

A new approach of dense stereo matching using possibility theory

Abstract : This manuscript presents a new approach to dense stereo matching problem. One of the main originality of this work is related to the use of the fuzzy set theory to replace precise similarity measure by imprecise resemblance measure in the matching process. This new matching measure is more robust to sampling and quantization due to a new modelling of a local digital luminance measure called "fuzzy pixel". This modelling associates to each pixel of the image a fuzzy domain and a fuzzy classification of the luminance.

Our method associate to each pixel of the first image a fuzzy domain of the second image based on luminance and epipolar information. Each domain is modified by an iterative and cooperative process in order to provide new domains that respect geometric constraint like ordering and uniqueness.

The validity of our approach is illustrated with experimentations on synthetic and real stereoscopic images.

Keywords : dense stereo matching, possibility theory, bipolar process, imprecision.
