



HAL
open science

Mise en correspondance stéréoscopique d'images couleur en présence d'occultations

Sylvie Chambon

► **To cite this version:**

Sylvie Chambon. Mise en correspondance stéréoscopique d'images couleur en présence d'occultations. Interface homme-machine [cs.HC]. Université Paul Sabatier - Toulouse III, 2005. Français. NNT : . tel-00011437

HAL Id: tel-00011437

<https://theses.hal.science/tel-00011437>

Submitted on 20 Jan 2006

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.

Mise en correspondance stéréoscopique d'images couleur en présence d'occultations

THÈSE

présentée et soutenue publiquement le 14 décembre 2005

pour l'obtention du

Doctorat de l'Université Paul Sabatier – Toulouse III

spécialité Informatique

par

Sylvie CHAMBON

Composition du jury

<i>Directeur de thèse :</i>	M. Alain CROUZIL	Maître de conférences UPS, Toulouse
<i>Rapporteurs :</i>	M. Xavier DESCOMBES	Chargé de recherche INRIA, INRIA Sophia-Antipolis
	M. Frédéric JURIE	Chargé de recherche CNRS, INRIA Rhône-Alpes
<i>Examineurs :</i>	M. Patrice DALLE	Professeur UPS, Toulouse
	M. Jean-José ORTEU	Professeur École des Mines, Albi
	M. Jean-Yves TOURNERET	Professeur INP, Toulouse

Résumé

Cette thèse se situe dans le cadre de la vision par ordinateur et concerne plus précisément l'étape de mise en correspondance de pixels en stéréovision binoculaire. Cette étape consiste à retrouver les pixels homologues dans deux images d'une même scène, prises de deux points de vue différents. Une des manières de réaliser la mise en correspondance est de faire appel à des mesures de corrélation. Les algorithmes utilisés se heurtent alors aux difficultés suivantes : les changements de luminosité, les bruits, les raccourcissements, les zones peu texturées et les occultations. Les travaux qui ont été réalisés sont une étude sur les méthodes à base de corrélation, en prenant en compte le problème des occultations et l'utilisation d'images couleur.

Dans un premier chapitre, nous établissons un état de l'art des méthodes de mise en correspondance de pixels. Nous donnons un modèle générique des méthodes s'appuyant sur la définition d'éléments constituants. Nous distinguons alors quatre catégories de méthodes : les méthodes locales, les méthodes globales, les méthodes mixtes et les méthodes à multiples passages.

Le second chapitre aborde le problème de l'évaluation des méthodes de mise en correspondance de pixels. Après avoir donné un état de l'art des protocoles existants, nous proposons un protocole d'évaluation et de comparaison qui prend en compte des images avec vérité terrain et qui distingue différentes zones d'occultations.

Dans le troisième chapitre, nous proposons une taxonomie des mesures de corrélation regroupées en cinq familles : les mesures de corrélation croisée, les mesures utilisant des outils des statistiques classiques, les mesures utilisant les dérivées des images, les mesures s'appuyant sur des outils des statistiques non paramétriques et les mesures exploitant des outils des statistiques robustes. Parmi cette dernière famille, nous proposons dix-sept mesures. Les résultats obtenus avec notre protocole montrent que ces mesures obtiennent les meilleurs résultats dans les zones d'occultations.

Le quatrième chapitre concerne la généralisation à la couleur des méthodes de mise en correspondance à base de corrélation. Après avoir présenté les systèmes de représentation de la couleur que nous testons, nous abordons la généralisation des méthodes à base de corrélation en passant par l'adaptation des mesures de corrélation à la couleur. Nous proposons trois méthodes différentes : fusion des résultats sur chaque composante, utilisation d'une analyse en composante principale et utilisation d'une mesure de corrélation couleur. Les résultats obtenus avec notre protocole mettent en évidence la meilleure méthode qui consiste à fusionner les scores de corrélation.

Dans le dernier chapitre, pour prendre en compte les occultations, nous proposons des méthodes hybrides qui s'appuient sur l'utilisation de deux mesures de corrélation : une mesure classique dans les zones sans occultation et une mesure robuste dans les zones d'occultations. Nous distinguons quatre types de méthodes à base de détection de contours, de corrélation pondérée, de post-détection des occultations et de fusion de cartes de disparités. Les résultats obtenus avec notre protocole montrent que la méthode la plus performante consiste à fusionner deux cartes de disparités.

Mots-clés : vision par ordinateur, stéréovision binoculaire, mise en correspondance, corrélation, protocole d'évaluation et de comparaison, occultations, couleur.

Abstract

This work deals with stereo-vision and more precisely matching of pixels using correlation measures. Matching is an important task in computer vision, the accuracy of the three-dimensional reconstruction depending on the accuracy of the matching. The problems of matching are: intensity distortions, noises, untextured areas, foreshortening and occlusions. Our research concerns matching color images and takes into account the problem of occlusions.

First, we distinguish the different elements that can compose a matching algorithm. This description allows us to introduce a classification of matching methods into four families: local methods, global methods, mixed methods and multi-pass methods.

Second, we set up an evaluation and comparison protocol based on fourteen image pairs, five evaluation areas and ten criteria. This protocol also provides disparity, ambiguity, inaccuracy and correct disparity maps. This protocol enables us to study the behavior of the methods we proposed.

Third, forty correlation measures are classified into five families: cross-correlation-based measures, classical statistics-based measures, derivative-based measures, non-parametric measures and robust measures. We also propose six new measures based on robust statistics. The results show us the most robust measures near occlusions: the robust measures including the six new measures.

Fourth, we propose to generalize dense correlation-based matching to color by choosing a color system and by generalizing the correlation measures to color. Ten color systems have been evaluated and three different methods have been compared: to compute the correlation with each color component and then to merge the results; to process a principal component analysis and then to compute the correlation with the first principal component; to compute the correlation directly with colors. We can conclude that the fusion method is the best.

Finally, in order to take into account the problem of occlusions, we present new algorithms that use two correlation measures: a classic measure in non-occluded area and a robust measure in the whole occlusion area. We introduce four different methods: edge detection methods, weighted correlation methods, post-detection methods and fusion method. This latter method is the most efficient.

Keywords: computer vision, stereoscopic matching, correlation, evaluation and comparison protocol, occlusions, color.

À mes parents

On ne voit bien qu'avec le coeur, l'essentiel est invisible pour les yeux.
[De Saint-Exupéry 43]

À Ivan

J'ai sur le bout de la langue
Ton petit coeur qui tangué.
Il m'a fait découvrir des trésors
Qui m'ont fait aimé encore
... Un peu plus.
[Bazbaz 04, Séchan 91]

Remerciements

Je tiens à remercier Xavier Descombes et Frédéric Jurie qui ont accepté d'être rapporteurs et dont les remarques positives et constructives m'ont encouragée à terminer ce travail plus sereinement.

Je remercie sincèrement Jean-José Orteu et Jean-Yves Tourneret qui ont bien voulu, malgré un emploi du temps chargé, être examinateurs.

Je souhaite remercier Patrice Dalle de m'avoir accueillie au sein de l'équipe TCI, d'avoir toujours pris le temps de répondre à mes questions et d'avoir accepté d'être examinateur.

Merci à Alain Crouzil d'avoir accepté d'être mon directeur de thèse. Grâce à ses qualités humaines et professionnelles, j'ai pu mener ces travaux de recherche dans les meilleures conditions possibles. Je n'oublie pas les encouragements tant au niveau personnel que professionnel. Merci Alain d'avoir su créer une ambiance chaleureuse et confiante, de m'avoir encadrée de manière rigoureuse en répondant patiemment à toutes mes questions.

Merci mille fois à Pierre Gurdjos pour mille raisons, entre autres, pour tous ses conseils qui m'ont permis d'y voir plus clair et plus loin dans mon avenir, et pour son aide précieuse à chaque rédaction d'article. Merci d'avoir compris que la réglisse est une des meilleures substances pour me stimuler, le comique de répétition un des meilleurs moyens pour me détendre, et la musique et le cinéma mes deux sujets de conversation préférés. Enfin, merci d'avoir pris en main l'aspect musical de cette thèse.

Merci aussi à Jean-Denis Durou pour ses encouragements durant tout l'été et toutes ses remarques précieuses à chaque présentation.

Merci à Frédéric Courteille qui m'a toujours apporté de l'aide quand j'en avais besoin et merci à Benoît Bocquillon d'avoir courageusement travaillé avec moi pendant son stage de DEA et de m'avoir gentiment encouragée dans les derniers moments de la rédaction.

Un grand merci à l'équipe TCI pour son soutien et son accueil, en particulier Fatima et Anne-Marie. Merci beaucoup Anne-Marie pour toutes ces petites conversations qui m'ont permis de me détendre pendant les pauses thé.

Merci beaucoup à Michèle Rommens d'avoir toujours tout mis en œuvre pour trouver tous les documents que je lui ai demandés depuis quatre ans.

Je souhaite aussi remercier Agathe Baritaud et Françoise Agar pour leur accueil chaleureux à chacune de mes visites.

Merci à Jean-Pierre Baritaud pour sa compréhension et sa bonne humeur communicative à chaque fois que je suis allée demander de l'aide.

Merci beaucoup à Jean-Pierre Ceccato d'avoir eu la patience d'imprimer et de relier ce document même si c'était toujours au dernier moment et avec des exigences particulières.

Merci à Martine Labruyère qui m'a toujours accueillie avec le sourire pour répondre patiemment à toutes mes questions.

Je tiens à remercier toutes les personnes qui ont amicalement partagé les bons et les mauvais moments de mes pauses café, thé et déjeuner : Cédric, Frédéric, José, Jérôme, Jean-Luc, Philippe, Tiago, Sihem et Ivan (du Brésil !). Un remerciement tout particulier à Marco polo qui m'a permis de m'évader à chacun de ses nouveaux récits. Je tiens aussi à remercier toutes les personnes qui m'ont encouragée à prendre l'air, manger et faire une pause ces derniers mois : Amélie, Anne, Florence, Nissou, Anne-Marie, Alain et Françoise, Benoît, Pierre, Lucille et Marie.

Merci à Gaël pour son amitié, sa gentillesse et sa franchise. Malgré mes sursauts, ça a toujours été un plaisir de te voir ouvrir avec entrain la porte de la salle des machines. Encore merci pour ta très précieuse et ta très courageuse relecture. Et puis merci aussi à Laure de m'avoir encouragée.

Merci à mon quatuor de femmes fortes : Amélie, Anne, Florence et Nissou, pour leur humour, leur bonne humeur communicative et leurs petites attentions.

Merci à ma petite Karine d'avoir toujours gardé le sourire pour aborder les sujets les plus sérieux comme les plus drôles. Merci surtout pour son amitié généreuse, merci d'être la seule personne à me proposer d'aller voir les films que personne ne veut voir (enfin à part elle et moi) et de m'avoir présenté des personnes adorables : Robert, Isabelle et Nicolas.

Merci Nathalie d'être toujours là dans les bons comme dans les mauvais moments. Même si le temps nous manque souvent, j'apprécie toujours les moments que nous passons ensemble, toutes les deux ou tous les quatre.

Je remercie tous mes amis qui malgré la distance ou leurs propres soucis ont toujours pensé à moi : Audrey, Eric, Clarisse et Frédéric, Cécile et Jérôme, Steph et Mathieu. Merci aussi à ma petite Marie-Cécile.

Un grand merci à Solenn et Sébastien pour la simplicité et la force de leur amitié et surtout d'avoir, au sens propre, mis la main à la pâte pour terminer cette thèse de manière agréable.

Merci à ma colocataire, amie et partenaire de voyage, d'avoir supporté mon absence et mes humeurs de ces premiers mois de cohabitation, de m'avoir encouragée et surtout d'être partie avec moi, avec autant d'enthousiasme, dans ma découverte de l'Italie.

Merci beaucoup à Marie d'être tout simplement une amie affectueuse, drôle et attentionnée.

Je remercie toute ma famille, pour l'amour et le soutien que vous m'avez toujours manifestés. Plein de petits mercis à mes chers Caroline et Alexandre, Nina et Marco, Laura et Margaux. Merci à mes parents, pour leur amour généreux et constant, et qui m'ont patiemment appris à relativiser dans les moments de stress. Merci aussi pour leur relecture attentive et leurs remarques judicieuses. Merci à ma petite Lulu pour son soutien quotidien, pour avoir égayé mes journées grâce à ses conseils déco, pour avoir elle aussi mis la main à la pâte et pour être tout simplement adorable. Merci encore Lulu pour toutes ces crêpes de ce long dimanche d'attente.

Enfin, plein de mercis pour ton incroyable générosité et ton immense tendresse, Ivan. Merci d'avoir perdu du temps, à me relire, à m'écouter et à préparer les cocktails ! Merci pour tous les soirs où je suis rentrée tard et où tu étais là pour me reconforter, que tu sois loin ou près de moi.

Table des matières

Table des figures	xviii
Liste des tableaux	xx
Liste des algorithmes	xxi
Notations	xxiii
Abréviations	xxvii
Introduction	1
1 Mise en correspondance stéréoscopique de pixels	9
1.1 Introduction	9
1.1.1 États de l’art existants sur la mise en correspondance stéréoscopique de pixels	9
1.1.2 Objectifs	10
1.1.3 Plan	10
1.2 Éléments constitutants de la mise en correspondance	10
1.3 Algorithme générique complet	12
1.4 Définitions et notations	13
1.4.1 Notations générales	13
1.4.2 Disparité	14
1.4.3 Fonction de disparité	14
1.5 Images et scènes traitées	14
1.6 Contraintes pour la mise en correspondance	15
1.6.1 Notion de contrainte	15
1.6.2 Contrainte épipolaire	15
1.6.3 Rectification épipolaire	16
1.6.4 Notations utilisées pour les contraintes de compatibilité	17
1.6.5 Contrainte d’unicité	18
1.6.6 Contrainte d’ordre	19
1.6.7 Contrainte de symétrie	19
1.6.8 Consistance faible	20
1.6.9 Synthèse et comparaison de quelques contraintes de compatibilité	21
1.6.10 Limite du gradient de disparité	22
1.6.11 Contrainte de rang	22
1.6.12 Contrainte de continuité figurale	23
1.6.13 Autres contraintes	23

1.7	Attributs des pixels	23
1.8	Prétraitement des images	24
1.8.1	Extraction des attributs des pixels	24
1.8.2	Segmentation en régions	25
1.9	Coûts de mise en correspondance, zones d'agrégation et support	25
1.9.1	Coût global de mise en correspondance	25
1.9.2	Méthodes locales – Méthodes par corrélation	27
1.9.3	Méthodes globales	29
1.9.4	Méthodes mixtes	31
1.10	Méthodes d'optimisation	34
1.10.1	Modèle des champs de Markov	35
1.10.2	Programmation dynamique	35
1.10.3	Recuit simulé	36
1.10.4	Relaxation	37
1.10.5	Coupage de graphe	38
1.10.6	Réseaux de neurones	39
1.10.7	Propagation de croyance	40
1.10.8	Algorithmes génétiques	41
1.11	Approche à passages multiples	41
1.11.1	Initialisation	41
1.11.2	Détermination automatique des paramètres des méthodes	42
1.11.3	Méthodes par croissance de germes	43
1.12	Affinement des résultats	44
1.12.1	Mise en correspondance au sous-pixel	44
1.12.2	Interpolation des disparités	44
1.12.3	Prise en compte d'erreurs d'appariement	45
1.13	Approche multirésolution	45
1.13.1	Construction de pyramides d'images	46
1.13.2	Propagation des résultats	47
1.14	Synthèse des méthodes de mise en correspondance	47
1.15	Conclusion	47
2	Protocole d'évaluation des méthodes de mise en correspondance stéréoscopique	49
2.1	Introduction	49
2.1.1	Évaluation des performances en vision par ordinateur	49
2.1.2	Caractéristiques d'un protocole d'évaluation	50
2.1.3	Objectifs	50
2.1.4	Plan	51
2.2	Définitions	51
2.2.1	Carte de disparités	51
2.2.2	Vérité terrain	51
2.3	Images testées	51
2.3.1	Images de synthèse	52
2.3.2	Images réelles sans vérité terrain	54
2.3.3	Images réelles avec vérité terrain	54
2.3.4	Résumé sur les images testées dans la littérature	59
2.4	Zones d'évaluation	60
2.5	Critères d'évaluation	60

2.5.1	Critères sans vérité terrain	60
2.5.2	Critères avec vérité terrain	61
2.6	Méthodes évaluées	62
2.7	Synthèse des critères et classement des méthodes	62
2.8	Protocole d'évaluation et de comparaison de Scharstein et Szeliski	63
2.8.1	Images testées	63
2.8.2	Zones d'évaluation	64
2.8.3	Critères d'évaluation	65
2.8.4	Synthèse des critères et classement des méthodes	65
2.8.5	Méthodes évaluées	66
2.9	Protocole d'évaluation et de comparaison proposé	68
2.9.1	Obtention de la vérité terrain	68
2.9.2	Obtention de nouvelles images avec vérité terrain	76
2.9.3	Images testées	77
2.9.4	Zones d'évaluation	78
2.9.5	Critères d'évaluation	79
2.9.6	Méthodes évaluées	80
2.9.7	Synthèse des critères et classement des méthodes	80
2.10	Conclusion	80
3	Mise en correspondance par corrélation des niveaux de gris – Corrélation 1D	81
3.1	Introduction	81
3.1.1	Avantages et inconvénients	81
3.1.2	Objectifs du chapitre	82
3.1.3	Plan	82
3.2	Présentation de la mise en correspondance par corrélation 1D	82
3.2.1	Notations	82
3.2.2	Définitions	84
3.2.3	Difficultés	85
3.2.4	Propriétés	86
3.3	Taxonomie des mesures 1D	87
3.3.1	Corrélation croisée	87
3.3.2	Statistiques classiques de la distribution des différences de niveaux de gris	88
3.3.3	Mesures utilisant les dérivées des images	91
3.3.4	Mesures non paramétriques	94
3.3.5	Mesures robustes	97
3.4	Synthèse des mesures	106
3.5	Techniques d'accélération des calculs	108
3.6	Résultats expérimentaux	109
3.6.1	Protocole d'évaluation	109
3.6.2	Présentation et analyse des résultats	112
3.7	Conclusion	116
4	Mise en correspondance par corrélation couleur – Corrélation 3D	125
4.1	Introduction	125
4.1.1	Objectifs	125
4.1.2	Stratégie de généralisation	125
4.1.3	Plan	126

4.2	Représentation de la couleur	126
4.2.1	Systèmes de primaires	126
4.2.2	Systèmes luminance-chrominance	128
4.2.3	Systèmes perceptuels	130
4.2.4	Systèmes d'axes indépendants	130
4.2.5	Systèmes hybrides	131
4.2.6	Synthèse des systèmes de représentation de la couleur	131
4.3	Généralisation des mesures de corrélation à la couleur	131
4.3.1	Notations	131
4.3.2	Généralisation d'opérateurs	131
4.3.3	Mesures 3D	135
4.4	Généralisation des méthodes de mise en correspondance à la couleur	139
4.4.1	Mise en correspondance sur chaque composante et fusion	139
4.4.2	Analyse en composantes principales	142
4.4.3	Mise en correspondance avec des mesures 3D	145
4.5	Résultats expérimentaux	145
4.5.1	Protocole d'évaluation	145
4.5.2	Présentation et analyse des résultats	148
4.6	Conclusion	150
5	Prise en compte des occultations	159
5.1	Introduction	159
5.1.1	Objectifs	159
5.1.2	Plan	160
5.2	Méthodes avec prétraitements	160
5.3	Méthodes locales	160
5.4	Méthodes globales	160
5.5	Méthodes locales à passages multiples	161
5.5.1	Fenêtres adaptatives	161
5.5.2	Mesures pondérées	162
5.5.3	Méthodes hybrides	163
5.6	Méthodes avec affinement	163
5.7	Méthodes hybrides utilisant deux mesures de corrélation	164
5.7.1	Zones des occultations considérées	164
5.7.2	Méthodes utilisant les contours	164
5.7.3	Méthodes utilisant le principe des mesures de corrélation pondérée	165
5.7.4	Méthodes avec post-détection	166
5.7.5	Méthodes de fusion de résultats	169
5.7.6	Récapitulatif des méthodes proposées	169
5.8	Résultats expérimentaux	169
5.8.1	Protocole d'évaluation	169
5.8.2	Présentation et analyse des résultats	172
5.9	Conclusion	176
	Conclusion	181
	Perspectives	185
	Annexe	187

Bibliographie

189

Table des figures

1	Dispositif faisant intervenir des techniques de vision par ordinateur.	1
2	Modèle géométrique de la caméra.	3
3	Modèle géométrique du capteur stéréoscopique binoculaire.	4
4	Problème des occultations.	5
5	Étude du voisinage.	6
6	Images couleur.	6
1.1	Géométrie épipolaire.	16
1.2	Contrainte d'unicité.	18
1.3	Contrainte d'ordre.	19
1.4	Contrainte de symétrie.	20
1.5	Comparaison des contraintes.	21
1.6	Contrainte de rang.	22
1.7	Programmation dynamique.	36
1.8	Technique des pixels distincts de Manduchi et Tomasi.	43
1.9	Multirésolution pour la mise en correspondance.	45
2.1	Génération d'un stéréogramme aléatoire.	52
2.2	Couples de synthèse.	53
2.3	Couples sans vérité terrain.	55
2.4	Couple avec vérité terrain obtenue manuellement.	56
2.5	Couples avec vérité terrain obtenue par vision active.	57
2.6	Couples avec vérité terrain obtenue à partir de contraintes géométriques.	59
2.7	Exemples de couples proposés par Scharstein et Szeliski.	64
2.8	Couples stéréoscopiques de synthèse proposés.	73
2.9	Vérité terrain théorique et vérité terrain calculée pour un couple d'images de synthèse.	74
2.10	Vérité terrain théorique et vérité terrain calculée pour des couples d'images réelles.	75
2.11	Couple stéréoscopique d'images réelles proposé.	77
2.12	Zones des occultations et zone des discontinuités considérées.	79
3.1	Zone de recherche et fenêtres de corrélation.	85
3.2	Mise en correspondance par corrélation 1D.	86
3.3	Principe des mesures robustes.	98
3.4	Graphe de visualisation des fonctions ρ_m des mesures utilisant des M-estimateurs.	104
3.5	Graphe de visualisation des fonctions J_m des mesures utilisant des R-estimateurs.	105
3.6	Boîtes filtrantes – Accélération des calculs grâce à la technique des boîtes filtrantes.	109
3.7	Mesure d'ambiguïté.	110
3.8	Mesure d'imprécision.	111
3.9	Cartes de résultats obtenus avec le stéréogramme aléatoire.	120

3.10	Cartes de résultats obtenus avec le couple d'images « murs ».	121
3.11	Cartes de résultats obtenus avec le couple d'images <i>cones</i> .	122
3.12	Cartes de résultats obtenus avec le couple d'images « livres ».	123
4.1	Méthode FUSION-SCORE.	140
4.2	Méthode FUSION-CARTE.	141
4.3	Méthode ACP-IMAGE.	143
4.4	Méthode ACP-CORRÉLATION.	144
4.5	Méthode DIRECTE.	146
4.6	Zones achromatiques et zones chromatiques dans les images testées.	147
4.7	Cartes de résultats obtenus avec le couple d'images « journaux ».	150
4.8	Cartes de résultats obtenus avec le couple d'images <i>barn1</i> .	152
4.9	Cartes de résultats obtenus avec le couple d'images <i>teddy</i> .	154
4.10	Cartes de résultats obtenus avec le couple d'images « livres ».	156
5.1	Méthode POST-DÉTECTION.	170
5.2	Cartes de résultats obtenus avec le couple d'images <i>tsukuba</i> .	176
5.3	Cartes de résultats obtenus avec le couple d'images <i>map</i> .	177
5.4	Cartes de résultats obtenus avec le couple d'images <i>cones</i> .	178
5.5	Cartes de résultats obtenus avec le couple d'images « livres ».	179

Liste des tableaux

1.1	Symboles utilisés dans les figures.	17
1.2	Récapitulatif sur les types de pixels.	24
1.3	Classement des méthodes de mise en correspondance.	48
2.1	Tableau récapitulatif des méthodes ayant participé au test de Scharstein et Szeliski.	67
2.2	Résultats obtenus pour les couples d'images testées.	74
3.1	Fonctions ρ_m des mesures utilisant des M-estimateurs.	103
3.2	Fonctions J_m des mesures utilisées par les R-estimateurs.	105
3.3	Famille CROISÉE.	106
3.4	Famille CLASSIQUE.	106
3.5	Famille DÉRIVÉE.	107
3.6	Famille NON PARAMÉTRIQUE.	107
3.7	Famille ROBUSTE.	108
3.8	Temps de calcul pour les différentes mesures de corrélation.	113
3.9	Variances des mesures sur les différentes scènes.	115
3.10	Résultats obtenus avec le stéréogramme aléatoire.	117
3.11	Résultats obtenus avec le couple d'images « murs ».	117
3.12	Résultats obtenus avec le couple d'images <i>cones</i>	118
3.13	Résultats obtenus avec le couple d'images « livres ».	118
3.14	Classement des mesures de corrélation.	119
4.1	Dix systèmes testés.	132
4.2	Famille CROISÉE couleur.	136
4.3	Famille CLASSIQUE couleur.	136
4.4	Famille DÉRIVÉE couleur.	137
4.5	Famille NON PARAMÉTRIQUE couleur.	137
4.6	Famille ROBUSTE couleur.	138
4.7	Classement des méthodes de mise en correspondance d'images couleur.	149
4.8	Résultats obtenus avec le couple d'images « journaux ».	151
4.9	Résultats obtenus avec le couple d'images <i>barn1</i>	153
4.10	Résultats obtenus avec le couple d'images <i>teddy</i>	155
4.11	Résultats obtenus avec le couple d'images « livres ».	157
5.1	Méthodes proposées pour prendre en compte le problème des occultations.	171
5.2	Classement des méthodes hybrides de mise en correspondance d'images couleur.	173
5.3	Résultats obtenus avec le couple d'images <i>tsukuba</i>	174
5.4	Résultats obtenus avec le couple d'images <i>map</i>	174

5.5	Résultats obtenus avec le couple d'images <i>cones</i>	175
5.6	Résultats obtenus avec le couple d'images « livres ».	175

Liste des algorithmes

1.1	Mise en correspondance stéréoscopique de pixels.	13
1.2	Méthode locale de mise en correspondance.	28
1.3	Méthode de mise en correspondance de Mayer.	28
1.4	Méthode globale de mise en correspondance.	29
1.5	Méthode de mise en correspondance de Sun <i>et al.</i>	32
1.6	Méthode de mise en correspondance de Hong et Chen.	33
1.7	Méthode générique d'optimisation pour la mise en correspondance.	34
1.8	Recuit simulé.	37
1.9	Relaxation.	38
2.1	Exploration de l'espace des polygones.	71
2.2	Recuit simulé pour l'estimation d'un polygone.	72
3.1	Mise en correspondance par corrélation 1D.	84
3.2	Algorithme minimal permettant de comparer les mesures de corrélation 1D.	112
4.1	Méthode FUSION-SCORE.	140
4.2	Méthode FUSION-CARTE.	141
4.3	Analyse en composantes principales.	142
4.4	Méthode ACP-IMAGE.	143
4.5	Méthode ACP-CORRÉLATION.	145
5.1	Mise en correspondance par corrélation avec des fenêtres adaptatives.	162
5.2	Méthode CONTOUR.	165
5.3	Méthode CORRÉLATION.	166
5.4	Méthode POST-DÉTECTION.	167

Notations

Nous ne donnons pas ici les notations qui sont utilisées de manière ponctuelle, c'est-à-dire seulement dans un paragraphe et qui ne sont plus employées ensuite. De plus, nous donnons, la plupart du temps, les notations dans l'ordre d'apparition dans le texte.

RÈGLES ET CONVENTIONS

x, N	Les scalaires sont toujours notés en caractères normaux.
\mathbf{v}	Les vecteurs sont notés en minuscules et en gras.
\mathbf{A}	Les matrices sont notées en majuscules et en gras.
\mathcal{E}	Les ensembles sont notés en lettres calligraphiées.
$l = g, d$	Les indices des images gauche et droite sont toujours notés de cette façon.
	Lorsqu'une barre verticale apparaît dans la marge, cela correspond à une définition.

NOTATIONS COMMUNES À TOUS LES CHAPITRES

I_l	Images
$I_l^{i,j}$	Niveau de gris du pixel de coordonnées $(i\ j)^T$
\mathbf{M}	Matrice de projection perspective
\mathbf{O}, \mathbf{O}_l	Centre optique de la caméra, centre optique de la caméra qui permet d'obtenir I_l
\mathbf{P}	Point 3D de la scène
\mathbf{p}_l	Point de I_l
$\mathbf{p}_l^{i,j}$	Pixel de I_l de coordonnées $(i\ j)^T$
$N_l^{\text{lig}}, N_l^{\text{col}}$	Nombre de lignes et nombre de colonnes de I_l
N_l	Nombre de pixels considérés (dont on cherche un correspondant) dans I_l .
N_z	Nombre de pixels dans la zone de recherche
N_f	Nombre de pixels dans la fenêtre de corrélation
N_v, N_h	Dimensions de la taille de fenêtre de corrélation avec $N_f = (2N_v + 1)(2N_h + 1)$
N_i	Nombre d'itérations
N_{OCM}	Nombre de niveaux possibles pour le codage utilisé dans la mesure OCM
t	Itération courante
\mathbf{A}^T	Transposée de \mathbf{A}

NOTATIONS COMMUNES À TOUS LES CHAPITRES (suite)

$\ \mathbf{v}\ $	Norme euclidienne du vecteur \mathbf{v}
$\text{card}(\mathcal{E})$	Nombre d'éléments de l'ensemble \mathcal{E}
d	Fonction de disparité
d_{cla}	Fonction de disparité associée à une méthode utilisant une mesure de corrélation classique
d_{rob}	Fonction de disparité associée à une méthode utilisant une mesure de corrélation robuste
d_{ref}	Fonction de disparité théorique
$\text{ZA}(\mathbf{p}_g^{i,j})$	Zone d'agrégation associée à $\mathbf{p}_g^{i,j}$ pour le coût local
$\text{ZAC}(\mathbf{p}_g^{i,j})$	Zone d'agrégation associée à $\mathbf{p}_g^{i,j}$ pour le coût des contraintes
$\text{Z}_d(\mathbf{p}_g^{i,j})$	Zone de recherche associée à $\mathbf{p}_g^{i,j}$
E_{global}	Coût global
$E_{\text{correspondance}}$	Coût d'attache aux données
$E_{\text{contrainte}}$	Coût des contraintes
E_{local}	Coût local
$E_{\text{dissimilarité}}$	Coût de dissimilarité
$E_{\text{voisinage}}$	Coût du voisinage
E_{lissage}	Coût de lissage
$f_{\text{dissimilarité}}$	Fonction de la différence des niveaux de gris
\mathcal{S}	Support du coût global
\mathcal{S}^i	Support du coût global associé aux méthodes globales
$\mathcal{S}^{i,j}$	Support du coût local associé aux méthodes locales
occ	Disparité attribuée lorsque le pixel est occulté
Prob	Probabilité
$\text{Err}_g^{i,j}$	Erreur de disparité associée au pixel $\mathbf{p}_g^{i,j}$
H	Matrice d'une homographie
\mathbf{m}_l	Coordonnées homogènes d'un point de l'image I_l
$\text{O}(\mathbf{p}_g^{i,j})$	Indique si $\mathbf{p}_g^{i,j}$ est un pixel occulté
$\text{PO}(\mathbf{p}_g^{i,j})$	Indique si $\mathbf{p}_g^{i,j}$ est un pixel proche d'une occultation
$\text{PD}(\mathbf{p}_g^{i,j})$	Indique si $\mathbf{p}_g^{i,j}$ est un pixel proche d'une discontinuité de profondeur
$\text{ZO}(I_l)$	Zone des occultations de I_l
$\text{ZI}(I_l)$	Zone d'influence des occultations de I_l
$\text{ZT}(I_l)$	Zone totale des occultations de I_l
$\text{ZD}(I_l)$	Zone des discontinuités de I_l
\mathbf{f}_l	Vecteur contenant les niveaux de gris des pixels des fenêtres de corrélation dans I_l
$\ \mathbf{f}_l\ _P$	P-normes
$\mathbf{f}_g \cdot \mathbf{f}_d$	Produit scalaire
$\bar{\mathbf{f}}_l$	Moyenne des valeurs de \mathbf{f}_l
$\text{var}(\mathbf{f}_l)$	Variance des valeurs de \mathbf{f}_l
$(f_l)_{k:N_f-1}$	Élément de rang k dans le vecteur \mathbf{f}_l
I_{max}	Niveau de gris le plus grand dans les images

NOTATIONS COMMUNES À TOUS LES CHAPITRES (suite)

$\nabla I_l^{i,j}$	Vecteur gradient situé au pixel $(i\ j)^T$ de l'image I_l
$\ \nabla I_l^{i,j}\ $	Norme du gradient
$\theta_l^{i,j}$	Orientation du gradient
$[x]$	Partie entière de x
\otimes	Concaténation
sgn	Fonction signe
$\text{Mes}(\mathbf{f}_g, \mathbf{f}_d)$	Mesure de corrélation entre deux vecteurs
D_{Ham}	Distance de Hamming
D_{OCM}	Distance définie pour la mesure <i>orientation code matching</i>
\mathbf{b}_l	Vecteur des valeurs binaires utilisées par les mesures ISC et SCC
\mathbf{e}_l	Vecteur des valeurs binaires utilisées par la mesure SCC
$\mathbf{R}_O(\mathbf{f}_l)$	Vecteur des directions des gradients dans la fenêtre \mathbf{f}_l calculé avec un opérateur O
$\mathbf{R}_{\text{Sobel}}(\mathbf{f}_l)$	Vecteur des directions des gradients dans la fenêtre \mathbf{f}_l calculé avec l'opérateur de Sobel
$\mathbf{R}_{\text{Kirsch}}(\mathbf{f}_l)$	Vecteur des directions des gradients dans la fenêtre \mathbf{f}_l calculé avec l'opérateur de Kirsch
$\mathbf{R}_{\text{LoG}}(\mathbf{f}_l)$	Vecteur des valeurs binaires du laplacien dans la fenêtre \mathbf{f}_l
$\mathbf{R}_{\text{Roberts}}(\mathbf{f}_l)$	Vecteur des valeurs de l'image binaire issue de la convolution par l'opérateur de Roberts dans la fenêtre \mathbf{f}_l
$\mathbf{R}_{\text{Pratt}}(\mathbf{f}_l)$	Vecteur des valeurs du laplacien dans la fenêtre \mathbf{f}_l
$\mathbf{R}_{\text{quad}}(\mathbf{f}_l)$	Vecteur des valeurs obtenues après application de la transformation associée à la <i>quadrant correlation</i>
$\mathbf{R}_{\text{rank}}(\mathbf{f}_l)$	Vecteur obtenu après application de la transformation de rang sur \mathbf{f}_l
$\mathbf{R}_\tau(\mathbf{f}_l)$	Vecteur obtenu après application de la transformation de recensement sur \mathbf{f}_l
$\mathbf{R}_m(\mathbf{f}_g)$	Vecteur des valeurs pondérées dans le cas des mesures de corrélation partielle
$\rho_m, \rho_m^{\min}, \rho_m^{\max}$	Fonction objet des M-estimateurs et valeurs minimales et maximales atteintes par ces fonctions
$J_m, J_m^{\min}, J_m^{\max}$	Fonction des R-estimateurs et valeurs minimales et maximales atteintes par ces fonctions
$s^{j,v}$	Score de corrélation entre le pixel situé en colonne j et le pixel situé en colonne v
x, y, z	Composantes couleur d'un système de représentation de la couleur
$\mathbf{c}_l^{i,j}$	Couleur du pixel de coordonnées $(i\ j)^T$ dans I_l
\mathbf{F}_l	Matrice contenant les composantes couleur des pixels de la fenêtre de corrélation
∇I_l^k	Matrice contenant les vecteurs gradients de chaque composante situés au pixel k de la fenêtre de corrélation dans I_l
$\frac{\partial x_l^k}{\partial i}, \frac{\partial y_l^k}{\partial i}, \frac{\partial z_l^k}{\partial i}, \frac{\partial x_l^k}{\partial j}, \frac{\partial y_l^k}{\partial j}, \frac{\partial z_l^k}{\partial j}$	Dérivées partielles suivant les lignes et les colonnes de chaque composante

NOTATIONS COMMUNES À TOUS LES CHAPITRES (suite)

$\ \nabla I_l^k\ $	Norme du gradient couleur
θ_l^k	Orientation du gradient couleur
\oslash	Division d'Hadamard
$\mathbf{x}_l, \mathbf{y}_l, \mathbf{z}_l$	Vecteurs contenant respectivement toutes les composantes x_l^k, y_l^k et z_l^k de la fenêtre de corrélation

Abréviations

ABRÉVIATIONS RELATIVES AUX MÉTHODES DE MISE EN CORRESPONDANCE

TYP	Type de la méthode (locale ou globale)
L	Locale
G	Globale
OPT	Méthode d'optimisation
AG	Algorithme génétique
CG	Coupure de graphe
PM	Passages multiples avec croissance de germes
PC	Propagation de croyance
PD	Programmation dynamique
RE	Relaxation
RG	Recherche gloutonne
RN	Réseau de neurones
RS	Recuit simulé
AC	Algorithme coopératif
WTA	<i>Winner take all</i>
AFF	Affinement
SP	Sous-pixel
Sym	Symétrie
PRÉ	Prétraitement
Fil	Filtrage
Seg	Segmentation
MUL	Approche à multiples passages
COU	Couleur
OCC	Occultations
COR	Mesure de corrélation
R	Rang
RMS	<i>Root Mean Square Error</i>

ABRÉVIATIONS UTILISÉES DANS LES TABLEAUX DE RÉSULTATS

COR	Pourcentage d'appariements corrects
ACC	Pourcentage d'appariements acceptés
MAU	Pourcentage d'appariements mauvais
ERR	Pourcentage d'appariements erronés
FPOS	Pourcentage de faux positifs
FNEG	Pourcentage de faux négatifs

 ABRÉVIATIONS UTILISÉES DANS LES TABLEAUX DE RÉSULTATS (suite)

ZT	Pourcentage d'appariements corrects dans la zone totale des occultations
ZO	Pourcentage d'appariements corrects dans la zone des occultations
ZI	Pourcentage d'appariements corrects dans la zone d'influence des occultations
ZD	Pourcentage d'appariements corrects dans la zone des discontinuités
RT	Rang de la méthode parmi toutes les méthodes évaluées
RF	Rang d'une mesure de corrélation au sein de sa famille

 ABRÉVIATIONS UTILISÉES POUR CARACTÉRISER LES INVARIANCES DES MESURES DE CORRÉLATION

G	Invariance de type gain
B	Invariance de type biais
GB	Invariance de type gain et biais

 FAMILLE CROISÉE

NCC	Corrélation croisée normalisée
ZNCC	Corrélation croisée centrée normalisée
MOR	Mesure de Moravec

 FAMILLE CLASSIQUE

D_P	Distances
D_1, SAD	Somme des valeurs absolues des différences
D_2, SSD	Somme des carrés des différences
D_∞	Maximum des valeurs absolues des différences
ZD_P	Distances centrées
$ZD_1, ZSAD$	Somme des valeurs absolues des différences centrée
$ZD_2, ZSSD$	Somme des carrés des différences centrée
ND_P	Distances normalisées
$ND_2, NSSD$	Somme des carrés des différences normalisée
ZND_P	Distances centrées normalisées
$ZND_2, ZNSSD$	Somme des carrés des différences centrée normalisée
LD_P	Distances localement centrées
$LD_1, LSAD$	Somme des valeurs absolues des différences localement centrée
$LD_2, LSSD$	Somme des carrés des différences localement centrée
VD	Variance des différences
VAD_P	Variance des valeurs absolues des différences élevées à la puissance P
$VAD_1, VOAD$	Variance des valeurs absolues des différences
$VAD_2, VOSD$	Variance des carrés des différences
K_4	<i>Kurtosis</i>

FAMILLE DÉRIVÉE

SEO _P	Mesure de Seitz
SES ₁ , SES ₁	Mesure de Seitz utilisant l'opérateur de Sobel
SEK ₁ , SEK ₁	Mesure de Seitz utilisant l'opérateur de Kirsch
NIS	Mesure de Nishihara
NA ₁ , NA ₂	Mesures de Nack
PRATT	Mesure de Pratt
OCM	<i>Orientation code matching</i> , mesure d'Ullah
GC	Corrélation des champs de gradients

FAMILLE NON PARAMÉTRIQUE

χ^2	Mesure de Puzicha
JEFF	Mesure de Jeffrey
ISC	<i>Increment sign correlation</i> , mesure de Kaneko
SCC	<i>Selective Coefficient Correlation</i> , mesure de Kaneko
RANK _P	Mesure de rang
CENSUS	Mesure de recensement
α	Mesure ordinaire de Bhat et Nayar

FAMILLE ROBUSTE

RM _m	Corrélation partielle
ZNCC _{F1} , ZNCC _{F2}	Corrélations partielles de Zoghلامي et Faugeras
RZSSD, RZNCC	Corrélation partielle de Lan
QUAD	<i>Quadrant correlation</i> , mesure de Huber
ZNCC _R	Corrélation croisée centrée normalisée robuste
D _P	Pseudo-norme
MAD	Variance robuste
LMS	Moindre médiane des carrés, <i>least median of squares</i>
LMP _P	Moindre médiane des puissances, <i>least median of powers</i>
LTS	Moindres carrés tronqués, <i>least trimmed squares</i>
LTP _P	Moindres puissances tronquées, <i>least trimmed powers</i>
SMAD	Estimation robuste de la variance <i>Smooth Median Absolute Deviation</i>
SMPD _P	<i>Smooth Median Powered Deviation</i>
ME _m	M-estimateur
RE _m	R-estimateur

SYSTÈMES DE REPRÉSENTATION DE LA COULEUR

RGB	Système de primaires, <i>Red Green Blue</i>
XYZ	Système de primaires proposé par la CIE
CMY	Système de primaires proposé pour la télévision et les imprimantes
L*u*v* ou CIELUV	Système perceptuellement uniforme proposé par la CIE
L*a*b* ou CIELAB	Système perceptuellement uniforme proposé par la CIE
AC ₁ C ₂	Système antagoniste proposé par Swain et Ballard
YCh ₁ Ch ₂	Système luminance-chrominance proposé par Carron
HSI	Système perceptuel

 SYSTÈMES DE REPRÉSENTATION DE LA COULEUR (suite)

$I_1I_2I_3$	Système d'axes indépendants proposé par Ohta
$H_1H_2H_3$	Système d'axes indépendants proposé Braquelaire et Brun

 MÉTHODES COULEUR

FUSION-SCORE	Fusion des scores de corrélation
FUSION-CARTE	Fusion des cartes de disparités
ACP-CORRÉLATION	Application d'une ACP sur la fenêtre de corrélation
ACP-IMAGE	Application d'une ACP sur toute l'image
DIRECTE	Utilisation d'une mesure de corrélation 3D

 MÉTHODES HYBRIDES

CONTOUR-SOBEL	Utilisation du filtre de Sobel
CONTOUR-SDEF	Utilisation de la méthode SDEF
CORRÉLATION-KANEKO	Utilisation de la méthode de Kaneko
CORRÉLATION-LAN	Utilisation de la méthode de Lan
POST-SEUIL	Utilisation d'un seuil sur le score de corrélation
POST-AMBIGUÏTÉ	Utilisation d'un seuil sur l'ambiguïté
POST-SYMMÉTRIE	Utilisation de la contrainte de symétrie
POST-UNICITÉ	Utilisation de la contrainte d'unicité
POST-ORDRE	Utilisation de la contrainte d'ordre
POST-MÉDIAN	Utilisation de la contrainte de symétrie suivie d'un filtrage médian
POST-DILATATION	Utilisation de la contrainte de symétrie suivie d'une dilatation des zones des occultations
POST-DILATATION-CONTOUR	Utilisation de la contrainte de symétrie suivie d'une dilatation des zones des occultations conditionnée par les contours
POST-ORDRE-SYMMÉTRIE	Utilisation des contraintes d'ordre et de symétrie
POST-AMBIGUÏTÉ-SYMMÉTRIE	Utilisation d'un seuil sur l'ambiguïté et de la contrainte de symétrie
FUSION-DISP	Fusion des deux cartes de disparités

Introduction

De tout temps, l'homme a tenté d'élaborer des machines capables de reproduire et de faciliter son travail. L'évolution des techniques et des connaissances dans de nombreux domaines, comme la médecine, la physique, les mathématiques et plus récemment l'informatique, a permis d'inventer toutes sortes d'outils, du plus rudimentaire au robot le plus perfectionné. Les progrès de la médecine ont permis de mieux comprendre le système visuel humain et, depuis quelques décennies, l'homme cherche à reproduire par une machine les mécanismes du système visuel humain pour des applications aussi variées que l'aide au déplacement d'un robot, l'exploration et l'étude des organes humains en médecine, la surveillance vidéo ou l'étude comportementale des animaux. Cette discipline constitue ce que nous appelons la *vision par ordinateur*.

Vision par ordinateur

De manière plus théorique, le but de la vision par ordinateur est de reproduire le système visuel humain et, plus précisément, de trouver un modèle qui possède des propriétés proches de la perception visuelle humaine. Pour cela, il faut une scène, une source de lumière et un capteur. Un exemple de dispositif où des techniques de vision par ordinateur interviennent est illustré sur la figure 1.

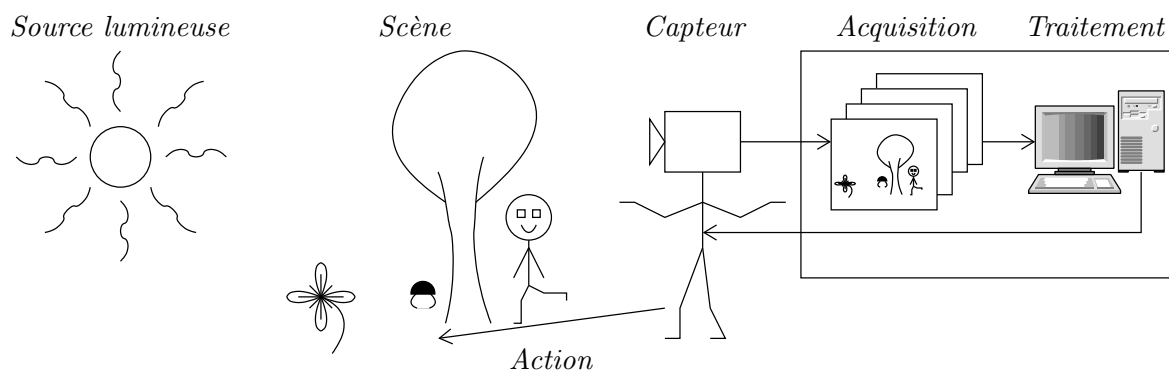


FIG. 1 – Dispositif faisant intervenir des techniques de vision par ordinateur.

Nous pouvons distinguer :

- *L'objectif* – Le robot doit aller cueillir la fleur ou le cèpe (en évitant si possible l'arbre et le personnage).
- *Le dispositif* – Il est constitué de la *scène* (une fleur, un cèpe, un arbre et un personnage en mouvement), de la *source lumineuse* (le soleil) et du *capteur* (la caméra du robot) qui permet de fournir des images de la scène qui sont analysées par un ordinateur.
- *Les étapes de traitement* – Elles comprennent : l'acquisition des images, le traitement des données (application d'un algorithme de vision par ordinateur) et la décision de l'action à réaliser.

- *Le résultat* – Si tout se déroule correctement, le robot va cueillir la fleur ou le cèpe en évitant l'arbre et le personnage.

La vision par ordinateur s'articule autour de trois objectifs principaux :

- *La reconnaissance des formes* – Avec l'exemple précédent, cela peut consister à distinguer la fleur et le cèpe.
- *L'analyse du mouvement* – Cela peut consister à détecter les mouvements du personnage.
- *La reconnaissance du relief* – Cela peut consister à déterminer que l'arbre est plus proche du robot que la fleur et le cèpe.

Pour tenter d'atteindre ces objectifs, différentes voies peuvent être suivies. Elles constituent des thèmes de recherche qui se distinguent par les caractéristiques du dispositif, par le type et le nombre d'images fournies par le capteur, par les informations disponibles sur la scène et sur le capteur et par la nature des résultats visés.

L'analyse du mouvement se caractérise par un dispositif dans lequel il existe un mouvement relatif entre le capteur et les objets de la scène. Des séquences d'images fournies par le capteur sont analysées afin d'extraire automatiquement des informations sur :

- *Le mouvement « perçu »* – Cela correspond à la projection du mouvement dans les images. On parle d'analyse du mouvement 2D qui peut, par exemple, passer par :
 - *Le calcul du flux optique* – On cherche à obtenir les champs des déplacements associés aux pixels des images de la séquence.
 - *Le suivi d'objets* – On cherche à repérer les positions d'un ou plusieurs objets dans chaque image de la séquence.
- *Le mouvement « réel »* – Cela correspond au mouvement 3D du capteur ou des objets de la scène.

La reconnaissance du relief concerne l'extraction automatique d'informations sur la structure 3D de la scène à partir d'une ou de plusieurs images. On parle alors de reconstruction 3D qui peut être réalisée à partir de stéréovision monoculaire, binoculaire ou multi-oculaire, suivant le nombre d'images disponibles.

Les travaux décrits dans ce mémoire se situent dans le cadre de la stéréovision binoculaire et concernent plus spécifiquement l'étape de mise en correspondance.

Stéréovision binoculaire

Nous pouvons considérer deux cas :

- Les images ont été acquises à deux instants différents (c'est le cas d'images obtenues depuis un satellite, à deux passages différents).
- Les images ont été acquises au même instant ou à des instants suffisamment proches pour qu'aucun changement ne se soit produit dans la scène.

Nous nous plaçons dans le second cas. Nous pouvons distinguer trois étapes fondamentales pour retrouver le relief de la scène :

- *Le calibrage* – Cela consiste à trouver les paramètres des capteurs.
- *La mise en correspondance* ou *appariement* – Le but est de retrouver les points homologues entre les deux images, c'est-à-dire les projections des mêmes points de la scène.
- *La reconstruction 3D* – À partir des paramètres des capteurs et des correspondances de points, un modèle 3D de la scène est reconstruit.

La mise en correspondance s'avère être une tâche délicate dont la qualité du résultat détermine directement la qualité de la reconstruction 3D. Elle rencontre de nombreuses difficultés, notamment en

présence de changements de luminosité entre les deux images, d'objets dont la surface est uniforme, d'occultations et de bruit dans les images. En raison de ces difficultés, les méthodes de mise en correspondance sont amenées à exploiter toutes les informations disponibles afin de faciliter la recherche et la détermination des correspondants. Des informations liées à des hypothèses concernant la formation des images sont utilisées pour réduire la taille des zones des images dans lesquelles les correspondants sont recherchés. Il s'agit d'informations qui découlent directement du modèle géométrique du capteur. C'est aussi sur ce modèle géométrique que repose l'étape de triangulation permettant la reconstruction 3D.

Modèle géométrique de la caméra

Le modèle *sténopé* (cf. figure 2), ou modèle du « trou d'aiguille » (*pinhole*), est constitué du plan image et du centre optique \mathbf{O} . Un point \mathbf{P} de la scène, de coordonnées (X, Y, Z) , se projette sur le plan image en un point \mathbf{p} , de coordonnées (u, v) , qui est l'intersection de la droite (\mathbf{OP}) avec le plan image. Cette projection peut être représentée par l'équation suivante :

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}^T = \mathbf{M} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}^T, \lambda \in \mathbb{R}^*, \quad (1)$$

où \mathbf{M} est la matrice de projection perspective qui contient les paramètres du modèle. Le *calibrage* d'une caméra peut consister à estimer cette matrice.

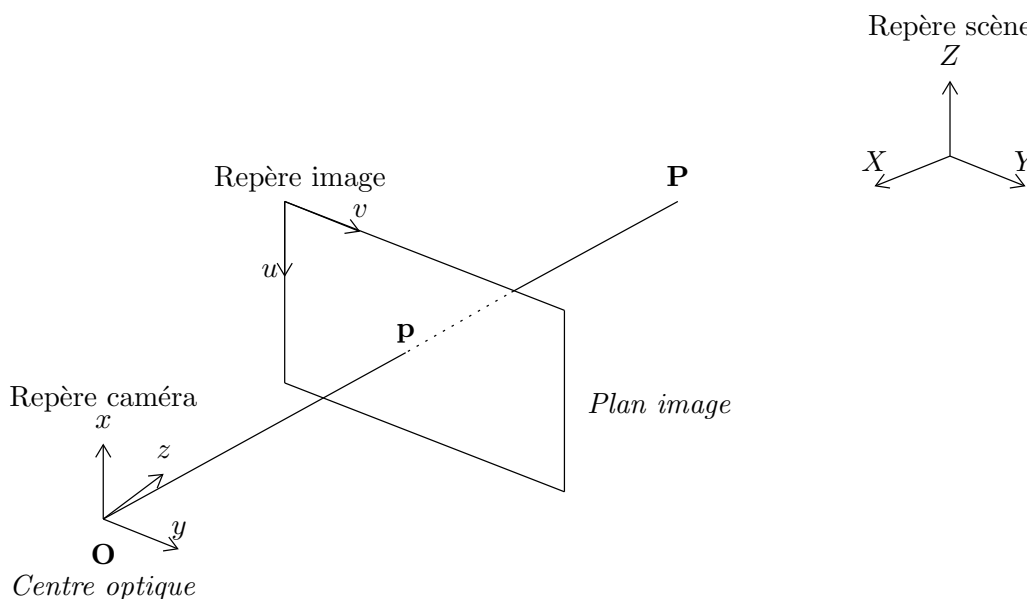


FIG. 2 – Modèle géométrique de la caméra.

Modèle du capteur stéréoscopique

Dans le cas particulier de la stéréovision binoculaire, deux images sont prises depuis deux points de vue différents. Pour cela, soit deux caméras sont utilisées, soit la caméra est déplacée entre les deux prises de vue (dans ce cas, les objets doivent être immobiles). La position relative des deux prises de vue doit être choisie de telle sorte qu'une grande partie de la scène soit visible dans les deux

images. La position de la deuxième prise de vue ne doit pas être obtenue par rotation autour du centre optique de la caméra, car, dans ce cas, la reconstruction 3D est impossible. Les paramètres du modèle géométrique du capteur stéréoscopique binoculaire (cf. figure 3) peuvent être représentés par les deux matrices de projection perspectives associées aux deux images.

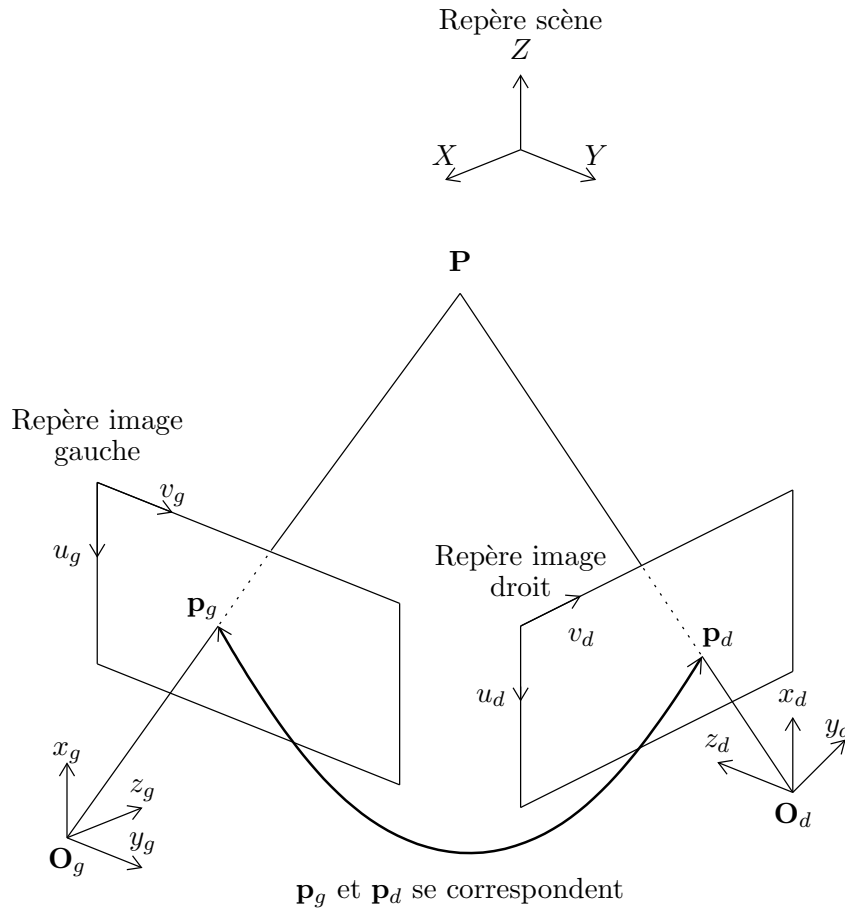


FIG. 3 – *Modèle géométrique du capteur stéréoscopique binoculaire.*

Mise en correspondance

La figure 3 montre qu'un point \mathbf{P} de la scène, visible dans les deux images se projette dans l'image gauche en \mathbf{p}_g de coordonnées (u_g, v_g) et dans l'image droite en \mathbf{p}_d de coordonnées (u_d, v_d) . On dit que les points \mathbf{p}_g et \mathbf{p}_d sont homologues, c'est-à-dire qu'ils se correspondent, car ils représentent le même point de la scène. La mise en correspondance consiste, à partir de deux images, que l'on appellera l'image gauche et l'image droite, à retrouver les couples de points qui se correspondent.

Si l'on dispose d'images de niveaux de gris, il paraît raisonnable de poser l'hypothèse suivante :

- | Si deux pixels se correspondent alors leurs niveaux de gris se ressemblent.
- Cependant, cette hypothèse n'étant pas assez discriminante, l'hypothèse suivante est très souvent ajoutée :
- | Si deux pixels se correspondent alors les niveaux de gris de leurs voisinages respectifs se ressemblent.
- Cette hypothèse est à la base des méthodes de *mise en correspondance par corrélation* qui sont au centre de notre travail. Bien qu'elles soient simples, leur efficacité a été plusieurs fois démontrée et

de nombreuses applications font intervenir une méthode de ce type. Nous pouvons mentionner par exemple :

- les travaux de Garcia *et al.* [Garcia 01b, Garcia 02, Orteu 03] qui étudient le comportement mécanique des matériaux ;
- les travaux de Lantagne *et al.* [Lantagne 03] consacrés au suivi de personnes dans un contexte de surveillance vidéo ;
- de nombreux travaux qui abordent la reconstruction de scènes urbaines à partir de MNE, Modèles Numériques d'Élévation, obtenus à partir d'images aériennes ou satellitaires [Ehlers 86, Cord 98, Baillard 00], notamment pour l'Institut Géographique National [Chehata 05].

Problématique

Les méthodes de mise en correspondance par corrélation reposent sur la ressemblance de deux ensembles de niveaux de gris qui est quantifiée grâce à une mesure de corrélation. De très nombreuses mesures de corrélation ont été proposées pour prendre en compte les différentes difficultés de la mise en correspondance, comme les bruits ou les occultations. Il s'avère donc très difficile de choisir une mesure de corrélation. Dans un premier temps, il nous a paru important d'effectuer un recensement, une description et une évaluation des mesures de corrélation existantes.

Par ailleurs, une des difficultés rencontrées par les méthodes de mise en correspondance concerne les occultations dont la figure 4 est une illustration.



FIG. 4 – Le bas de la boîte désigné par un point blanc dans l'image gauche, n'est plus visible dans l'image droite.

On parle d'occultations lorsque des parties de la scène ne sont visibles que dans l'une des deux images. Les occultations sont une source d'erreur lors de la mise en correspondance car, dans le voisinage d'une occultation, l'hypothèse des voisinages semblables n'est plus vérifiée (cf. figure 5).

Face à ce problème, nous pouvons nous poser les questions suivantes : quelle est l'influence des occultations sur le comportement des méthodes de mise en correspondance ? Comment prendre en compte cette difficulté ?

Par ailleurs, la plupart des capteurs actuels permettent l'acquisition d'images en couleur. Globalement, la couleur apporte une information supplémentaire, plus discriminante, comme nous pouvons le voir sur la figure 6.

Dans le cadre de la mise en correspondance stéréoscopique de pixels, on peut se demander dans quelle mesure l'utilisation de la couleur apporte un gain de performance. Mais cela entraîne les questions suivantes : quels systèmes de représentation de la couleur faut-il utiliser ? Comment adapter les méthodes à la couleur ?

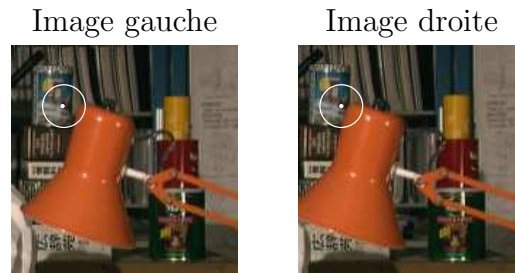


FIG. 5 – *Les voisinages de ces deux pixels correspondants, représentés par les deux cercles blancs, sont différents, ce qui rend leur comparaison difficile.*



FIG. 6 – *L'image en niveaux de gris ne nous permet pas de distinguer qu'il y a une différence de luminosité entre les lettres en haut du cône et les lettres en bas du cône. De même on ne distingue pas sur le morceau de tissu jaune, la différence de couleur entre les deux bandes imprimées.*

Enfin, nous sommes naturellement amenés à évaluer et à comparer différentes méthodes. Mais, la mise en place d'un protocole d'évaluation des performances soulève des questions concernant le choix des données et des critères.

Nous souhaitons tenter de répondre à ces questions. Pour cela, les objectifs des travaux présentés dans ce mémoire sont les suivants :

- Établir un état de l'art des méthodes de mise en correspondance stéréoscopique de pixels – Nous souhaitons définir clairement les différents éléments qui constituent les méthodes et fournir une classification des méthodes existantes en fonction de ces éléments.
- Évaluer et comparer les méthodes locales de mise en correspondance stéréoscopique de pixels – Notre but est de mettre en place un protocole d'évaluation et de comparaison rigoureux.
- Analyser la technique de mise en correspondance par corrélation – D'une part, nous souhaitons caractériser et classer les mesures de corrélation existantes et, d'autre part, nous avons pour objectif de proposer de nouvelles mesures.
- Prendre en compte la couleur – Nous désirons adapter les mesures de corrélation et les méthodes de mise en correspondance par corrélation à la couleur.
- Prendre en compte le problème des occultations – Nous souhaitons proposer de nouvelles méthodes de mise en correspondance qui s'appuient sur des mesures de corrélation robustes.

Organisation du mémoire

Ce mémoire s'articule autour de cinq chapitres.

- Chapitre 1 – État de l'art sur la mise en correspondance stéréoscopique de pixels – Nous donnons un modèle générique des méthodes de mise en correspondance s'appuyant sur la définition d'éléments constitutants. Notre état de l'art s'attache à distinguer les différentes méthodes par les instances de ces éléments constitutants. Il conduit à une classification en quatre catégories : les méthodes locales, les méthodes globales, les méthodes mixtes et les méthodes à multiples passages.
- Chapitre 2 – État de l'art et protocole d'évaluation et de comparaison de méthodes de mise en correspondance – Notre but étant d'obtenir un protocole d'évaluation précis et pertinent, nous commençons par donner un état de l'art sur les méthodes d'évaluation dans le cadre de la mise en correspondance de pixels, puis nous détaillons notre démarche pour obtenir de nouvelles images et pour choisir les différentes caractéristiques de notre protocole qui sera utilisé dans tous les chapitres suivants.
- Chapitre 3 – Mise en correspondance par corrélation – Après avoir défini la mise en correspondance par corrélation, nous présentons une taxonomie des mesures groupées en cinq familles : les mesures utilisant la corrélation croisée, les mesures utilisant des outils des statistiques classiques, les mesures s'appuyant sur les dérivées des images, les mesures exploitant des outils des statistiques non paramétriques et les mesures à base d'outils des statistiques robustes. Dans la dernière de ces familles, nous proposons des mesures qui prennent en compte le problème des occultations.
- Chapitre 4 – Mise en correspondance par corrélation couleur – Ce chapitre se divise en trois parties : le choix du système de représentation, la généralisation des mesures de corrélation et la généralisation des méthodes de mise en correspondance stéréoscopique par corrélation.
- Chapitre 5 – Mise en correspondance avec prise en compte des occultations – Nous utilisons les mesures proposées au chapitre 3 pour établir quatre nouveaux types de méthodes hybrides. Ces méthodes utilisent une mesure robuste dans les zones des occultations et une mesure classique dans les autres zones.

Chapitre 1

Mise en correspondance stéréoscopique de pixels

1.1 Introduction

De manière générale, la mise en correspondance stéréoscopique consiste à retrouver dans les images gauche et droite, les primitives homologues, c'est-à-dire, les primitives qui sont la projection de la même entité de la scène. Dans le cadre de nos travaux, les primitives que nous considérons sont les pixels des images.

Même si elles sortent *a priori* du cadre de notre étude, nous avons aussi considéré certaines méthodes de calcul du flux optique qui peuvent être utilisées pour faire de la mise en correspondance stéréoscopique de pixels. Aggarwal et Davis [Aggarwal 88] ont, à ce sujet, effectué un parallèle entre ces deux thèmes.

1.1.1 États de l'art existants sur la mise en correspondance stéréoscopique de pixels

Généralement, les méthodes de mise en correspondance sont décrites sous la forme d'un problème de minimisation d'une fonction d'énergie, le coût global, ou de plusieurs fonctions d'énergie, les coûts locaux. Des états de l'art ont été réalisés et nous citons les deux les plus aboutis :

- Scharstein et Szeliski [Scharstein 02] distinguent quatre éléments constitutants pour caractériser les méthodes de mise en correspondance stéréoscopique de pixels :
 - le coût local qui évalue le coût d'une correspondance ;
 - la zone d'agrégation qui correspond à l'ensemble des pixels pris en compte pour le calcul du coût local ;
 - la méthode d'optimisation ;
 - l'affinement des résultats.

Ces quatre éléments vont être détaillés au paragraphe 1.2. Les auteurs donnent un tableau récapitulatif de plus d'une trentaine de méthodes classées en quatre groupes :

- Les méthodes qui se différencient par leur coût (zone d'agrégation réduite au pixel étudié et même méthode d'optimisation utilisée, *winner take all*, une méthode de recherche exhaustive).
- Les méthodes qui se différencient par leur zone d'agrégation (même coût, différence des niveaux de gris au carré et même méthode d'optimisation, *winner take all*).

- Les méthodes qui se différencient par leur méthode d'optimisation (même coût et zone d'agrégation réduite au pixel étudié).
- Les méthodes qui ne rentrent dans aucune des trois catégories citées.
- Brown *et al.* [Brown 03] différencient :
 - les méthodes locales, en distinguant les méthodes par corrélation, les méthodes fondées sur l'équation du flux optique et les méthodes d'appariement de primitives structurées comme les segments ;
 - les méthodes globales en distinguant celles utilisant la programmation dynamique (avec des pixels ou avec des primitives structurées), les coupures de graphes et les méthodes probabilistes (avec des pixels).

Les auteurs distinguent aussi les méthodes par leur façon de détecter les occultations (nous nous intéresserons à cet aspect dans le chapitre 5).

1.1.2 Objectifs

Les objectifs de l'état de l'art que nous présentons dans ce chapitre sont de répertorier les approches qui peuvent être appliquées au cas de la mise en correspondance stéréoscopique de pixels et de les caractériser par leurs éléments constituants. Nous avons réalisé un état de l'art qui s'appuie sur ceux exposés dans [Scharstein 02] et [Brown 03] et les complète, c'est pourquoi :

- Nous avons repris la caractérisation en éléments constituants utilisée dans [Scharstein 02] à laquelle nous avons ajouté les éléments suivants : primitives et attributs, prétraitement, support, zone de recherche, passages multiples et approche multirésolution. De plus, nous abordons la notion de coût global et nous distinguons différentes zones d'agrégation.
- Même si dans [Brown 03], de nombreuses méthodes d'optimisation sont énumérées nous allons tenter d'être plus précis et plus exhaustif, notamment en évoquant les techniques de recuit simulé, les techniques de relaxation, les réseaux de neurones et les algorithmes génétiques.

1.1.3 Plan

Tout d'abord, nous énumérons les éléments constituants et nous abordons les définitions, les notations et l'algorithme générique complet pour définir le cadre de cet état de l'art sur la mise en correspondance stéréoscopique de pixels. Puis, nous énumérons brièvement le type d'images et le type de scènes qui ont été traitées dans les publications concernées par cet état de l'art. Par la suite, nous présentons les contraintes les plus employées dans le cadre de la mise en correspondance stéréoscopique de pixels. Ensuite, nous détaillons les différents éléments constituants : les attributs des pixels, le prétraitement, les différents coûts de mise en correspondance, les méthodes d'optimisation, les passages multiples, l'affinement des résultats et la multirésolution. Enfin, nous présentons une synthèse de cet état de l'art.

1.2 Éléments constituants de la mise en correspondance

Pour caractériser les méthodes de mise en correspondance, nous avons identifié les éléments constituants suivants :

- *Les primitives à apparier et leurs attributs* – Comme dans [Jones 97], nous distinguons deux catégories :
 - *Les pixels de l'image* – Nous pouvons, par exemple, considérer tous les pixels de l'image, c'est le cas dans ce mémoire, ou seulement une partie, comme des points d'intérêt. Leurs attributs sont la plupart du temps les niveaux de gris ou les composantes couleur.

- *Les primitives structurées* – Jawahar et Narayanan [Jawahar 02] ont réalisé une synthèse des algorithmes effectuant la mise en correspondance des primitives structurées, appelée *feature-based matching*. Nous pouvons distinguer :
 - *Les segments* – Les attributs généralement utilisés sont, entre autres, la position, l'orientation, la longueur, les niveaux de gris du segment, le contraste sur le segment et les relations avec les segments adjacents [Ayache 89, Horaud 89, Tao 01, Jurie 01].
 - *Les contours* – La plupart du temps, les niveaux de gris, les composantes couleur ou les gradients des pixels qui forment le contour sont utilisés [Petrakis 02].
 - *Les régions* – Les attributs utilisés peuvent être les niveaux de gris ou les couleurs de la région, comme dans [Matas 02], une mesure globale sur la région comme la variance des niveaux de gris utilisée dans [Han 01] ou des moments, cf. [Tuytelaars 00].

La plupart des publications sur la mise en correspondance de primitives structurées datent de la fin des années 1980. Toutefois, des publications récentes concernent la mise en correspondance de régions pour guider la mise en correspondance de pixels. L'utilisation de primitives structurées limite la combinatoire lors de la recherche des correspondances et offre la possibilité d'utiliser des attributs discriminants permettant de lever des ambiguïtés. En revanche, il est nécessaire de détecter au préalable ces primitives structurées. Les détecteurs utilisés doivent être suffisamment répétables pour extraire les primitives homologues des images. De plus, les primitives doivent être caractérisées par des attributs à la fois invariants et discriminants, ce qui est souvent contradictoire. Enfin, la mise en correspondance de primitives structurées ne permet pas d'obtenir une reconstruction dense de la scène.

- *Le prétraitement des images* – Un traitement des images peut être effectué avant de chercher les correspondances, notamment pour tenter de supprimer le bruit ou effectuer le calcul des attributs.
- *Le coût global de mise en correspondance* – Lorsqu'un ensemble de correspondances entre primitives est établi, il est nécessaire de déterminer à quel point cet ensemble de correspondances est en accord avec le modèle choisi, c'est-à-dire à quel point les hypothèses sont vérifiées et les contraintes sont satisfaites. Pour cela, un coût global de mise en correspondance est associé à l'ensemble des correspondances. Ce coût se compose de deux termes :
 - *Le coût de correspondance* – Il s'agit de la somme des coûts locaux de toutes les correspondances. Chaque coût local permet de déterminer la dissimilarité entre deux primitives correspondantes.
 - *Le coût des contraintes* – Il quantifie le degré de respect des contraintes prises en compte pour l'ensemble des correspondances. Il est constitué de la somme des coûts de voisinage. Chaque coût de voisinage prend en compte les correspondants attribués au voisinage de la primitive considérée.

Par exemple, dans le cas où les primitives sont des pixels et où les niveaux de gris sont leurs attributs, le coût local peut consister en une simple fonction de la différence de niveaux de gris et le coût de voisinage en la distance qui sépare le correspondant et le correspondant de l'un des pixels voisins. Certaines méthodes s'appuient sur un coût global réduit au coût local de mise en correspondance et sont appelées les méthodes locales. Les autres méthodes sont appelées les méthodes globales. Le coût global est généralement appelé énergie. Dans ce mémoire, nous préférons le terme « coût », mais nous utiliserons indifféremment les deux termes.

- *Les zones d'agrégation* – Elles concernent :
 - *Le coût local* – Des méthodes prennent en compte des primitives voisines aux primitives étudiées pour calculer le coût local. L'ensemble des primitives prises en compte constitue la zone d'agrégation.

- *Le coût de voisinage* – Dans la plupart des cas, la zone d’agrégation associée au coût des contraintes est différente de celle associée au coût local.
- *Le support* – C’est l’ensemble des primitives considérées au moment du calcul du coût global. Le support peut être, par exemple, dans le cas des méthodes globales l’ensemble des primitives de l’image ou les primitives qui se trouvent sur la même droite épipolaire (cf. paragraphe 1.6.2) et, dans le cas des méthodes locales, le singleton constitué du pixel étudié.
- *La zone de recherche* – Elle est associée à chaque primitive dont on cherche le correspondant et est constituée des primitives candidates.
- *La méthode d’optimisation* – Il s’agit de la méthode utilisée pour trouver la valeur minimale du coût global, dans le cas des méthodes globales, ou des coûts locaux, dans le cas des méthodes locales, et déterminer ainsi les couples de primitives retenues comme correspondantes.
- *L’affinement des résultats* – Après avoir établi les correspondances, les résultats peuvent aussi être traités, notamment pour tenter de corriger *a posteriori* d’éventuelles erreurs.
- *Les passages multiples* – Certaines méthodes effectuent plusieurs fois la mise en correspondance avec des méthodes différentes pour combiner les avantages de chacune ; c’est ce que nous appelons les méthodes à passages multiples. Par exemple, certaines méthodes effectuent une mise en correspondance locale, suivie d’une mise en correspondance globale.
- *La multirésolution* – Pour diminuer les temps de calcul, la mise en correspondance peut être réalisée selon une approche multirésolution : une pyramide d’images à différentes résolutions est construite, l’appariement est effectué au niveau le plus grossier et les résultats sont propagés jusqu’au niveau le plus fin.

Cette caractérisation nous semble convenir pour décrire la plupart des méthodes que nous citons dans ce chapitre. Cependant, il est parfois difficile de distinguer dans certains articles ce qui correspond au coût global de mise en correspondance et ce qui est propre à la méthode d’optimisation. Le choix de tous ces éléments constitutifs est largement influencé par le type d’images traitées (images à une ou plusieurs composantes) et le type de scènes considérées (scène d’extérieur, scène d’intérieur, scène polyédrique). Enfin, il faut noter que certains de ces éléments constitutifs sont essentiels. En effet, on ne peut pas définir une méthode de mise en correspondance sans : déterminer les primitives et leurs attributs, définir un coût global de mise en correspondance avec un support et des zones d’agrégation associées et choisir une méthode d’optimisation.

1.3 Algorithme générique complet

Chacun des éléments constitutifs que nous avons présentés intervient dans une étape du schéma de description fourni par l’algorithme 1.1. Nous appelons cet algorithme, l’algorithme générique complet. Nous nous plaçons dans le cadre de la mise en correspondance de pixels, c’est pourquoi l’algorithme 1.1 prend en compte des pixels. Les différentes méthodes existantes ne passent pas automatiquement par toutes les étapes de l’algorithme générique complet car elles ne prennent pas forcément en compte tous les éléments constitutifs :

- Les étapes 1 et 2.5 correspondent au schéma multirésolution et elles ne sont réalisées que pour les méthodes utilisant la multirésolution.
- L’étape 2.1 qui correspond au prétraitement n’est effectuée que par les méthodes qui nécessitent des calculs préliminaires ou qui améliorent les images.
- L’étape 2.3 est très différente selon le type de méthode :
 - Les méthodes locales – Les étapes B à E ne sont pas réalisées et il n’y a qu’une seule itération par support.
 - Les méthodes globales – Toutes les étapes sont réalisées.

- L'étape 2.4 n'est réalisée que par les méthodes qui traitent les erreurs d'appariement après avoir établi les correspondances.

Pour chaque passage faire

1. Construire la pyramide d'images
2. Pour chaque niveau de la pyramide faire
 - 2.1. Prétraiter les images
 - 2.2. Partitionner l'ensemble des pixels dont on cherche les correspondants en supports
 - 2.3. Pour chaque support faire
 - 2.3.1. Pour chaque itération faire
 - A. Calculer les coûts locaux en utilisant la zone d'agrégation du coût local et en tenant compte de la zone de recherche
 - B. Calculer le coût de correspondance à partir des coûts locaux sur le support
 - C. Calculer les coûts de voisinage en utilisant la zone d'agrégation du coût de voisinage et en tenant compte de la zone de recherche
 - D. Calculer le coût des contraintes à partir des coûts de voisinage sur le support
 - E. Calculer le coût global à partir du coût de correspondance et du coût des contraintes
 - F. Déterminer les correspondances sur le support selon la méthode d'optimisation
 - 2.4. Affiner les résultats
 - 2.5. Propager les résultats

ALGO. 1.1 – *Mise en correspondance stéréoscopique de pixels* – Cet algorithme générique complet reprend tous les éléments constitutants qui se dégagent de l'ensemble des méthodes que nous avons répertoriées. Tous les algorithmes existants ne suivent pas strictement toutes ces étapes.

1.4 Définitions et notations

1.4.1 Notations générales

Les images gauche et droite, de taille $N_l^{\text{lig}} \times N_l^{\text{col}}$, sont notées I_l , avec $l = g, d$. Les notations suivantes sont utilisées :

- Le vecteur transposé de \mathbf{v} est noté \mathbf{v}^T .
- Le pixel de l'image I_l de coordonnées $(i \ j)^T$ est noté $\mathbf{p}_l^{i,j}$.
- Le terme N_l est le nombre de pixels considérés dans l'image I_l .
- Le niveau de gris associé au pixel de coordonnées $(i \ j)^T$ dans l'image I_l est $I_l^{i,j}$.
- Si $\mathcal{E} = \{e_i\}_{i=0..N-1}$, alors $\text{card}(\mathcal{E})$ est le nombre d'éléments de l'ensemble \mathcal{E} .
- Le terme $\|\mathbf{v}\|$ correspond à la norme euclidienne du vecteur \mathbf{v} .

1.4.2 Disparité

Nous nous plaçons dans le cas où la mise en correspondance est effectuée de l'image gauche vers l'image droite mais toutes les explications et les formules peuvent être utilisées dans le cas d'une mise en correspondance de la droite vers la gauche. Une manière de représenter le résultat d'une mise en correspondance consiste à associer à chaque pixel $\mathbf{p}_g^{i,j} = (i \ j)^T$ de l'image gauche un vecteur appelé disparité, défini par ses composantes $(u - i \ v - j)^T$ où $(u \ v)^T$ sont les coordonnées dans l'image droite du point correspondant au pixel $\mathbf{p}_g^{i,j}$.

1.4.3 Fonction de disparité

Nous pouvons assimiler le problème de la mise en correspondance à la recherche d'une fonction de disparité d qui attribue une disparité à chaque pixel $\mathbf{p}_g^{i,j}$. Dans le cas général, nous avons :

$$d : \mathbb{N}^2 \longrightarrow \mathbb{R}^2 \quad (1.1)$$

$$\mathbf{p}_g^{i,j} \longmapsto d(\mathbf{p}_g^{i,j}) = (u - i \ v - j)^T. \quad (1.2)$$

Lorsqu'une disparité réelle est calculée, il s'agit d'un appariement au sous-pixel. La plupart des méthodes effectuent une mise en correspondance de pixels à pixels, c'est-à-dire qu'elles calculent des disparités entières, et dans ce cas, nous avons :

$$d : \mathbb{N}^2 \rightarrow \mathbb{Z}^2. \quad (1.3)$$

1.5 Images et scènes traitées

Nous nous sommes limités, dans cette étude, aux images optiques (nous ne parlerons pas d'images radar et des techniques associées). La méthode de mise en correspondance dépend avant tout de l'acquisition des images et du type de scènes traitées. Voici les types d'images que nous pouvons rencontrer :

- *Les images en niveaux de gris ou en couleur* – Ce sont surtout des images en niveaux de gris qui ont été utilisées, mais, ces dernières années, les images couleur sont de plus en plus utilisées.
- *Les images multimodales* – Elles correspondent aux images médicales acquises par différents procédés : IRM (image par résonance magnétique), images échographiques (utilisation des ultrasons) et images acquises par scanner à rayon X. Dans [Hermosillo Valadez 02, Rogelj 03], les auteurs présentent notamment des méthodes pour des IRM, dans [Roche 01], pour des images échographiques et, dans [Faghihi 02], il s'agit d'IRM et d'images acquises par scanner à rayon X.
- *Les images aériennes et satellitaires* – Elles correspondent généralement à des images prises d'avion ou d'un satellite. Dans [Gennert 88, Lotti 94, Geiger 95, Papparoditis 98, Chahata 05], un pixel représente entre un mètre et dix mètres au sol. Certains auteurs utilisent des images à très haute résolution, entre dix à trente centimètres, comme [Fuchs 01, Jibrini 04]. Du fait de leur haute définition, ces images sont souvent traitées par multirésolution pour réduire les temps de calcul.
- *Les images réciproques* – Dans [Zickler 03], les images réciproques sont obtenues en échangeant la source lumineuse et le capteur entre les deux prises de vues. Ceci permet de supposer que la relation entre l'intensité de deux pixels qui se correspondent est indépendante de la réflectance de la surface des objets de la scène. Ce type d'acquisition combine des aspects de la stéréovision (deux capteurs sont utilisés) à des aspects de la stéréophotométrie (on déplace les sources de lumière).

Les types de scènes que nous pouvons distinguer sont :

- *Les scènes complexes d'intérieur ou d'extérieur* – Ce sont les scènes les plus difficiles à traiter, car les images présentent de nombreuses difficultés comme, par exemple, les changements de luminosité d'une image à l'autre et de nombreuses occultations.
- *Les scènes aquatiques* – Ces scènes sont plus rares mais, ces dernières années, on a vu des applications apparaître dans ce domaine. Dans [Laurence 98, Garcia 01c, Semani 05], les auteurs proposent des méthodes pour le suivi de poissons dans des aquariums ou dans leur milieu naturel. Les images acquises sont en couleur et la composante dominante étant le bleu, les auteurs n'utilisent, en général, plus que cette composante dans l'analyse de la scène.
- *Les scènes planes par morceaux* – Dans le cas de scènes polyédriques, certains auteurs, comme dans [Scharstein 02], se sont servis des propriétés particulières liées à la géométrie de ce type de scènes pour faciliter la mise en correspondance et notamment pour obtenir des couples stéréoscopiques servant à l'évaluation des performances des algorithmes de mise en correspondance (cf. chapitre 2).

1.6 Contraintes pour la mise en correspondance

1.6.1 Notion de contrainte

Une contrainte est une propriété liée à une correspondance qui découle d'hypothèses faites sur :

- *La géométrie du capteur* – La principale contrainte utilisée est la contrainte épipolaire. Par abus de langage, on appelle cette contrainte une contrainte géométrique (on ne précise pas qu'il s'agit de la géométrie du capteur) et nous conserverons cette appellation dans ce mémoire.
- *La géométrie de la scène* – Il s'agit des contraintes d'unicité, d'ordre et de symétrie. Jones [Jones 97] les appelle des contraintes de compatibilité. Elles ne sont pas appliquées seulement à des pixels, mais aussi à d'autres types de primitives, comme les segments. Nous allons décrire, à partir du paragraphe 1.6.4, les contraintes de compatibilité les plus utilisées, mais cette liste n'est pas exhaustive.
- *La réflectance de la surface des objets ainsi que le type et la position de la source lumineuse* – Cela concerne la contrainte de dissimilarité. Nous avons choisi de ne pas traiter la contrainte de dissimilarité dans ce paragraphe car nous considérons qu'elle constitue l'élément essentiel du coût de mise en correspondance. Nous aborderons donc la notion de dissimilarité au paragraphe 1.9.

Les buts de l'utilisation de ces contraintes sont :

- limiter le nombre de correspondants potentiels pour chaque pixel ;
- choisir entre plusieurs correspondants potentiels ;
- vérifier et, le cas échéant, supprimer des correspondances.

Pour une contrainte donnée, il y a deux aspects à considérer : la définition (la règle) elle-même et les buts de l'utilisation de cette contrainte.

Nous pouvons aussi distinguer, comme dans [Zhang 93, Jones 97], les contraintes unaires qui ne prennent en compte qu'une correspondance, des contraintes binaires, qui utilisent les informations relatives à deux correspondances.

1.6.2 Contrainte épipolaire

La contrainte épipolaire, contrainte géométrique et unaire, peut être utilisée pour faciliter la mise en correspondance en réduisant la zone de recherche. La figure 1.1 résume la géométrie épipolaire. Le point \mathbf{P} de la scène et les centres de projection des deux caméras, \mathbf{O}_g et \mathbf{O}_d , définissent le plan

épipolaire. Les intersections de ce plan avec les deux plans images sont deux droites appelées épipolaires conjuguées. La contrainte épipolaire est donc définie par :

Le correspondant \mathbf{p}_d d'un point \mathbf{p}_g , qui se trouve sur la droite épipolaire gauche \mathcal{D}_g , se trouve nécessairement sur la droite \mathcal{D}_d , image de \mathcal{D}_g dans le plan droit. La droite \mathcal{D}_d est appelée la droite épipolaire droite associée au point \mathbf{p}_g .

Les droites épipolaires possèdent la propriété suivante :

Toutes les droites épipolaires concourent en un point appelé l'épipôle, noté \mathbf{e}_l .

La contrainte épipolaire permet d'obtenir une relation entre les points de l'image gauche et les points de l'image droite qui ne dépend que des paramètres des caméras et des coordonnées image et qui est indépendante des coordonnées du point de la scène. La matrice qui représente cette relation est appelée la matrice fondamentale. Zhang [Zhang 96] et Armangué et Salvi [Armangué 03] donnent des états de l'art sur les manières d'estimer cette matrice fondamentale.

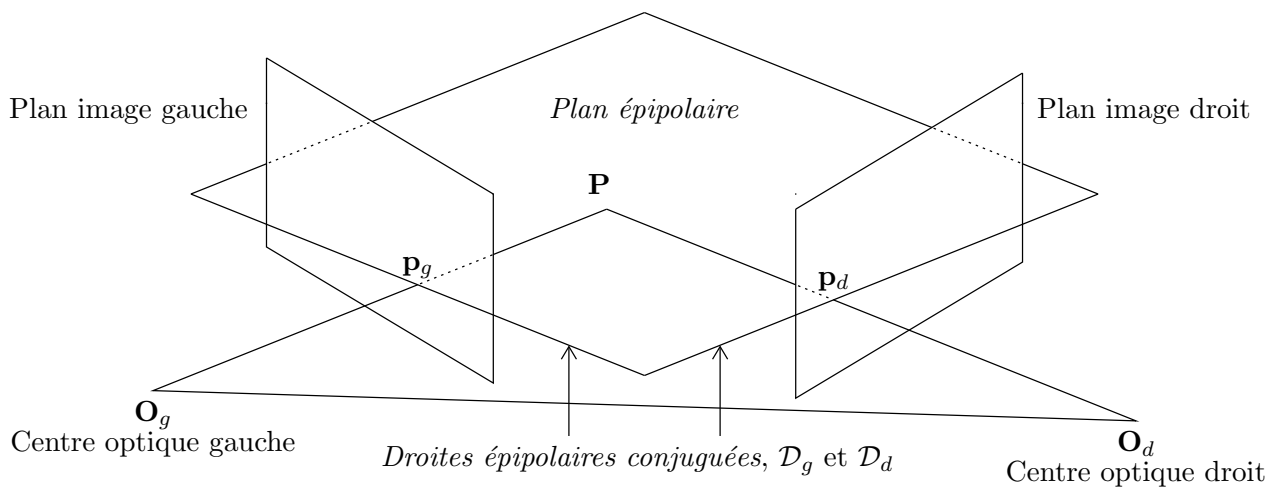


FIG. 1.1 – Géométrie épipolaire.

1.6.3 Rectification épipolaire

La rectification épipolaire consiste à se ramener à une configuration particulière du capteur, appelée configuration parallèle, dans laquelle :

les droites épipolaires sont toutes parallèles et horizontales dans les deux images et deux pixels qui se correspondent se trouvent sur la même ligne. Ainsi le déplacement entre un pixel dans l'image gauche et son correspondant dans l'image droite correspond à une simple translation.

Ainsi, la rectification épipolaire est une transformation géométrique que l'on fait subir aux deux images. La difficulté réside dans l'estimation des paramètres de cette transformation. Deux techniques sont habituellement utilisées pour rectifier des images stéréoscopiques, selon l'information dont on dispose :

- Dans [Ayache 89, Fusiello 00] et [Hartley 99, 41–49], les paramètres de calibrage pour les deux images sont disponibles, alors les matrices de projection perspective associées à chaque image peuvent être estimées et les auteurs peuvent en déduire les transformations à appliquer pour n'avoir qu'une translation entre les deux images.
- Lorsque les informations complètes de calibrage ne sont pas connues, le calcul des paramètres de la transformation à appliquer sur les images passe par l'estimation de la matrice fondamen-

tales à partir de correspondance de points. Il s'agit de la méthode employée dans [Zhang 96] et [Devernay 97, chap. 2].

Cette configuration est très intéressante car elle simplifie la méthode de recherche des correspondants. Garcia [Garcia 01a] et Devernay [Devernay 97] montrent que le processus de rectification génère deux nouvelles images dont les niveaux de gris ont été calculés par interpolation ; ce qui entraîne une dégradation du contenu des images. Cependant, dans la plupart des images utilisées pour faire de la mise en correspondance stéréoscopique, les auteurs rectifient les images et les algorithmes proposés utilisent la contrainte épipolaire. Cette contrainte géométrique est utilisée pendant la mise en correspondance alors que les contraintes de compatibilité peuvent être utilisées pendant ou après l'étape de mise en correspondance. Lorsque les images sont rectifiées, la disparité correspond seulement à la différence de colonnes et dans ce cas, l'expression (1.2) devient :

$$\mathbf{p}_g^{i,j} \mapsto d(\mathbf{p}_g^{i,j}) = (0 \quad v - j)^T. \quad (1.4)$$

Cette configuration implique que la zone de recherche dans l'image droite est constituée des pixels sur la ligne i et situés sur les colonnes dont l'indice est inférieur ou égal à j . De plus, cette zone de recherche est souvent réduite si on fixe les limites inférieure et supérieure aux disparités qui correspondent aux objets de la scène respectivement les plus éloignés et les plus proches du capteur.

1.6.4 Notations utilisées pour les contraintes de compatibilité

À partir de ce paragraphe, nous supposons toujours que nous travaillons avec des images rectifiées. Tous les exemples que nous donnons sont dans le cas où les droites épipolaires conjuguées correspondent à la même ligne dans les deux images étudiées. Nous avons choisi des représentations communes pour toutes les figures illustrant les contraintes de compatibilité ; elles sont résumées dans le tableau 1.1.

REPRÉSENTATION	SIGNIFICATION
•	représente un pixel.
○	indique qu'il y a une correspondance entre deux pixels.
×	indique que cette correspondance est interdite ou qu'un pixel est occulté.
●—————>	représente une correspondance entre deux pixels qui respecte une contrainte.
●- - - - ->	représente une correspondance entre deux pixels qui ne respecte pas une contrainte.

TAB. 1.1 – *Symboles utilisés dans les figures.*

De plus, à partir du paragraphe suivant, nous nous plaçons dans le cas des méthodes de mise en correspondance de pixel à pixel.

1.6.5 Contrainte d'unicité

La contrainte d'unicité, contrainte binaire, est largement employée en stéréovision (cf. figure 1.2.(a)) ; elle est définie par :

$$\text{Si } \mathbf{p}_g^{i,j_1} + d(\mathbf{p}_g^{i,j_1}) = \mathbf{p}_d^{i,v} \text{ alors } \forall j_2 \neq j_1 \mathbf{p}_g^{i,j_2} + d(\mathbf{p}_g^{i,j_2}) \neq \mathbf{p}_d^{i,v}. \quad (1.5)$$

Si deux pixels différents ont le même correspondant, alors la contrainte d'unicité n'est pas vérifiée. Cependant, cette contrainte peut être violée : quand un plan de la scène est très incliné par rapport à l'une des deux caméras, l'effet de raccourcissement peut apparaître et ainsi tous les pixels n'ont pas forcément un correspondant unique. Un exemple est fourni sur la figure 1.2.(b).

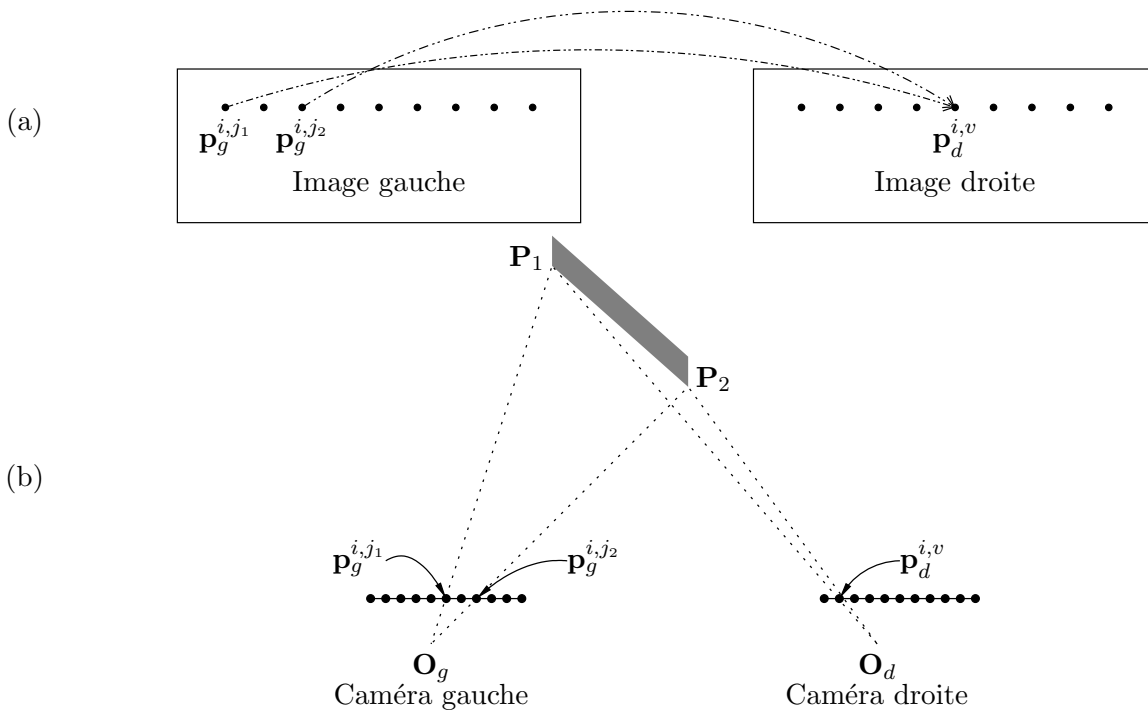


FIG. 1.2 – Contrainte d'unicité – Dans la figure (a), nous pouvons voir les correspondances entre \mathbf{p}_g^{i,j_1} et $\mathbf{p}_d^{i,v}$ et entre \mathbf{p}_g^{i,j_2} et $\mathbf{p}_d^{i,v}$ qui ne respectent pas la contrainte d'unicité. En (b), Il s'agit d'une vue de dessus d'un capteur binoculaire en configuration parallèle et d'une scène. C'est un exemple de violation de la contrainte provoquée par un raccourcissement : tous les points qui se trouvent sur le segment de droite $[\mathbf{P}_1\mathbf{P}_2]$ se projettent sur le pixel $\mathbf{p}_d^{i,v}$ dans l'image droite.

1.6.6 Contrainte d'ordre

La contrainte d'ordre, contrainte binaire, est parfois utilisée en stéréovision (cf. figure 1.3.(a)) et elle est définie par :

$$\text{Si } \mathbf{p}_g^{i,j_1} + d(\mathbf{p}_g^{i,j_2}) = \mathbf{p}_d^{i,v_1} \text{ et } \mathbf{p}_g^{i,j_2} + d(\mathbf{p}_g^{i,j_1}) = \mathbf{p}_d^{i,v_2} \text{ alors } (j_1 - j_2)(v_1 - v_2) \geq 0. \quad (1.6)$$

Elle signifie que l'ordre des pixels de l'image gauche le long de la droite épipolaire doit être le même que celui de leurs correspondants. Cette contrainte peut aussi être violée lorsqu'il y a, dans la scène, des plans transparents fortement inclinés par rapport au plan de l'une des caméras. Un exemple est montré sur la figure 1.3.(b).

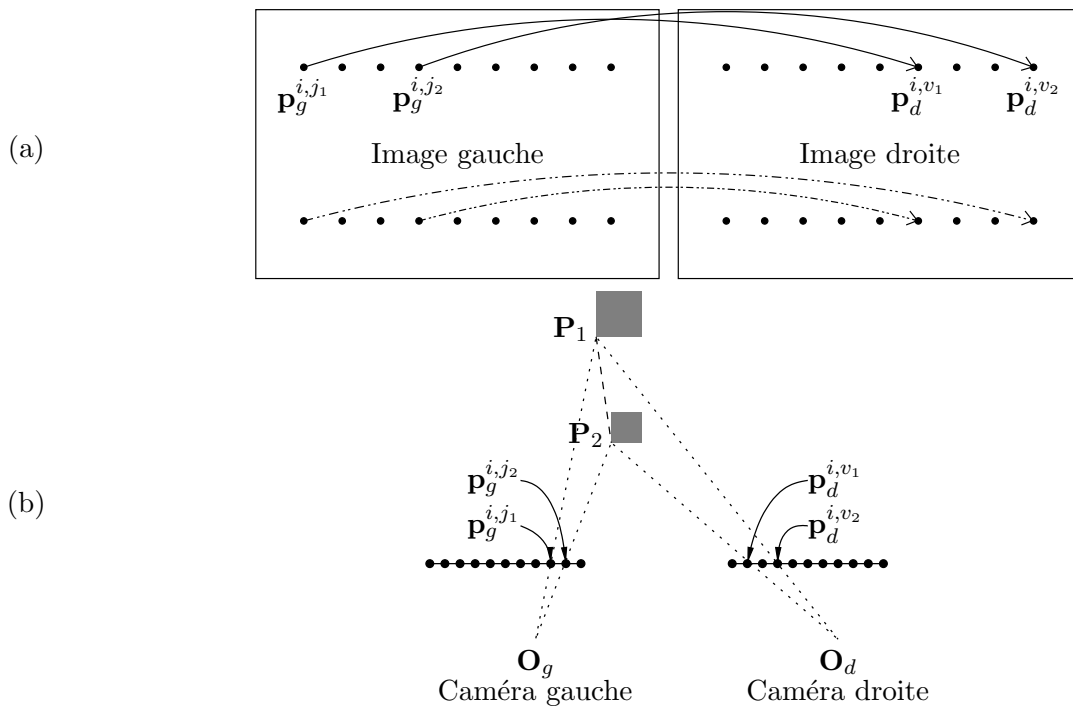


FIG. 1.3 – Contrainte d'ordre – En (a), nous pouvons voir deux correspondances qui respectent la contrainte d'ordre et deux correspondances qui ne la respectent pas. En (b), une vue de dessus d'un capteur binoculaire en configuration parallèle et d'une scène, montre un exemple de violation de la contrainte d'ordre (le plan transparent vertical passant par les points \mathbf{P}_1 et \mathbf{P}_2 , représenté en tirets, est fortement incliné par rapport au plan des images).

1.6.7 Contrainte de symétrie

Un des affinements des résultats d'une mise en correspondance les plus employés est la vérification bidirectionnelle (cf. figure 1.4.(a)), qui met en œuvre la contrainte de symétrie. Cette contrainte binaire s'écrit :

$$\text{Si } \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ alors } \mathbf{p}_d^{i,v} + d(\mathbf{p}_d^{i,v}) = \mathbf{p}_g^{i,j}. \quad (1.7)$$

Deux mises en correspondance sont effectuées, de la gauche vers la droite puis de la droite vers la gauche. Si un pixel $\mathbf{p}_g^{i,j}$ a pour correspondant le pixel $\mathbf{p}_d^{i,v}$, alors, lors de la seconde mise en correspondance, $\mathbf{p}_d^{i,v}$ doit avoir pour correspondant $\mathbf{p}_g^{i,j}$. Si ce n'est pas le cas, la contrainte de symétrie n'est pas

respectée. Nous avons pu trouver une définition différente, mais équivalente, de cette contrainte sous l'appellation de contrainte de consistance forte définie dans [Gong 03, Kostková 03]. Elle est définie comme la conjonction de la contrainte d'unicité de la droite vers la gauche et de la gauche vers la droite. La contrainte de symétrie assure la contrainte d'unicité, ce qui signifie que si un ensemble de correspondances vérifient la contrainte de symétrie, alors elles vérifient aussi la contrainte d'unicité.

Cette contrainte est très forte et, dans certains travaux [Crouzil 96, Belli 00], les auteurs ont proposé une version qui tolère une erreur de quelques pixels que l'on l'appelle la contrainte de symétrie faible (cf. figure 1.4.(b)) et qui est définie par :

$$\text{Si } \mathbf{p}_g^{i,j_1} + d(\mathbf{p}_g^{i,j_1}) = \mathbf{p}_d^{i,v} \text{ et } \mathbf{p}_d^{i,v} + d(\mathbf{p}_d^{i,v}) = \mathbf{p}_g^{i,j_2} \text{ alors } |j_2 - j_1| < T_j, \quad (1.8)$$

avec T_j est un seuil à fixer.

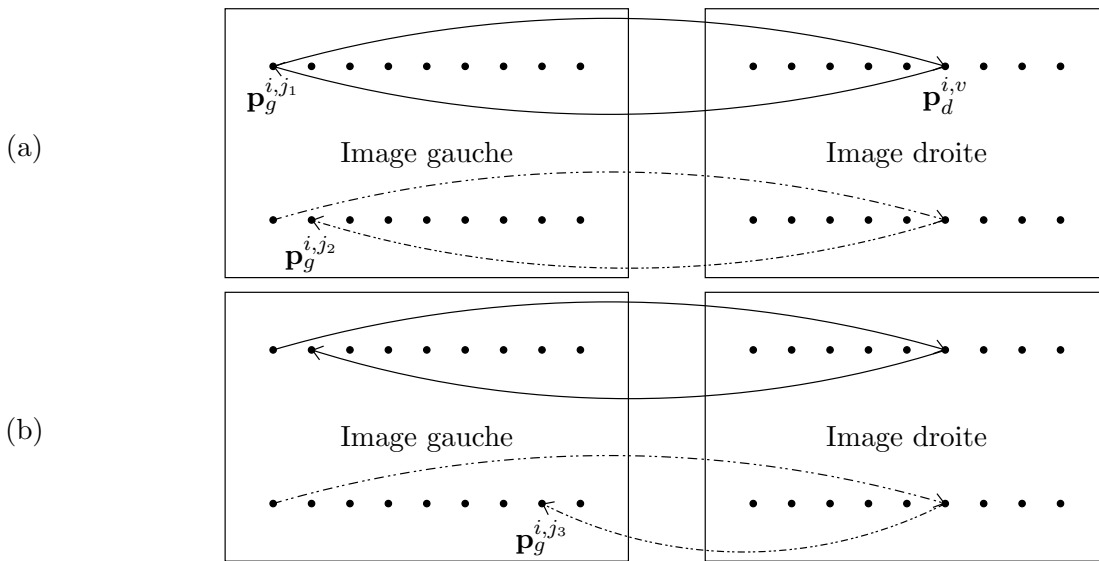


FIG. 1.4 – Contrainte de symétrie (vérification bidirectionnelle) – La contrainte de symétrie est illustrée en (a) et nous pouvons la comparer à la contrainte de symétrie faible en (b). Nous pouvons remarquer que les correspondances entre \mathbf{p}_g^{i,j_1} et $\mathbf{p}_d^{i,v}$ et entre $\mathbf{p}_d^{i,v}$ et \mathbf{p}_g^{i,j_2} sont rejetées lorsqu'on applique la contrainte de symétrie alors qu'elles ne le sont pas quand il s'agit de la contrainte de symétrie faible.

Les contraintes d'ordre et d'unicité ont été très utilisées ; une variante de ces contraintes, la consistance faible, a même été proposée dans [Gong 03, Kostková 03].

1.6.8 Consistance faible

La consistance faible s'appuie sur la contrainte d'unicité et la contrainte d'ordre. Elle est donc binaire et elle est définie par :

$$\begin{aligned} \text{Si } \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ alors} \\ \forall j' > j \mathbf{p}_g^{i,j'} + d(\mathbf{p}_g^{i,j'}) \neq \mathbf{p}_d^{i,v} \text{ et } \forall v' < v \mathbf{p}_d^{i,v'} + d(\mathbf{p}_d^{i,v'}) \neq \mathbf{p}_g^{i,j}. \end{aligned} \quad (1.9)$$

Cela signifie que si $\mathbf{p}_g^{i,j}$ a pour correspondant $\mathbf{p}_d^{i,v}$ alors la contrainte de symétrie doit être vérifiée pour tous les pixels sur la même ligne que $\mathbf{p}_g^{i,j}$ tels que leurs indices de colonne soient supérieurs à j et pour tous les pixels sur la même ligne que $\mathbf{p}_d^{i,v}$ tels que leurs indices de colonnes soient inférieurs à v . Cette contrainte a été proposée pour tenter de limiter l'effet des contraintes d'unicité et d'ordre, c'est-à-dire pour interdire moins de correspondances.

1.6.9 Synthèse et comparaison des contraintes d'unicité, de symétrie, d'ordre et de consistance faible

Toutes ces contraintes peuvent être utilisées pour rejeter des appariements ou conditionner la mise en correspondance, c'est-à-dire, diminuer l'espace de recherche ou pénaliser certaines correspondances. Pour illustrer les différences entre toutes ces contraintes (unicité, ordre, symétrie et consistance faible), nous pouvons étudier une correspondance entre deux pixels et analyser l'impact de chaque contrainte, si elle est appliquée, sur l'appariement des autres pixels de l'image. Sur la figure 1.5, on voit que l'effet de la consistance faible est beaucoup moins important que celui de la contrainte d'ordre et qu'elle permet de tolérer certaines configurations dans lesquelles la contrainte d'unicité et la contrainte d'ordre sont violées.

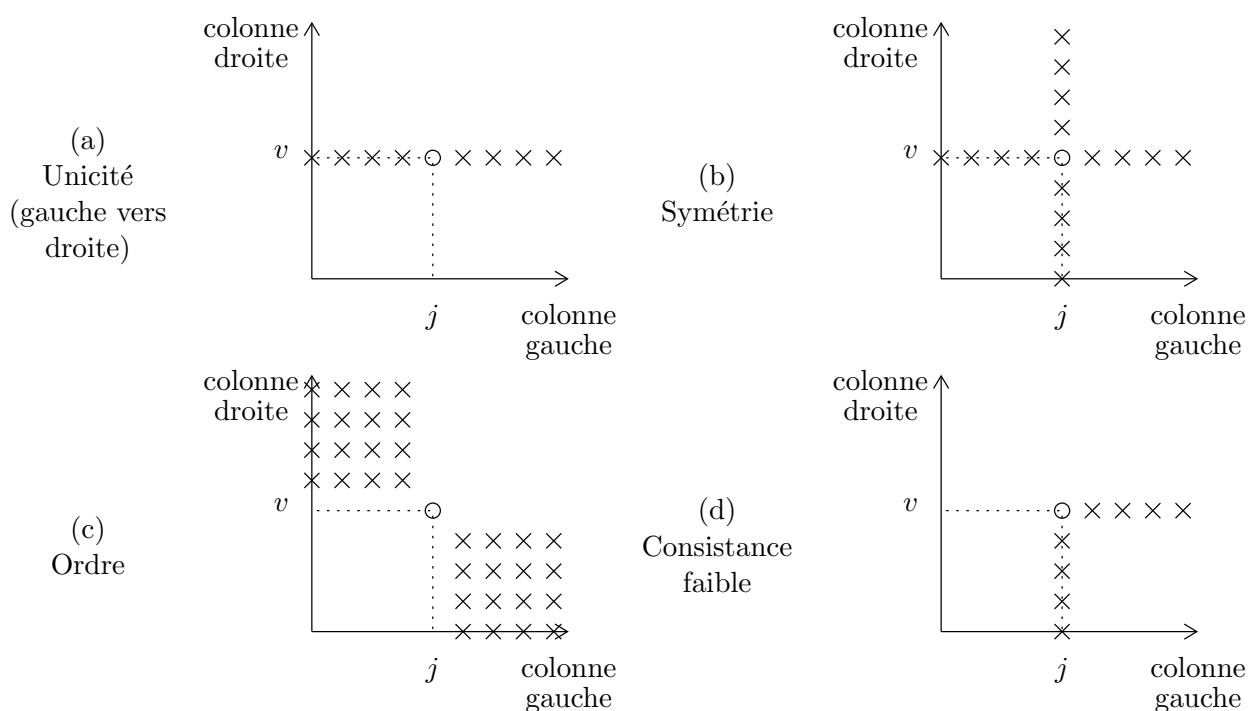


FIG. 1.5 – Comparaison des contraintes – Ces schémas illustrent l'impact des contraintes sur l'ensemble des appariements possibles, lorsque la correspondance entre un pixel de colonne j , dans l'image gauche, et un pixel de colonne v , dans l'image droite, a été établie. Les axes des abscisses et des ordonnées correspondent respectivement à la colonne des pixels de l'image de gauche et de droite sur leur droites épipolaires associées. Dans l'exemple de la contrainte d'unicité, en (a), si on établit la correspondance entre les pixels de colonnes j et v , alors toutes les correspondances sur la colonne v sont interdites. Grâce aux figures (a) et (b), on peut illustrer que la contrainte de symétrie est bien équivalente à la contrainte d'unicité appliquée de la gauche vers la droite et de la droite vers la gauche. La contrainte de symétrie implique aussi la contrainte de consistance faible. La contrainte d'ordre (c) est la plus forte, dans le sens où elle permet d'interdire le plus grand nombre de correspondances. Les contraintes d'unicité (a) et de consistance faible (d) sont les moins fortes.

1.6.10 Limite du gradient de disparité

La limite du gradient de disparité est aussi très utilisée [Barnard 89, Jones 97, Zhang 00, Mayer 03, Wei 04]. Cette contrainte binaire s'appuie sur l'hypothèse d'une continuité dans les disparités, ce qui implique que le gradient de disparité, noté G_d , ne doit pas dépasser une certaine valeur. Une approximation du gradient de disparité est définie comme la différence des disparités divisée par la séparation cyclopéenne (différence de la distance entre deux pixels dans l'image gauche et de la distance entre leurs correspondants respectifs dans l'image droite). Cette contrainte est définie dans [Horaud 93, p. 185] par :

$$\begin{aligned} &\text{Si } \mathbf{p}_g^{i,j_1} + d(\mathbf{p}_g^{i,j_1}) = \mathbf{p}_d^{i,v_1} \text{ et } \mathbf{p}_g^{i,j_2} + d(\mathbf{p}_g^{i,j_2}) = \mathbf{p}_d^{i,v_2} \text{ alors} \\ G_d(\mathbf{p}_g^{i,j_1}, \mathbf{p}_d^{i,v_1}, \mathbf{p}_g^{i,j_2}, \mathbf{p}_d^{i,v_2}) &= \frac{\left\| (\mathbf{p}_g^{i,j_1} - \mathbf{p}_g^{i,j_2}) - (\mathbf{p}_d^{i,v_1} - \mathbf{p}_d^{i,v_2}) \right\|}{\left\| \frac{1}{2} (\mathbf{p}_g^{i,v_1} + \mathbf{p}_d^{i,v_1}) - (\mathbf{p}_g^{i,j_2} + \mathbf{p}_d^{i,v_2}) \right\|} < T_G, \end{aligned} \quad (1.10)$$

où T_G est un seuil à fixer.

1.6.11 Contrainte de rang

Banks et Bennamoun [Banks 01] proposent la contrainte de rang (cf. figure 1.6). Cette contrainte unaire est aussi utilisée dans [Rao 03, Zickler 03]. Elle s'appuie sur la comparaison du niveau de gris du pixel central avec celui des pixels de la zone d'agrégation et se traduit par la règle suivante :

$$\begin{aligned} &\text{Si } \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ alors} \\ \forall \mathbf{p}_g^{i',j'} \in \text{ZAC}^c(\mathbf{p}_g^{i,j}), \mathbf{p}_d^{i',v'} \in \text{ZAC}^c(\mathbf{p}_d^{i,v}) \text{ tels que } j' - j = v' - v & \\ (I_g^{i',j'} - I_g^{i,j})(I_d^{i',v'} - I_d^{i,v}) > 0. & \end{aligned} \quad (1.11)$$

Le terme $\text{ZAC}^c(\mathbf{p}_l^{i,j})$ correspond à l'ensemble des pixels qui se trouvent dans la zone d'agrégation de la contrainte du pixel $\mathbf{p}_l^{i,j}$. Cette contrainte suppose que le signe de la différence entre les niveaux de gris du pixel étudié et de ceux de sa zone d'agrégation doit être le même pour $\mathbf{p}_g^{i,j}$ et $\mathbf{p}_d^{i,v}$.

$$\begin{array}{ccc|ccc} \text{(a)} & & & \text{(b)} & & \text{(c)} \\ \begin{pmatrix} 250 & 200 & 198 \\ 100 & \mathbf{104} & 110 \\ 50 & 48 & 78 \end{pmatrix} & & & \begin{pmatrix} 255 & 205 & 203 \\ 105 & \mathbf{109} & 115 \\ 55 & 53 & 93 \end{pmatrix} & & \begin{pmatrix} 255 & 205 & 100 \\ 105 & \mathbf{109} & 105 \\ 55 & 53 & 110 \end{pmatrix} \end{array}$$

FIG. 1.6 – Contrainte de rang – Sur cette figure, chaque matrice représente le niveau de gris du pixel étudié, au centre, en gras, et les niveaux de gris des pixels de la zone d'agrégation (ici, nous avons choisi une fenêtre de taille 3×3 centrée sur le pixel considéré). En (a), il s'agit du pixel dont on cherche le correspondant et de son voisinage 3×3 . En (b) et (c), il s'agit de deux correspondants possibles et de leurs zones d'agrégation respectives. Entre (a) et (b), la contrainte de rang est respectée. Entre (a) et (c), les pixels de la colonne de droite (en italique), ne respectent pas la contrainte.

1.6.12 Contrainte de continuité figurale

La contrainte de continuité figurale [Ohta 85, Mohan 89, Horaud 93, p. 196], contrainte binaire, est définie par :

Soient \mathcal{C}_g un contour dans l'image gauche et \mathcal{C}_d un contour dans l'image droite. On note :

$$\mathcal{P}_{\mathcal{C}_g} = \{\mathbf{p}_g^{i,j} \in \mathcal{C}_g\} \text{ et } \mathcal{P}'_{\mathcal{C}_g} = \{\mathbf{p}_g^{i,j} \in \mathcal{P}_{\mathcal{C}_g} \mid \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j}) \in \mathcal{C}_d\}. \quad (1.12)$$

Si $\text{card}(\mathcal{P}'_{\mathcal{C}_g}) \geq \frac{1}{2} \text{card}(\mathcal{P}_{\mathcal{C}_g})$ alors $\forall (i',j')$ tels que $\mathbf{p}_g^{i',j'} \in \mathcal{P}_{\mathcal{C}_g}$, $(\mathbf{p}_g^{i',j'} + d(\mathbf{p}_g^{i',j'})) \in \mathcal{C}_d$.

Si un pixel $\mathbf{p}_g^{i,j}$ appartient à un contour \mathcal{C}_g dont la majorité des pixels ont été appariés avec des pixels appartenant au même contour \mathcal{C}_d , dans l'image droite, alors le correspondant du pixel $\mathbf{p}_g^{i,j}$ appartient à \mathcal{C}_d . Cette contrainte a été utilisée dans [Ohta 85] dans le cadre de la programmation dynamique. Il s'agit de l'exemple le plus significatif de l'utilisation de cette contrainte. Mohan *et al.* [Mohan 89] ont utilisé cette contrainte pour tenter de corriger les erreurs d'appariement après avoir calculé toutes les correspondances.

1.6.13 Autres contraintes

Il existe d'autres contraintes, moins populaires que celles que nous venons de présenter. Nous pouvons toutefois citer les contraintes d'ambiguïté et d'imprécision [De Joinville 01]. Ces contraintes s'utilisent avec des méthodes locales, c'est pourquoi nous les détaillerons dans le chapitre 3, paragraphe 3.6.1.2.

1.7 Attributs des pixels

Dans cet état de l'art, nous considérons les méthodes s'appliquant sur des images en niveaux de gris et des images couleur. Les correspondances peuvent être déterminées pour une partie des pixels – mise en correspondance partielle (ou non dense) – ou pour tous les pixels – mise en correspondance dense. Nous prendrons en compte ces deux types de méthodes.

Jones [Jones 97] considère qu'il existe deux catégories de pixels à étudier (cf. tableau 1.2) :

- *Tous les pixels de l'image* – Dans ce cas, les attributs utilisés sont assez limités. Cela peut être le niveau de gris (c'est le plus populaire) ou la couleur.
- *Des pixels particuliers de l'image* – Nous appelons pixels particuliers des pixels qui correspondent à des points particuliers dans la scène. Dans ce contexte, la mise en correspondance ne sera pas dense. Comme type de pixels particuliers, nous pouvons citer :
 - *Les points d'intérêt* – Ce sont les points qui possèdent des caractéristiques particulières les rendant plus discriminants pour la mise en correspondance. De très nombreuses méthodes s'appuient sur une détection de points d'intérêt [Weng 92, Jung 01]. Les détecteurs de points d'intérêt les plus populaires sont celui de Harris et Stephens [Harris 88], utilisé notamment dans [Gouet 98, Smith 98, Gouet 00, Lhuillier 04, Benhimane 04] et celui de Shi et Tomasi [Shi 94], utilisé dans [Megyesi 04]. D'autres méthodes effectuent une détection de coins comme dans [Ishikawa 98, Stock 03, Kaplan 04], en utilisant notamment le détecteur SUSAN proposé dans [Smith 97].
 - *Les points contour* – Les points contour, utilisés notamment dans [Ishikawa 98, Shao 02, Petrakis 02], sont détectés le plus souvent par le détecteur de Canny [Canny 86] ou l'opérateur de Shen et Castan [Shen 92]. Dans [Xu 97, Koch 98, Paparoditis 98, Rziza 00, Xiao 03, Torr 04], l'opérateur de Canny est utilisé alors que dans [Belli 00], on se sert de

celui de Shen et Castan. Dans [Pajares 03], Pajares et De La Cruz s'appuient sur une image filtrée par un laplacien de gaussien et dans [Ohta 85], Ohta et Kanade appliquent une combinaison de plusieurs opérateurs permettant le calcul des dérivées premières des niveaux de gris.

Dans la plupart des cas, les attributs utilisés sont les mêmes que ceux utilisés pour les pixels. Nous pouvons toutefois citer des méthodes qui ont été proposées pour prendre en compte de grandes déformations entre les images afin d'effectuer notamment de l'indexation d'images ou du suivi d'objets :

- Dans [Carneiro 04], à chaque point d'intérêt est associé un vecteur d'attributs qui dépend de l'histogramme local (le pixel étudié et ses pixels voisins sont pris en compte).
- Dans [Stock 03], les caractéristiques des droites qui forment un coin sont considérées.
- Dans [Schmid 95, Gros 97], chaque pixel particulier est caractérisé par des invariants locaux.

CATÉGORIE	PRIMITIVE	ATTRIBUTS	PRÉTRAITEMENT
<i>Pixels</i>	-	Niveaux de gris ou couleur	-
<i>Pixels particuliers</i>	Points d'intérêt	Niveaux de gris ou couleur Histogramme local	Détecteur de Harris et Stephens [Harris 88] ou opérateur de SUSAN [Smith 97]
	Points contour	Invariants locaux	Opérateur de Canny [Canny 86], de Shen et Castan [Shen 92]

TAB. 1.2 – Récapitulatif sur les types de pixels, leurs attributs et l'opérateur de détection choisi.

1.8 Prétraitement des images

Le prétraitement des images est lié aux images utilisées, aux pixels pris en compte et aux attributs associés aux pixels choisis. Les différents cas que nous pouvons distinguer sont les suivants :

- *Pour améliorer les images* – Pour tenter d'éliminer d'éventuels bruits présents dans les images, dans [Jones 92, Kim 05], les auteurs recommandent d'utiliser un filtre gaussien.
- *Pour détecter les pixels et extraire leurs attributs* – Il est nécessaire d'appliquer un opérateur de détection au préalable pour extraire les pixels particuliers (cf. paragraphe 1.7) ou les attributs associés.
- *Pour extraire des informations utilisées dans les coûts locaux et globaux* – Dans certaines méthodes, une segmentation en régions des images est réalisée, pour être ensuite utilisée dans le calcul des coûts de mise en correspondance.

1.8.1 Extraction des attributs des pixels

Certaines méthodes n'utilisent pas directement les niveaux de gris pour caractériser les pixels et nous pouvons distinguer :

- *Les méthodes locales s'appuyant sur le calcul des dérivées des images* – Ces méthodes effectuent donc un préfiltrage de l'image, en utilisant, notamment, le laplacien de gaussien [Nishihara 88, Kanade 96], le filtre de Sobel [Seitz 89, Ullah 01], le filtre de Kirsch [Seitz 89], le filtre de Roberts [Nack 75] et le filtre de Shen-Castan [Crouzil 96].
- *Les méthodes locales utilisant des transformations de rang* – Zabih et Woodfill [Zabih 94] introduisent ce type de méthodes (nous détaillerons cette technique dans le paragraphe 3.3.4.5).

- *Les méthodes à base d'images intrinsèques* – Ces images, proposées dans [Silva 00], sont obtenues en effectuant, pour chaque pixel de chaque ligne de l'image, la somme des dérivées premières du début de la ligne jusqu'au pixel étudié.
- *Les méthodes locales, utilisant la variance ou la moyenne des niveaux de gris sur la zone d'agrégation* – Ces variances ou ces moyennes sont calculées, en prétraitement, pour tous les pixels de l'image [Falkenhagen 97, Menard 97, Arcara 00].

1.8.2 Segmentation en régions

Ces dernières années ont vu se multiplier les publications présentant des méthodes coopératives, *segment-based methods*, qui intègrent une segmentation en régions à la mise en correspondance [Birchfield 99, Han 01, Jahn 01, Tao 01, Lin 02, Zhang 04]. Voici les différents types de régions considérées et la méthode de segmentation associée :

- *Régions homogènes* – Dans [Gutiérrez 04], les régions homogènes (non texturées) sont distinguées des régions non homogènes en utilisant un seuil sur l'amplitude de la norme du gradient sur chaque image. Le gradient est calculé en utilisant les dérivées du filtre gaussien.
- *Régions achromatiques* – Dans [Tseng 92, Koschan 93], les auteurs séparent, dans des images couleur, les régions achromatiques des régions chromatiques en utilisant des seuils spécifiques sur les composantes I et S du système de couleur HSI (ce système est décrit dans le chapitre 4 au paragraphe 4.2.3).
- *Régions extrêmes à stabilité maximale* – Dans [Matas 02], Matas *et al.* s'appuient sur le niveau de gris des pixels pour déterminer ces régions qui sont sélectionnées en comparant le niveau de gris des pixels sur le contour des régions et ceux qui sont adjacents.
- *Segmentation par mean-shift* – Proposée dans [Comaniciu 99], une segmentation en régions par *mean-shift* s'appuyant sur un critère d'homogénéité de couleur est souvent utilisée [Bleyer 04, Hong 04, Wei 04].

1.9 Coûts de mise en correspondance, zones d'agrégation et support

1.9.1 Coût global de mise en correspondance

Le coût global de mise en correspondance qui évalue à quel point un ensemble de correspondances est valide a pour forme générale :

$$E_{\text{global}}(d) = (1 - \lambda)E_{\text{correspondance}}(d) + \lambda E_{\text{contrainte}}(d), \quad (1.13)$$

où $\lambda \in [0; 1]$ permet d'ajuster l'influence des poids entre les deux termes :

- $E_{\text{correspondance}}$ qui est le coût de correspondance et qui est le terme d'attache aux données.
- $E_{\text{contrainte}}$ qui est le coût des contraintes et qui modélise les interactions entre les pixels considérés.

Résoudre le problème de la mise en correspondance, formulé de cette manière, consiste à trouver la fonction de disparité d qui minimise ce coût E_{global} .

1.9.1.1 Coût de correspondance

Ce terme évalue à quel point un ensemble de correspondances est fiable ; il est déterminé par la somme sur le support \mathcal{S} des coûts locaux et il est de la forme :

$$E_{\text{correspondance}}(d) = \sum_{\mathbf{p}_g^{i,j} \in \mathcal{S}} E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) \quad (1.14)$$

$$\text{avec } \mathbf{p}_d^{i,v} = \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j}).$$

Le coût local correspond au coût d'une correspondance et il est défini par :

$$E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) = \sum_{\substack{\mathbf{p}_g^{i',j'} \in \text{ZA}(\mathbf{p}_g^{i,j}) \\ \mathbf{p}_d^{i',v'} \in \text{ZA}(\mathbf{p}_d^{i,v})}} E_{\text{dissimilarité}}(\mathbf{p}_g^{i',j'}, \mathbf{p}_d^{i',v'}), \quad (1.15)$$

où $\text{ZA}(\mathbf{p}_l^{i,j})$ correspond à la zone d'agrégation considérée (pixels pris en compte pour le calcul du coût local). La zone d'agrégation d'un pixel $\mathbf{p}_l^{i,j}$ est le plus souvent un voisinage de $\mathbf{p}_l^{i,j}$, c'est-à-dire qu'elle correspond à un ensemble de pixels connexes qui contient $\mathbf{p}_l^{i,j}$. Un exemple classique de voisinage est une fenêtre carrée de taille $(2t + 1) \times (2t + 1)$ centrée sur $\mathbf{p}_l^{i,j}$. Le terme $E_{\text{dissimilarité}}$ est le coût de dissimilarité : il évalue à quel point deux pixels ne se ressemblent pas. Nous verrons par la suite qu'il s'agit la plupart du temps d'une dissimilarité des niveaux de gris.

1.9.1.2 Coût des contraintes

Il modélise les contraintes entre les pixels du support \mathcal{S} et les pixels de la zone d'agrégation. Il y a N_c contraintes et ainsi N_c zones d'agrégation associées que nous notons ZAC car cette zone est différente de la zone d'agrégation associée au coût local, ZA, et elle correspond la plupart du temps à un voisinage. Le coût des contraintes est donné par :

$$E_{\text{contrainte}}(d) = \sum_{c=0}^{N_c-1} E_{\text{contrainte}}^c(d) = \sum_{c=0}^{N_c-1} \sum_{\mathbf{p}_g^{i,j} \in \mathcal{S}} E_{\text{voisinage}}^c(\mathbf{p}_g^{i,j}), \quad (1.16)$$

Le coût du voisinage permet d'évaluer l'effet de la contrainte sur le pixel étudié et ses voisins ; il est défini par :

$$E_{\text{voisinage}}^c(\mathbf{p}_g^{i,j}) = \sum_{\mathbf{p}_g^{i',j'} \in \text{ZAC}^c(\mathbf{p}_g^{i,j})} E_{\text{lissage}}^c(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'}). \quad (1.17)$$

Le terme E_{lissage}^c permet de comparer les disparités associées aux deux pixels considérés et il est appelé coût de lissage. La zone d'agrégation associée à la $c^{\text{ème}}$ contrainte est notée ZAC^c .

Nous avons présenté les deux termes du coût global de mise en correspondance, nous pouvons à présent reprendre l'équation (1.13) et la détailler :

$$E_{\text{global}}(d) = \sum_{\mathbf{p}_g^{i,j} \in \mathcal{S}} \left((1 - \lambda) \underbrace{\sum_{\substack{\mathbf{p}_g^{i',j'} \in \text{ZA}(\mathbf{p}_g^{i,j}) \\ \mathbf{p}_d^{i',v'} \in \text{ZA}(\mathbf{p}_d^{i,v})}} E_{\text{dissimilarité}}(\mathbf{p}_g^{i',j'}, \mathbf{p}_d^{i',v'})}_{\text{coût local}} + \lambda \sum_{c=0}^{N_c-1} \underbrace{\sum_{\mathbf{p}_g^{i',j'} \in \text{ZAC}^c(\mathbf{p}_g^{i,j})} E_{\text{lissage}}^c(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'})}_{\text{coût du voisinage}} \right) \quad (1.18)$$

$$\text{avec } \mathbf{p}_d^{i,v} = \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j}).$$

Les choix à faire pour ce coût global de mise en correspondance sont :

- λ – Le plus souvent, les publications n'abordent pas cette pondération.
- $E_{\text{dissimilarité}}$ – Ce coût s'appuie souvent sur les niveaux de gris des pixels, mais nous verrons au chapitre 3 qu'il existe d'autres possibilités.
- E_{lissage}^c – Ce coût, dans la majorité des cas, est lié à la disparité associée à chaque pixel ou au fait que le pixel est considéré comme occulté ou non.
- ZA – Nous distinguons les méthodes qui utilisent une zone d'agrégation réduite à un seul pixel (le pixel étudié), ce sont les méthodes globales, des méthodes qui utilisent une zone d'agrégation plus importante, ce sont les méthodes locales et les méthodes mixtes.
- ZAC^c – Les zones d'agrégation des contraintes ne sont, en général, pas identiques aux zones d'agrégation du coût local. Dans de nombreux cas, la zone d'agrégation pour le coût local est limitée au pixel étudié alors que la zone d'agrégation des contraintes est un 4-voisinage.
- \mathcal{S} – Le support peut être le pixel considéré ou l'ensemble des N_g pixels considérés dans l'image ou les N_g^{col} pixels qui se trouvent sur la même ligne (dans le cas des images rectifiées).

Une manière très répandue de classer les méthodes de mise en correspondance est de distinguer les méthodes locales des méthodes globales, c'est ce qui est proposé dans [Scharstein 02, Brown 03]. Les méthodes locales effectuent N_g minimisations des coûts locaux et la zone d'agrégation est toujours strictement supérieure à un pixel (au pixel étudié), alors que les méthodes globales effectuent une minimisation du coût global qui est une fonction des coûts locaux et la zone d'agrégation est souvent réduite au pixel considéré. Nous reprenons cette classification en la complétant. Nous considérons :

- les méthodes locales ;
- les méthodes globales dont la zone d'agrégation est réduite à un pixel ;
- les méthodes globales dont la zone d'agrégation n'est pas réduite à un pixel et que nous appelons les méthodes mixtes.

1.9.2 Méthodes locales – Méthodes par corrélation

1.9.2.1 Coût local

Les méthodes locales correspondent à N_g minimisations des coûts locaux (cf. algorithme 1.2). Elles sont aussi appelées méthodes par corrélation. Une corrélation indique un degré de ressemblance entre deux ensembles de données. Par abus de langage, dans ce chapitre, quand nous parlons de méthode par corrélation, la mesure de corrélation utilisée évalue le degré de dissimilarité entre deux ensembles de pixels (score de corrélation). Pour ces méthodes, nous avons :

- \mathcal{S} – Il y a N_g supports, chacun étant réduit à un singleton que nous notons $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$.
- ZA – Il y a deux possibilités : utiliser une fenêtre carrée centrée sur le pixel considéré ou des fenêtres adaptatives (nous abordons cette notion au chapitre 5, dans le paragraphe 5.5.1).
- $E_{\text{dissimilarité}}$ – Il existe de très nombreuses variantes [Aschwanden 92, Zabih 94, Kaneko 02], une des plus utilisées étant :

$$E_{\text{dissimilarité}}(\mathbf{p}_g^{i',j'}, \mathbf{p}_d^{i',v'}) = f_{\text{dissimilarité}}(I_g^{i',j'} - I_d^{i',v'}) = (I_g^{i',j'} - I_d^{i',v'})^2, \quad (1.19)$$

où $f_{\text{dissimilarité}}$ est une fonction de la différence des deux niveaux de gris. Nous ne détaillons pas toutes les formes que peut prendre ce coût de dissimilarité dans les méthodes locales puisqu'elles sont largement décrites au chapitre 3.

Les méthodes locales n'ont pas de coût des contraintes et seul le coût local, équation (1.15), est utilisé. Ainsi la fonction de disparité d est déterminée de la façon suivante :

$$\forall \mathbf{p}_g^{i,j} \in \mathcal{P}_g \quad d(\mathbf{p}_g^{i,j}) = (0 \ v - j)^T \text{ avec } \mathbf{p}_d^{i,v} = \underset{\mathbf{p}_d^{i,v'} \in \mathcal{Z}_d(\mathbf{p}_g^{i,j})}{\text{argmin}} E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v'}), \quad (1.20)$$

où \mathcal{P}_g est l'ensemble des pixels considérés dans l'image, $Z_d(\mathbf{p}_g^{i,j})$ correspond à la zone de recherche associée au pixel $\mathbf{p}_g^{i,j}$. La méthode d'optimisation employée est la méthode de recherche exhaustive, *winner take all*, notée WTA, c'est-à-dire que le pixel qui obtient le meilleur coût est sélectionné.

Pour chaque passage faire

1. Construire la pyramide d'images
2. Pour chaque niveau de la pyramide faire
 - 2.1. Prétraiter les images
 - 2.2. $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$
 - 2.3. Pour chaque $\mathcal{S}^{i,j}$ faire
 - A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire
 - Calculer $E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v})$ en tenant compte de $ZA(\mathbf{p}_g^{i,j})$
 - F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA
 - 2.4. Affiner les résultats
 - 2.5. Propager les résultats

ALGO. 1.2 – *Méthode locale de mise en correspondance stéréoscopique de pixels* – Cet algorithme générique est une instance de l'algorithme 1.1. Les notations utilisées sont celles présentées dans le paragraphe 1.9.2. Nous pouvons voir que les étapes B à E ne sont pas effectuées et que la boucle à l'étape 2.3.1 est absente et c'est pourquoi, par la suite, tous les algorithmes qui sont des instances de cet algorithme ne feront plus mention de cette étape.

1.9.2.2 Exemple de méthode locale

Un exemple de méthode locale est fourni avec l'algorithme 1.3 qui a obtenu le meilleur résultat parmi les méthodes locales, dans le classement de Scharstein et Szeliski (nous abordons ce classement dans le chapitre 2, au paragraphe 2.8). Cet algorithme illustre le type d'algorithme utilisé actuellement dans le cadre des méthodes locales.

Pour chaque passage faire

- 2.1 Calculer le gradient des images
- 2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$
- 2.3 Pour chaque $\mathcal{S}^{i,j}$ faire
 - A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire
 - Calculer $E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v})$ en tenant compte de $ZA(\mathbf{p}_g^{i,j})$, en utilisant SAD et NCC, cf. équations (3.19) et (3.15), pondérées en fonction du gradient de l'image et en utilisant la couleur
 - F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA
- 2.4 Détecter les occultations en utilisant un seuil sur le coût local
Effectuer un calcul des disparités au sous-pixel

ALGO. 1.3 – *Méthode de mise en correspondance de Mayer* – Cet algorithme est une instance de l'algorithme 1.2. Mayer et al. n'utilisent pas une approche multirésolution, c'est la raison pour laquelle, les étapes 1 et 2.5 sont absentes et la boucle 2 a disparu.

1.9.3 Méthodes globales

<p><u>Pour</u> chaque passage <u>faire</u></p> <ol style="list-style-type: none"> 1. Construire la pyramide d'images 2. <u>Pour</u> chaque niveau de la pyramide <u>faire</u> <ol style="list-style-type: none"> 2.1. Prétraiter les images 2.2. $\mathcal{S}^i = \{\mathbf{p}_g^{i,j}\}_{i=0..N_g^{\text{lig}}-1, j=0..N_g^{\text{col}}-1}$ (1 support) ou $\mathcal{S}^i = \{\mathbf{p}_g^{i,j}\}_{j=0..N_g^{\text{col}}-1}$ (N_g^{lig} supports) ou $\mathcal{S}^i = \{\mathbf{p}_g^{i',j}\}_{i'=i-1..i+1, j=0..N_g^{\text{col}}-1}$ (N_g^{lig} supports) 2.3. <u>Pour</u> chaque \mathcal{S}^i <u>faire</u> <ol style="list-style-type: none"> 2.3.1. <u>Pour</u> chaque itération t <u>faire</u> <ol style="list-style-type: none"> A. <u>Pour</u> chaque $\mathbf{p}_g^{i,j} \in \mathcal{S}^i$ <u>faire</u> <u>Pour</u> chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ <u>faire</u> Calculer $f_{\text{dissimilarité}}^t(I_g^{i,j} - I_d^{i,v})$ B. Calculer $E_{\text{correspondance}}^t(d)$ sur \mathcal{S}^i C. <u>Pour</u> chaque $\mathbf{p}_g^{i,j} \in \mathcal{S}^i$ <u>faire</u> <u>Pour</u> chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ <u>faire</u> <u>Pour</u> chaque contrainte c <u>faire</u> Calculer $E_{\text{voisinage}}^{c,t}(\mathbf{p}_g^{i,j})$ en tenant compte de $ZAC^c(\mathbf{p}_g^{i,j})$ D. Calculer $E_{\text{contrainte}}^t(d)$ sur \mathcal{S}^i E. Calculer $E_{\text{global}}^t(d)$ sur \mathcal{S}^i F. Déterminer les correspondances sur \mathcal{S}^i selon la méthode d'optimisation 2.4. Affiner les résultats 2.5. Propager les résultats
--

ALGO. 1.4 – Méthode globale de mise en correspondance – Cet algorithme est une instance de l'algorithme 1.1. Les notations utilisées sont celles présentées dans le paragraphe 1.9.3. Pour les différents coûts, nous ajoutons l'indice t qui correspond à l'itération courante de la méthode d'optimisation.

Dans le cas des méthodes globales (cf. algorithme 1.4), la zone d'agrégation dans le coût local est réduite à un pixel et l'équation (1.15) devient :

$$E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) = E_{\text{dissimilarité}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}). \quad (1.21)$$

Ainsi, le coût global, équation (1.18), devient :

$$E_{\text{global}}(d) = \sum_{\mathbf{p}_g^{i,j} \in \mathcal{S}} \left((1 - \lambda) E_{\text{dissimilarité}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) + \lambda \sum_{c=0}^{N_c-1} \sum_{\mathbf{p}_g^{i',j'} \in ZAC^c(\mathbf{p}_g^{i,j})} E_{\text{lissage}}^c(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i',j'}) \right). \quad (1.22)$$

Le support \mathcal{S} peut être constitué de :

- *L'image entière* – Cela signifie que $\text{card}(\mathcal{S}) = N_g$ et qu'il y a une seule optimisation à réaliser [Anandan 89, Black 93, Black 96, Boykov 01, Gong 02, Veksler 03].

- *Les droites épipolaires* – Lorsque les images sont rectifiées, les correspondants des pixels appartenant à une même ligne de l'image de gauche appartiennent à la même ligne dans l'image de droite, ainsi E_{global} n'est minimisée que N_g^{lig} fois et $\text{card}(\mathcal{S}) = N_g^{\text{col}}$. De nombreuses méthodes [Park 01, Scharstein 02, Agrawal 04, Forstmann 04] s'appuient sur ce principe car cela permet de réduire considérablement le nombre de calculs.
- *Les droites épipolaires, ainsi que les droites épipolaires adjacentes* – Les droites épipolaires adjacentes sont les deux droites qui se situent respectivement au dessus et au dessous de la droite épipolaire considérée. Cette technique, introduite dans [Ohta 85] et employée dans [Belhumeur 92, Roy 98, Bobick 99, Gong 05, Kim 05], est aussi appelée mise en correspondance intra et inter-droites épipolaires (*intra and inter-scanline matching*). Dans ce cas, nous avons $\text{card}(\mathcal{S}) = 3N_g^{\text{col}}$ et il y a N_g^{lig} minimisations à réaliser.

Nous allons à présent détailler les coûts de dissimilarité et les coûts des contraintes utilisés.

1.9.3.1 Coût de dissimilarité

Tous les coûts de dissimilarité que nous avons pu relever s'appuient sur une fonction de la différence des niveaux de gris [Deng 05] et même dans le cas de la couleur, dans [Scharstein 02], la moyenne des trois composantes couleur est préalablement calculée, ce qui revient à utiliser des niveaux de gris. Ainsi, dans tous les cas que nous citons dans ce paragraphe, nous avons :

$$E_{\text{dissimilarité}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) = f_{\text{dissimilarité}}(I_g^{i,j} - I_d^{i,v}).$$

Les différentes formes de ce coût s'appuient sur :

- *L'utilisation de la fonction exponentielle* – Elle est utilisée dans [Tao 00, Tao 01] et le coût de dissimilarité est donné par :

$$f_{\text{dissimilarité}}(I_g^{i,j} - I_d^{i,v}) = e^{\frac{(I_g^{i,j} - I_d^{i,v})^2}{2\sigma^2}}, \quad (1.23)$$

où σ est une constante.

- *La valeur absolue des différences des niveaux de gris* – Elle est utilisée dans [Bovyrin 03, Veksler 05] et donnée par :

$$f_{\text{dissimilarité}}(I_g^{i,j} - I_d^{i,v}) = |I_g^{i,j} - I_d^{i,v}|. \quad (1.24)$$

- *Le carré de la différence des niveaux de gris* – La définition de cette différence est employée dans [Barnard 89, Belhumeur 92, Ishikawa 98, Boykov 01, Kolmogorov 01, Park 01, Kolmogorov 02] et elle est donnée par l'équation (1.19).
- *La dissimilarité s'appuyant sur les niveaux de gris interpolés linéairement autour des deux pixels mis en correspondance* – Au cours de l'échantillonnage de l'image, un même pixel correspond à des points différents de la scène. Ce phénomène entraîne des erreurs, parfois importantes, sur le calcul des coûts de dissimilarité. La technique proposée dans [Birchfield 98] a pour but de proposer un coût de dissimilarité insensible à l'échantillonnage de l'image. Elle est utilisée aussi dans [Veksler 03] et repose sur la définition suivante :

$$f_{\text{dissimilarité}}(I_g^{i,j} - I_d^{i,v}) = \min \left(\min_{v-\frac{1}{2} < v' < v+\frac{1}{2}} |I_g^{i,j} - I_d^{i,v'}|, \min_{j-\frac{1}{2} < j' < j+\frac{1}{2}} |I_g^{i,j'} - I_d^{i,v}| \right). \quad (1.25)$$

Dans ce cas, il faut donc calculer les niveaux de gris interpolés $I_g^{i,j'}$ et $I_d^{i,v'}$ donnés par $I_l^{i,j+t} = (I_l^{i,j+1} - I_l^{i,j})t + I_l^{i,j}$ avec $j' = j + t$.

1.9.3.2 Coût des contraintes

La zone d'agrégation des contraintes, ZAC, peut être :

- *Les quatre voisins* – Il s'agit du voisinage considéré dans [Barnard 89, Goulermas 03, Bleyer 04].
- *Le pixel précédent* – Cela correspond au pixel adjacent au pixel étudié sur la même ligne et/ou au pixel adjacent sur la même colonne. Il s'agit du voisinage considéré dans [Agrawal 04].

Le coût de lissage E_{lissage}^c a pour but de pénaliser les disparités différentes des disparités du voisinage [Barnard 89, Belhumeur 92, Gong 04]. Il peut prendre les deux formes suivantes :

- *Comparaisons avec les disparités voisines* :

$$E_{\text{lissage}}^c(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'}) = f_{\text{lissage}}^c(\|d(\mathbf{p}_g^{i,j}) - d(\mathbf{p}_g^{i',j'})\|). \quad (1.26)$$

Dans [Roy 98, Veksler 99, Agrawal 04], la fonction f_{lissage} peut être la valeur absolue, la valeur absolue tronquée ou une fonction constante par morceaux.

- *Termes de pénalité* :

$$E_{\text{lissage}}^c(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'}) = \begin{cases} \lambda_c & \text{si } \mathbf{p}_g^{i,j} \text{ et } \mathbf{p}_g^{i',j'} \text{ satisfont le critère } c \\ 0 & \text{sinon.} \end{cases} \quad (1.27)$$

Dans ce cas, une première estimation des disparités est préalablement effectuée. Nous pouvons distinguer :

- Dans [Bobick 99, Kolmogorov 02, Bovyrin 03], les discontinuités de disparités sont utilisées pour détecter les pixels occultés, ainsi, le critère c utilisé est : $\|d(\mathbf{p}_g^{i,j}) - d(\mathbf{p}_g^{i',j'})\| > S_o$ avec S_o un seuil à fixer.
- Dans [Ohta 85, Ishikawa 98, Tsai 99, Lee 02, Zhang 02, Sun 03, Wei 04], les auteurs exploitent une segmentation en régions, les pixels des frontières des régions sont considérés comme des pixels occultés [Tao 00, Tao 01, Bleyer 04]. Le critère c utilisé est : si $\mathbf{p}_g^{i,j}$ est occulté.
- Dans [Nasrabadi 92b, Yang 93, Goulermas 03, Kolmogorov 01] (cf. paragraphe 1.6.5), les correspondances considérées ne respectent pas la contrainte d'unicité.
- Dans [Cox 95, Goulermas 03] (cf. paragraphe 1.6.6), les correspondances considérées ne respectent pas la contrainte d'ordre.
- Dans [Luo 95], les correspondances considérées ne respectent pas la contrainte de symétrie (cf. paragraphe 1.6.7).

1.9.3.3 Exemple de méthode globale

L'algorithme 1.5 est celui de Sun *et al.* [Sun 05] dont la méthode a obtenu les meilleurs résultats dans l'évaluation de Scharstein et Szeliski (nous aborderons cette évaluation dans le chapitre 2, au paragraphe 2.8). Cet algorithme illustre le type d'algorithme utilisé actuellement dans le cadre des méthodes globales.

1.9.4 Méthodes mixtes

Il existe aussi des méthodes qui correspondent à une combinaison d'une méthode locale et d'une méthode globale et que nous appelons méthodes mixtes. Le coût global est alors défini par l'équation générale (1.18).

Pour chaque passage faire

2.1 Segmentation couleur par *mean-shift* des images

2.2 $\mathcal{S}^i = \{\mathbf{p}_g^{i,j}\}_{j=0..N_g^{\text{col}}-1}$ (N_g^{lig} supports)

2.3 Pour chaque \mathcal{S}^i faire

2.3.1 Pour chaque itération t faire

A. Pour chaque $\mathbf{p}_g^{i,j} \in \mathcal{S}^i$ faire

Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire

Calculer $f_{\text{dissimilarité}}^t(I_g^{i,j} - I_d^{i,v}) = (I_g^{i,j} - I_d^{i,v})^2$

B. Calculer $E_{\text{correspondance}}^t(d)$ sur \mathcal{S}^i

C. Pour chaque $\mathbf{p}_g^{i,j} \in \mathcal{S}^i$ faire

Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire

Pour chaque contrainte c faire

Calculer $E_{\text{voisinage}}^{c,t}(\mathbf{p}_g^{i,j})$ en tenant compte de $ZAC^c(\mathbf{p}_g^{i,j})$

(ce terme comprend un coût de lissage et un coût de visibilité)

D. Calculer $E_{\text{contrainte}}^t(d)$ sur \mathcal{S}^i

E. Calculer $E_{\text{global}}^t(d)$ sur \mathcal{S}^i

F. Déterminer les correspondances sur \mathcal{S}^i selon la méthode de propagation de croyance (cf. paragraphe 1.10.7)

Les différents passages :

Passage 0 -- Estimation des occultations dans I_g en fonction des disparités estimées dans I_d

Passage 1 -- Estimation des disparités en tenant compte des occultations détectées

ALGO. 1.5 – *Méthode de mise en correspondance de Sun et al.* – *Cet algorithme est une instance de l'algorithme 1.4. Il est à passages multiples. Il n'utilise pas d'approche multirésolution, c'est pourquoi, les étapes 1 et 2.5 sont absentes et la boucle 2 a disparu. De plus, il ne contient pas d'étape d'affinement (disparition de l'étape 2.4).*

1.9.4.1 Coût local

Dans la plupart des cas, le coût local, qui est fonction des coûts de dissimilarité, correspond à une mesure de corrélation, [Black 93, Le Besnerais 98, Rziza 00, Scharstein 02, Shao 02, Sun 02, Gong 02, Criminisi 03, Leung 04, Torr 04, Agrawal 04]. Dans [Broekers 04b], pour obtenir une mesure de corrélation plus générale (c'est-à-dire qui prend en compte davantage de caractéristiques que les niveaux de gris seuls), Broekers *et al.* ajoutent, dans le coût de dissimilarité entre deux pixels, un terme qui mesure la dissimilarité des gradients et un autre qui mesure la dissimilarité de la texture.

1.9.4.2 Coût des contraintes

Les contraintes utilisées sont, entre autres, les mêmes que celles décrites dans le paragraphe 1.9.3.2. La mesure de corrélation intervient parfois dans les contraintes utilisées : Anandan [Anandan 89] ajoute une contrainte qui correspond à une mesure de confiance qui s'appuie sur la variance des scores de corrélation dans le voisinage du pixel considéré.

1.9.4.3 Exemple de méthode mixte

L'algorithme 1.6 est celui de Hong et Chen [Hong 04] dont la méthode a obtenu les meilleurs résultats parmi les méthodes mixtes dans l'évaluation de Scharstein et Szeliski (nous aborderons cette évaluation dans le chapitre 2, au paragraphe 2.8). Cet algorithme illustre le type d'algorithme utilisé actuellement dans le cadre des méthodes mixtes.

<p><u>Pour</u> chaque passage <u>faire</u></p> <p>2.1 Segmentation couleur par <i>mean-shift</i> des images</p> <p>2.2 $\mathcal{S}^i = \{\mathbf{p}_g^{i,j}\}_{j=0..N_g^{\text{col}}-1}$, $k = 0..N_g^{\text{lig}} - 1$ (N_g^{lig} supports)</p> <p>2.3 <u>Pour</u> chaque \mathcal{S}^i <u>faire</u></p> <p>2.3.1 <u>Pour</u> chaque itération t <u>faire</u></p> <p>A. <u>Pour</u> chaque $\mathbf{p}_g^{i,j} \in \mathcal{S}^i$ <u>faire</u> <u>Pour</u> chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ <u>faire</u> Calculer $E_{\text{local}}^t(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v})$, si passage 0 en utilisant la mesure SAD, cf. équation (3.19)</p> <p>B. <u>Si</u> passage 1 <u>alors</u> calculer $E_{\text{correspondance}}^t(d)$ sur \mathcal{S}^i</p> <p>C. <u>Pour</u> chaque $\mathbf{p}_g^{i,j} \in \mathcal{S}^i$ <u>faire</u> <u>Pour</u> chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ <u>faire</u> <u>Pour</u> chaque contrainte c <u>faire</u> <u>Si</u> passage 1 <u>alors</u> Calculer $E_{\text{voisinage}}^{c,t}(\mathbf{p}_g^{i,j})$ en tenant compte des $ZAC^c(\mathbf{p}_g^{i,j})$ (ce terme comprend un coût de lissage)</p> <p>D. <u>Si</u> passage 1 <u>alors</u> calculer $E_{\text{contrainte}}^t(d)$ sur \mathcal{S}^i</p> <p>E. <u>Si</u> passage 1 <u>alors</u> calculer $E_{\text{global}}^t(d)$ sur \mathcal{S}^i</p> <p>F. <u>Si</u> passage 0 <u>alors</u> déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA <u>Sinon</u> déterminer les correspondances sur le support selon la méthode de coupure de graphe (cf. paragraphe 1.10.5)</p>
<p><i>Les différents passages :</i></p> <p>Passage 0 -- Utilisation d'une méthode locale pour initialiser les disparités</p> <p>Passage 1 -- Utilisation d'une méthode globale</p>

ALGO. 1.6 – Méthode de mise en correspondance de Hong et Chen – Cet algorithme est une instance de l'algorithme 1.1. Il est à passages multiples. Il n'utilise pas d'approche multirésolution, c'est la raison pour laquelle, les étapes 1 et 2.5 sont absentes et la boucle 2 a disparu. Il ne contient pas d'étape d'affinement (disparition de l'étape 2.4).

1.10 Méthodes d'optimisation

La forme du coût global de mise en correspondance étant choisie, il faut trouver la méthode pour minimiser ce coût. Nous pouvons distinguer les méthodes à variables continues, c'est-à-dire effectuant un appariement au sous-pixel, et dans ce cas, la disparité estimée est réelle, des méthodes à variables discrètes, c'est-à-dire effectuant un appariement au sous-pixel et, dans ce cas, la disparité estimée est entière. Parmi les méthodes à variables continues, nous pouvons citer les méthodes par descente de gradient [Wei 98, Xiao 03] et, plus précisément, les méthodes modélisant le problème de la mise en correspondance par des équations aux dérivées partielles [Gennert 88, Gökstorp 95, Robert 96, Alvarez 00, Hermsillo Valadez 02, Maier 03, Devernay 05], les méthodes par décomposition en ondelettes [Pan 96] et les méthodes par corrélation de phase [Fleet 91, Smith 96]. Dans les cas des méthodes à variables continues, il est nécessaire d'effectuer une interpolation des niveaux de gris, étape délicate qui peut introduire de des erreurs, cf. [Schreier 00].

Dans ce mémoire, nous nous intéressons aux méthodes à variables discrètes car, en vision par ordinateur, elles sont les plus utilisées. Nous nous plaçons dans le cas d'une optimisation discrète et l'espace des solutions est fini car nous utilisons une zone de recherche bornée. La méthode la plus naïve est la recherche gloutonne qui consiste à effectuer une recherche exhaustive dans l'ensemble des solutions. Dans le cas des méthodes locales, la recherche exhaustive, *winner take all*, a une complexité en $N_z^{N_g}$ où N_z est le nombre de pixels considérés dans la zone de recherche et N_g est le nombre de pixels considérés dans l'image gauche. Dans le cas des méthodes globales, la recherche exhaustive a une complexité en $N_d^{N_g}$ où N_d est le nombre de pixels considérés dans l'image droite. En effectuant une recherche exhaustive, les méthodes locales sont donc coûteuses mais beaucoup moins que si on utilisait la recherche exhaustive avec les méthodes globales, car on peut dire que N_z est petit comparé à N_d . La recherche exhaustive pour les méthodes globales n'est pas du tout populaire, même dans le cas où on utilise la contrainte épipolaire, il y a une complexité en $N_d^{\text{col}N_g}$. Cependant, certains auteurs se sont permis de l'utiliser dans des cas très particuliers où l'espace des solutions a été préalablement très réduit. C'est le cas dans [Tao 00, Tao 01, Bleyer 04] où les auteurs effectuent une segmentation en régions de l'image puis considèrent sur chaque droite épipolaire les segments de pixels connexes appartenant à la même région.

1. Initialiser les disparités (fonction d)
 2. Calculer l'énergie globale initiale $E_{\text{global}}^0(d)$
 3. Répéter
 - 3.1. Mettre à jour la fonction d
 - 3.2. Mettre à jour l'énergie globale, $E_{\text{global}}^t(d)$
 - 3.3. Calculer la variation d'énergie $\Delta E_{\text{global}}(d)$
- Jusqu'à ($t > N_i$) ou ($\Delta E_{\text{global}}(d) < T_o$)

ALGO. 1.7 – Méthode générique d'optimisation pour la mise en correspondance – Cet algorithme générique des méthodes d'optimisation pour la mise en correspondance est itératif et son objectif est de minimiser $E_{\text{global}}(d)$. Le terme N_i est le nombre d'itérations maximal, t est l'itération courante et T_o est un seuil à fixer.

Depuis 1985, de nombreuses méthodes de mise en correspondance ont utilisé des méthodes d'optimisation. La première méthode d'optimisation utilisée dans le cadre de la mise en correspondance, et, en même temps, la plus populaire est la programmation dynamique [Ohta 85, Cox 92, Geiger 95, Koch 98, Rziza 00, Petrakis 02, Criminisi 03, Zickler 03, Forstmann 04, Veksler 05]. Parmi toutes les publications que nous avons recensées et qui utilisent une technique d'optimisation, la moitié des

méthodes proposées s'appuient sur une optimisation par programmation dynamique. Des techniques aussi anciennes que la programmation dynamique ont été utilisées de manière plus marginale, comme la relaxation [Anandan 89, Nasrabadi 92a, Brockers 04b] ou le recuit simulé [Barnard 89, Lee 02]. Une technique plus récente qui tend à se développer depuis 1998 (cf. tableau 2.1 du chapitre 2) concerne les coupures de graphes [Ishikawa 98, Veksler 99, Kolmogorov 01, Lin 02, Agrawal 04, Deng 05]. Enfin, l'utilisation d'autres techniques pour la mise en correspondance reste très ponctuelle et très récente, comme les réseaux de neurones utilisés dans [Marr 76, Nasrabadi 92b, Wang 99, Gong 05], voire très marginale, comme les algorithmes génétiques employés dans [Saito 95, Goulermas 03] ou l'algorithme de propagation de croyance [Tappen 03, Sun 03, Sun 05]. Les méthodes d'optimisation sont toutes itératives et suivent l'algorithme général 1.7. Il faut souligner que toutes ces méthodes font appel à la même modélisation de départ : les champs de Markov aléatoires, *Markov Random Fields*. C'est pourquoi, avant de décrire les différentes techniques d'optimisation, nous allons définir le modèle des champs de Markov.

1.10.1 Modèle des champs de Markov

Le modèle des champs de Markov est très utilisé en analyse d'images et en vision par ordinateur, cf. [Descombes 04]. Il constitue souvent le modèle *a priori* dans un cadre bayésien permettant d'apporter des connaissances sur la solution au problème posé. La solution recherchée est généralement une réalisation du champ de Markov correspondant au maximum *a posteriori* (MAP). La recherche du maximum de ce critère est faite grâce à une méthode d'optimisation.

Concernant la mise en correspondance, les champs de Markov sont utilisés par plusieurs auteurs qui font appel à des méthodes d'optimisation différentes :

- Dans [Luo 95], la méthode de relaxation est employée.
- Dans [Boykov 01, Tappen 03], les auteurs utilisent des coupures de graphes.
- Dans [Gong 02], une technique de programmation dynamique est utilisée.
- Dans [Le Besnerais 98, Lee 02], les méthodes proposées s'appuient sur le recuit simulé.
- Dans [Sun 03, Sun 05], les auteurs s'appuient sur la propagation de croyance.

1.10.2 Programmation dynamique

Bellman et Dreyfus [Bellman 62] sont les fondateurs de la programmation dynamique. Nous avons répertorié un très grand nombre d'articles utilisant la programmation dynamique [Belhumeur 92, Intille 94, Cox 95, Dhond 95, Cox 96, Birchfield 98, Bobick 99, Tsai 99, Fielding 00, Park 01, Gong 02, Sun 02, Zhang 02, Bovyrin 03, Criminisi 03, Gong 03, Veksler 03, Gong 04, Leung 04, Torr 04, Gong 05, Kim 05]. En vision par ordinateur, c'est l'algorithme de Viterbi [Viterbi 67, Forney 73], initialement développé dans le domaine du codage d'information, qui est exploité pour mettre en œuvre des méthodes de programmation dynamique. L'algorithme de Viterbi est vu comme une solution de l'estimation du MAP dans un modèle des champs de Markov. La programmation dynamique pour la mise en correspondance, présentée notamment dans [Zhang 93] et [Horaud 93, p. 203–204], est appliquée entre les deux séquences de pixels sur les droites épipolaires conjuguées. Les indices de colonnes des pixels de ces deux séquences sont les indices d'une matrice \mathbf{M}_P , de taille $N_g^{\text{col}} \times N_d^{\text{col}}$ qui contient les coûts des correspondances possibles entre les pixels. Cet algorithme s'appuie sur deux contraintes : la contrainte d'ordre et la contrainte d'unicité. Le but est de trouver un chemin optimal du pixel de colonne 0 au pixel de colonne $N_g^{\text{lig}} - 1$ (cf. figure 1.7). L'algorithme de programmation

dynamique se compose de deux étapes :

1. Le calcul des éléments de la matrice \mathbf{M}_P qui est défini dans [Criminisi 03], par :

$$M_P(j,v) = \min \left\{ \begin{array}{l} M_P(j-1,v) + \beta, \\ M_P(j-1,v-1) + E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}), \\ M_P(j,v-1) + \beta \end{array} \right\}, \quad (1.28)$$

où β est une constante à définir. Le premier terme de l'opérateur min correspond au cas où $\mathbf{p}_g^{i,j}$ est occulté. Le second terme correspond au cas où une correspondance est établie entre $\mathbf{p}_g^{i,j}$ et $\mathbf{p}_d^{i,v}$. Le troisième terme correspond au cas où il y a une discontinuité de profondeur (c'est-à-dire où le pixel $\mathbf{p}_d^{i,v}$ est occulté).

2. Le calcul du chemin optimal dans la matrice \mathbf{M}_P , en la parcourant de l'élément $M_P(N_g^{\text{lig}} - 1, N_d^{\text{lig}} - 1)$ à l'élément $(0,0)$.

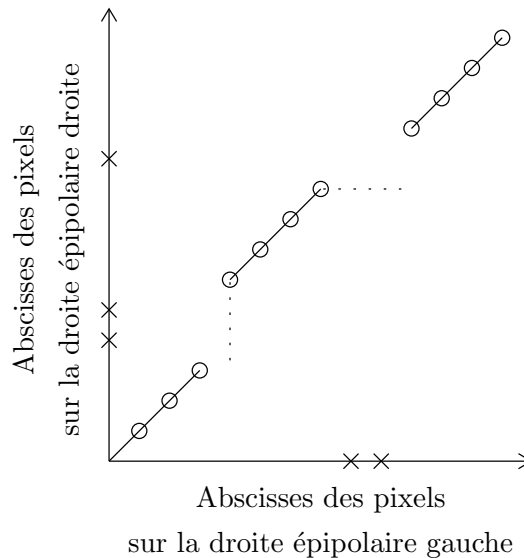


FIG. 1.7 – Programmation dynamique – Il s'agit d'un exemple de chemin optimal associé à la matrice \mathbf{M}_P des coûts cumulés. Les symboles utilisés sur la figure sont ceux présentés dans le tableau 1.1 et les pointillés correspondent à des discontinuités de profondeur.

1.10.3 Recuit simulé

Une des manières de rechercher la réalisation du champ de Markov correspondant au MAP consiste à faire appel aux méthodes du type Monte-Carlo par chaînes de Markov (MCMC). L'algorithme du recuit simulé s'inscrit dans ce cadre. Il consiste à tirer au sort une nouvelle configuration. Si celle-ci fait diminuer l'énergie à minimiser (liée au MAP), alors elle est acceptée, sinon, elle est acceptée avec une probabilité paramétrée par une température (distribution tempérée). Cette probabilité est d'autant plus grande que la température est élevée. Cette technique est appliquée de manière itérative en partant d'une configuration aléatoire et d'une température élevée, puis en diminuant progressivement la température jusqu'à convergence.

Deux variantes sont utilisées : l'échantillonneur de Gibbs et l'algorithme de Metropolis. Dans le cadre de la mise en correspondance, les publications les plus significatives qui font appel au recuit simulé

sont [Barnard 89, Belhumeur 92, Starink 95, Le Besnerais 98, Lee 02]. L'algorithme (1.8) présente les étapes du recuit simulé pour la mise en correspondance. Il s'appuie sur le calcul d'une probabilité qui dépend d'une température. Le calcul de l'énergie est itéré jusqu'à ce que l'équilibre soit atteint, c'est-à-dire que cette probabilité tende vers 1. L'avantage majeur de cette technique est que l'initialisation n'influence pas le résultat. L'inconvénient est la difficulté pour ajuster la température, c'est-à-dire, pour l'initialiser puis pour la faire évoluer de manière à converger vers le minimum global en un temps raisonnable. En effet, en théorie, la température doit baisser très lentement, suivant une décroissance logarithmique, mais, en pratique, on utilise une décroissance géométrique dont la raison n'est pas simple à régler. De plus, les méthodes de mise en correspondance utilisant le recuit simulé sont coûteuses en temps de calcul. Dans [Barnard 89, Lee 02], les auteurs ont toutefois proposé une approche multirésolution (cf. paragraphe 1.13).

1. Initialiser les correspondances et initialiser la température T^0

2. Répéter

2.1. Effectuer un changement sur une correspondance

2.2. Diminuer la température

2.3. Calculer la variation d'énergie

$$\Delta E = E_{\text{global}}^t(d) - E_{\text{global}}^{t-1}(d) \quad (1.29)$$

2.4. Si $\Delta E < 0$ alors accepter le changement

2.5. Si $\Delta E \geq 0$ alors accepter le changement avec la probabilité $e^{\left(\frac{-\Delta E}{kT^t}\right)}$

Jusqu'à obtenir un équilibre

ALGO. 1.8 – Recuit simulé – Il s'agit de l'algorithme général du recuit simulé suivant la technique de Metropolis [Metropolis 53].

1.10.4 Relaxation

Davis et Rosenfeld [Davis 81] sont les premiers à avoir proposé une technique de relaxation en vision par ordinateur. Une fois de plus, cette méthode peut être vue comme une solution à un problème modélisé par des champs de Markov [Luo 95]. Dans [Anandan 89, Luo 95, Black 96, Gouet 00, Shao 02, Brockers 04b], des méthodes de mise en correspondance utilisant le principe de la relaxation ont été proposées. Le principe de la relaxation, exposé dans [Jones 97, Zhang 93] et [Horaud 93, p. 204–207] (cf. algorithme 1.9), est le suivant :

À chaque pixel $\mathbf{p}_g^{i,j}$, un ensemble de correspondances possibles $\{\mathbf{p}_d^{i,v}\}$ est associé et à chaque pixel de cet ensemble une mesure de confiance (probabilité) $\text{Prob}(\mathbf{p}_d^{i,v}/\mathbf{p}_g^{i,j})$ est attribuée. De manière itérative, cet ensemble évolue en tenant compte des pixels voisins et de certaines contraintes.

Un exemple de mesure de confiance est celle donnée dans [Horaud 93, p. 205] :

$$\text{Prob}(\mathbf{p}_d^{i,v}/\mathbf{p}_g^{i,j}) = \sum_{\mathbf{p}_g^{i',j'} \in \text{ZAC}^c(\mathbf{p}_g^{i,j})} \max_{\mathbf{p}_d^{i',v'} \in \mathcal{P}_d} \frac{E_{\text{local}}(\mathbf{p}_g^{i',j'}, \mathbf{p}_d^{i',v'}) G_d(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}, \mathbf{p}_g^{i',j'}, \mathbf{p}_d^{i',v'})}{f(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'})}, \quad (1.30)$$

où \mathcal{P}_d est l'ensemble des pixels correspondants possibles du pixel $\mathbf{p}_g^{i',j'}$ et $f(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'})$ est une fonction proportionnelle à la distance entre $\mathbf{p}_g^{i,j}$ et $\mathbf{p}_g^{i',j'}$. Le terme G_d est le gradient de disparité, cf. équation (1.10). L'algorithme de relaxation nécessite une bonne initialisation. De plus, pour obtenir des temps de calcul raisonnables, dans [Luo 95, Shao 02], une approche multirésolution a été proposée

(cf. paragraphe 1.13). Parmi ces méthodes, celles proposées dans [Anandan 89, Shao 02, Brockers 04b] sont des méthodes mixtes.

1. Initialisation

1.1. Calculer $E_{\text{local}}(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}) \quad \forall i, j, v$

1.2. Construire l'ensemble $\mathcal{X} = \{n_g^{i,j}\}_{i=0..N_d^{\text{lig}}-1, j=0..N_d^{\text{col}}-1}$ où $n_g^{i,j}$ un noeud associé au pixel $\mathbf{p}_g^{i,j}$ contenant l'ensemble des couples $(\mathbf{p}_d^{i,v}, \text{Prob}(\mathbf{p}_d^{i,v} / \mathbf{p}_g^{i,j}))$ (cf. équation (1.30))

2. Répéter

2.1. Mémoriser les correspondances ayant la meilleure probabilité

2.2. Mettre à jour l'ensemble \mathcal{X} , c'est-à-dire supprimer tous les noeuds faisant intervenir les pixels dans les correspondances mémorisées

Jusqu'à $\mathcal{X} = \emptyset$

ALGO. 1.9 – *Relaxation* – Il s'agit de l'algorithme général des méthodes de mise en correspondance qui utilisent une technique de relaxation et se qui distinguent par le choix de $\text{Prob}(\mathbf{p}_d^{i,v} / \mathbf{p}_g^{i,j})$.

1.10.5 Coupure de graphe

Récemment, des travaux s'appuient sur l'utilisation de flot de graphe et plus précisément sur des techniques de coupure de graphe [Ishikawa 98, Birchfield 99, Veksler 99, Ishikawa 00, Banks 01, Kang 01, Paris 02, Lin 02, Hong 04, Deng 05]. Introduite par Roy et Cox [Roy 98], l'utilisation de flot de graphe en vision par ordinateur a été formalisée par Veksler [Veksler 99] et développée dans [Boykov 01, Kolmogorov 01, Kolmogorov 02] sous la forme d'un problème d'étiquetage visant à attribuer une disparité à chaque pixel de l'image. Tappen et Freeman [Tappen 03] ont montré que l'algorithme des coupures de graphes permet aussi de proposer une solution optimale à un problème modélisé par des champs de Markov.

Le problème du flot de graphe peut être symbolisé par un écoulement d'eau dans un réseau de tuyaux, c'est-à-dire qu'étant donné une source et un puits, on cherche le flot maximum qui peut passer dans le réseau. Les méthodes de coupure de graphe utilisent un graphe, noté \mathcal{G} , qui symbolise le réseau, composé d'un ensemble de sommets \mathcal{S}_o , qui correspondent aux connexions, et d'un ensemble d'arcs orientés $\mathcal{A} \in \mathcal{S}_o^2$ qui représentent les tuyaux. La formulation théorique repose sur les définitions suivantes :

- Deux sommets particuliers sont définis : la source s et le puits p .
- Pour un sommet x , on distingue les arcs entrants $\mathcal{A}_+(x)$ et les arcs sortants $\mathcal{A}_-(x)$.
- La capacité, $\text{cap}(a) : \mathcal{A} \rightarrow \mathbb{R}$, et le flot, $\text{flot}(a) : \mathcal{A} \rightarrow \mathbb{R}$, d'un arc a sont tels que :

$$\begin{aligned} \forall a \in \mathcal{A} \quad \text{flot}(a) &\leq \text{cap}(a) \\ \forall x \in \mathcal{S}_o - \{s, p\} \quad \sum_{a_+ \in \mathcal{A}_+(x)} \text{flot}(a_+) &= \sum_{a_- \in \mathcal{A}_-(x)} \text{flot}(a_-). \end{aligned} \quad (1.31)$$

- Un flot de graphe est la somme des flots des arcs :

$$\text{flot}(\mathcal{G}) = \sum_{a \in \mathcal{A}_-(s)} \text{flot}(a). \quad (1.32)$$

- Une coupure $\mathcal{C}_{\mathcal{G}}$, d'un graphe \mathcal{G} , est une partition de l'ensemble des sommets \mathcal{S}_o , en deux ensembles connexes, \mathcal{S}_s et \mathcal{S}_p , tels que $s \in \mathcal{S}_s$ et $p \in \mathcal{S}_p$. Une valeur $\text{coup}(\mathcal{C}_{\mathcal{G}})$ est associée à cette coupure :

$$\text{coup}(\mathcal{C}_{\mathcal{G}}) = \sum_{(x,y) \in \mathcal{S}_s \times \mathcal{S}_p} \text{cap}(x,y) \text{ où } (x,y) \text{ est un arc coupé.} \quad (1.33)$$

- Le théorème du flot maximal et de la coupure minimale est le suivant : le flot maximal est atteint pour la coupure minimale :

$$\underset{\mathcal{G}}{\text{argmax}} \text{flot}(\mathcal{G}) = \underset{\mathcal{G}}{\text{argmin}} \text{coup}(\mathcal{C}_{\mathcal{G}}). \quad (1.34)$$

L'idée directrice des méthodes de mise en correspondance à base de coupure de graphe est de créer un graphe dont la coupure représente le coût de mise en correspondance globale. Chaque sommet correspond au couple $(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v})$, la source et le puits correspondent respectivement aux pixels qui sont les projections des points les plus proches et les plus éloignés du plan image (ce qui correspond respectivement à la disparité maximale et la disparité minimale). Les sommets sont 4-connexes car ils sont reliés au puits, à la source et aux deux sommets voisins qui correspondent aux couples de correspondances obtenus pour les deux pixels voisins dans l'image gauche sur la droite épipolaire (dans [Roy 98], des sommets 6-connexes qui considèrent en plus deux voisins sur les deux lignes épipolaires adjacentes sont utilisés). La capacité d'un arc correspond au coût de mise en correspondance associé aux deux sommets qu'il relie.

Un des avantages de cette méthode est de pouvoir l'appliquer sur toute l'image avec des temps de calcul raisonnables, contrairement à la programmation dynamique qui est très coûteuse. En réalité, cette technique a été souvent utilisée, comme la programmation dynamique, sur les droites épipolaires conjuguées [Ishikawa 98, Lin 02, Agrawal 04].

1.10.6 Réseaux de neurones

Hopfield [Hopfield 82] a été l'un des premiers à aborder cette technique d'optimisation. Nous pouvons cependant aussi citer l'article de Marr et Piogo [Marr 76]. Bien qu'il ne fasse pas référence à des réseaux de neurones, la technique qu'ils proposent est très proche de la technique des réseaux de neurones. L'utilisation pour la mise en correspondance est beaucoup plus marginale que celles que nous avons déjà présentées. Très peu de publications ont été réalisées dans ce domaine [Nasrabadi 92b, Hu 93, Wang 99, Jahn 01]. La mise en correspondance est effectuée seulement sur des points d'intérêt préalablement détectés. Le but est de formuler le problème de la mise en correspondance de telle manière qu'il puisse être assimilé à un réseau de neurones de Hopfield. Pour cela, il faut que l'énergie à minimiser soit une fonction de Lyapunov qui représente le comportement collectif du réseau, avec des poids d'interconnexion neuronale qui correspondent aux contraintes imposées par la mise en correspondance. Lorsque le réseau est stable, la fonction d'énergie est à son minimum. Le principe du réseau de neurones est d'engendrer une décision coopérative entre tous les neurones, c'est-à-dire que chaque neurone reçoit des informations des autres neurones. Le réseau est représenté par une matrice de taille $N_g \times N_d$. L'état, $n_t^{j,v}$, à l'instant t , de chaque neurone (j,v) dans cette matrice prend une valeur binaire : 0 lorsqu'il n'y a pas de correspondance, 1 sinon. La fonction de Lyapunov, qui correspond à l'énergie totale, est la suivante :

$$E_{\text{global}}^t(d) = \frac{-1}{2} \sum_{j=0}^{N_g-1} \sum_{v=0}^{N_d-1} \sum_{j'=0}^{N_g-1} \sum_{v'=0}^{N_d-1} t^{j,v,j',v'} n_t^{j,v} n_t^{j',v'} - \sum_{j=0}^{N_g-1} \sum_{v=0}^{N_d-1} n_0^{j,v} n_t^{j,v}, \quad (1.35)$$

avec $t^{j,v,j',v'} = t^{j',v',j,v}$ le poids d'interconnexion entre deux neurones et $n_0^{j,v}$ la valeur initiale du neurone (j,v) . Un changement d'état $\Delta n_t^{j,v}$ du neurone (j,v) induit une variation d'énergie $\Delta E_{\text{local}}^t$

définie par :

$$\Delta E_{\text{local}}^t(\mathbf{P}_g^{i,j}, \mathbf{P}_d^{i,v}) = \left(\sum_{j'=0}^{N_g-1} \sum_{v'=0}^{N_d-1} t^{j,v,j',v'} n_t^{j',v'} + n_0^{j,v} \right) \Delta n_t^{j,v}. \quad (1.36)$$

L'équation suivante décrit l'évolution des neurones :

$$n_{t+1}^{j,v} = \begin{cases} 0 & \text{si } \left(\sum_{j'=0}^{N_g-1} \sum_{v'=0}^{N_d-1} t^{j,v,j',v'} n_t^{j',v'} + n_0^{j,v} \right) > 0 \\ 1 & \text{si } \left(\sum_{j'=0}^{N_g-1} \sum_{v'=0}^{N_d-1} t^{j,v,j',v'} n_t^{j',v'} + n_0^{j,v} \right) < 0 \\ n_t^{j,v} & \text{sinon.} \end{cases} \quad (1.37)$$

Ainsi, le coût de mise en correspondance globale, à l'instant t , est donné par :

$$E_{\text{global}}^t(d) = \sum_{j=0}^{N_g-1} \sum_{v=0}^{N_d-1} \sum_{j'=0}^{N_g-1} \sum_{v'=0}^{N_d-1} c_t^{j,v,j',v'} p_t^{j,v} \text{Prob}_t^{j',v'} + \sum_{j=0}^{N_g-1} \left(1 - \sum_{v=0}^{N_d-1} \text{Prob}_t^{j,v} \right)^2 + \sum_{v=0}^{N_d-1} \left(1 - \sum_{j=0}^{N_g-1} \text{Prob}_t^{j,v} \right)^2, \quad (1.38)$$

où $c_t^{j,v,j',v'}$ est une mesure de compatibilité qui dépend de la différence de disparités associées aux neurones (j,v) et (j',v') et de la différence de distance entre les pixels j et j' et les pixels v et v' . Les termes $\text{Prob}_t^{j,v}$ sont des probabilités dont les valeurs sont données par un coût local de mise en correspondance. Une fois de plus, cet algorithme a été proposé en utilisant la contrainte épipolaire et donc en réduisant la recherche des correspondants à la droite épipolaire associée. De plus, l'espace de recherche est encore plus réduit puisque seuls des points d'intérêt sont mis en correspondance.

1.10.7 Propagation de croyance

La propagation de croyance, *belief propagation*, a été introduite par Pearl [Pearl 88] dans le cadre de l'intelligence artificielle. Depuis, cette technique a été reprise en vision par ordinateur, essentiellement dans [Weiss 01, Yedidia 05]. Dans le cadre de la mise en correspondance, c'est une technique très récente et c'est seulement dans [Tappen 03, Sun 03, Gutiérrez 04, Sun 05] qu'elle est utilisée. Bien qu'elle soit très marginale, nous pensons important de la décrire puisque c'est la technique employée par la méthode de mise en correspondance la mieux classée par le protocole d'évaluation de Scharstein et Szeliski [Scharstein 02] (c'est ce que nous verrons dans le chapitre 2 et plus précisément sur le tableau 2.1). Cette technique s'appuie sur le modèle des champs de Markov. Elle est très proche de la technique des réseaux de neurones (décrite au paragraphe 1.10.6). L'énergie globale est de la forme de l'équation (1.18) avec un coût de dissimilarité, cf. paragraphe 1.9.3.1, et un coût de lissage, cf. équation (1.27). Un neurone représente un pixel et sa disparité associée. À chaque itération, l'énergie est mise à jour en prenant en compte les nouvelles valeurs affectées à chaque neurone, sachant que chaque neurone est mis à jour en assimilant les messages reçus de la part de ses voisins. Les messages dépendent du coût de dissimilarité et du coût de lissage associés à chaque neurone. Cette méthode est comparée à celle des coupure de graphe par Tappen et Freeman [Tappen 03] qui montrent que ces deux techniques sont aussi performantes l'une que l'autre.

1.10.8 Algorithmes génétiques

Holland [Holland 75] est le premier à avoir exposé des algorithmes génétiques fondés sur la théorie de l'évolution de Darwin. Ils consistent à faire évoluer une population d'individus à l'aide de différents opérateurs qui sont la sélection, le croisement et la mutation. Ces opérateurs correspondent à des tirages aléatoires. Les algorithmes génétiques, qui sont itératifs, s'appuient sur trois étapes :

- *Production et évolution* – Cela correspond à une étape d'initialisation. Les individus sont représentés par une chaîne de bits. La population est créée, de manière aléatoire ou non, lors de la première itération de l'algorithme. Lors des itérations suivantes, cette étape consiste en une modification de la population par croisement (échange aléatoire entre les chromosomes de deux individus différents) et mutation (changement aléatoire d'un bit), sachant que, naturellement, un croisement est beaucoup plus probable qu'une mutation.
- *Sélection* – Grâce à une fonction d'ajustement, *fitness*, on sélectionne les meilleurs individus.
- *Vérification* – Si la population obtenue est optimale, suivant certains critères, on arrête, sinon, on continue.

Les méthodes de mise en correspondance s'appuyant sur les algorithmes génétiques sont assez rares et surtout assez récentes [Saito 95, Han 01, Gong 02, Goulermas 03]. L'adaptation des algorithmes génétiques au cas particulier de la mise en correspondance se traduit par les règles suivantes :

- Chaque individu correspond à une carte de disparités. Un individu n'est donc plus représenté par une chaîne de bits mais par un ensemble de disparités.
- Le croisement et la mutation s'effectuent de manière aléatoire.
- La fonction de sélection s'appuie sur le coût global de mise en correspondance.

1.11 Approche à passages multiples

Nous distinguons les méthodes itératives, qui correspondent toujours aux méthodes par optimisation globale, des méthodes à passages multiples. Ces dernières correspondent aux méthodes qui effectuent plusieurs fois le calcul des disparités mais avec des techniques de mise en correspondance différentes. Pour les méthodes qui utilisent la contrainte de symétrie, il faut calculer les disparités de la gauche vers la droite puis de la droite vers la gauche pour pouvoir sélectionner les correspondances qui respectent la contrainte. En dehors de ce cas très particulier et très répandu, de plus en plus de méthodes à passages multiples se développent actuellement et il nous a paru intéressant de les répertorier et de les classer en trois catégories : les méthodes qui nécessitent une initialisation, les méthodes fixant les conditions aux limites et les méthodes par croissance de germes.

1.11.1 Initialisation

En observant l'algorithme 1.7, nous voyons que les méthodes par optimisation globale peuvent entrer dans la catégorie des méthodes à passages multiples (sauf les méthodes par recuit simulé). En effet, chacune de ces méthodes a besoin d'une estimation initiale des disparités et la plupart utilisent une autre technique de mise en correspondance pour cette initialisation. La manière de procéder la plus courante est d'effectuer une première mise en correspondance avec une technique locale puis d'effectuer une deuxième mise en correspondance globale. Nous pouvons citer les cas d'une mise en correspondance par corrélation (SAD et ZNCC, cf. équations (3.19 et 3.16)) suivie d'une mise en correspondance par relaxation proposée dans [Brockers 04b, Brockers 04a] ou par programmation dynamique [Park 01].

D'autres méthodes utilisent plusieurs méthodes globales. Les publications les plus significatives sont :

- [Belhumeur 92] – Les deux techniques d'optimisation utilisées sont la programmation dynamique

et le recuit simulé. D’abord, une optimisation avec un support 1D, c’est-à-dire le long des droites épipolaires est effectuée. Bien que le recuit simulé ne nécessite par d’initialisation, la solution obtenue sur chaque droite épipolaire est utilisée comme initialisation pour le recuit simulé avec un support 2D, c’est-à-dire sur toute l’image. Nous supposons que les auteurs ont fait ce choix dans le but de faire converger l’algorithme de recuit simulé plus vite.

- [Luo 95] – Il s’agit de deux méthodes par relaxation. Dans les deux cas, les énergies sont différentes, la première prend en compte seulement un coût de dissimilarité et la seconde prend aussi en compte une contrainte liée aux occultations.

1.11.2 Détermination automatique des paramètres des méthodes

Dans le but de déterminer automatiquement certains paramètres de la méthode qu’ils ont choisie, certains auteurs effectuent une mise en correspondance initiale grâce à une méthode locale. Cette dernière est généralement une méthode simple et rapide qui donne une première approximation des disparités permettant de calculer les paramètres, comme l’intervalle de variation des disparités. Nous distinguons les méthodes qui utilisent en deuxième passage une méthode locale de celles qui utilisent une méthode globale.

1.11.2.1 Méthodes globales à passages multiples

Pour les méthodes globales que nous pouvons citer, la première étape sert à diminuer l’espace de recherche des correspondances :

- Dans [Zickler 03], Zickler *et al.* effectuent une mise en correspondance de points contour suivie d’une mise en correspondance par programmation dynamique.
- Dans [Intille 94, Bobick 99, Tsai 99, Gong 03], les auteurs effectuent une mise en correspondance initiale par corrélation (ZSSD, cf. équation (3.24)) qui permet de choisir les points de contrôle (correspondances qui vérifient la contrainte de symétrie et dont le coût local dépasse un seuil fixé) suivie d’une mise en correspondance par programmation dynamique.
- Dans [Ohta 85, Kim 05], après avoir effectué une mise en correspondance de points d’intérêt, les auteurs réalisent une mise en correspondance par programmation dynamique en deux passages : un passage suivant les droites épipolaires, puis un passage en prenant en compte les droites épipolaires adjacentes.

1.11.2.2 Méthodes locales à passages multiples

Le principe de toutes ces méthodes est d’effectuer une mise en correspondance initiale avec une mesure de corrélation classique (ZNCC, SSD ou SAD, cf. paragraphes 3.3.1.2 et 3.3.2.1) puis de réexaminer les résultats obtenus sur tous les pixels ou sur une partie des pixels :

- Dans [Manduchi 99], après la première mise en correspondance, Manduchi et Tomasi détectent les pixels distincts, c’est-à-dire, les pixels dont la mise en correspondance n’est pas ambiguë (cf. paragraphe 3.6.1). Une mise en correspondance de ces pixels distincts permet de limiter la zone de recherche pour les pixels ambigus. Plus formellement :

Si \mathbf{p}_g^{i,j_1} et \mathbf{p}_g^{i,j_2} sont deux pixels distincts dont les correspondants sont respectivement \mathbf{p}_d^{i,v_1} et \mathbf{p}_d^{i,v_2} et $\forall j \in [j_1 + 1; j_2 - 1]$ et si $\mathbf{p}_g^{i,j}$ est un pixel ambigu, alors on effectue la mise en correspondance entre les pixels $\{\mathbf{p}_g^{i,j}\}$, $j \in [j_1 + 1; j_2 - 1]$ et $\{\mathbf{p}_d^{i,v}\}$, $v \in [v_1 + 1; v_2 - 1]$.

La figure 1.8 permet d’illustrer cette méthode.

- Dans [Kostková 02, Kostková 03, Eklund 03], les auteurs déterminent une nouvelle zone d’agrégation en fonction des résultats obtenus après la première mise en correspondance.

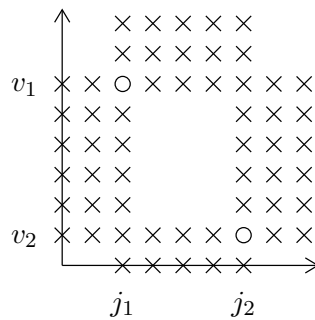


FIG. 1.8 – *Technique des pixels distincts de Manduchi et Tomasi* – Les notations utilisées sont les mêmes que dans la figure 1.5. L'utilisation de la contrainte des pixels distincts permet d'interdire un grand nombre de correspondances pour les pixels compris entre \mathbf{p}_g^{i,j_1} et \mathbf{p}_d^{i,v_1} . L'espace de recherche est donc très réduit.

- Dans [Zoghلامي 96], Zoghلامي *et al.* utilisent les coûts locaux obtenus au cours de la première étape pour effectuer une deuxième mise en correspondance avec des coûts locaux pondérés (les poids sont calculés en fonction des coûts locaux calculés à la première étape).
- Dans [Zitnick 00], Zitnick et Kanade proposent un coût local coopératif et comme dans [Zoghلامي 96], les coûts locaux sont modifiés en fonction des coûts obtenus à l'étape précédente. La différence avec la méthode dans [Zoghلامي 96] est que le nombre d'itérations n'est pas fixe. Cette méthode est aussi appelée méthode coopérative, mais elle est très différente des méthodes coopératives abordées au paragraphe 1.8.2. Dans [Jones 92, Mayer 03], les auteurs proposent une technique qui s'appuie sur le même principe que dans [Zitnick 00] mais avec un coût local différent.

1.11.3 Méthodes par croissance de germes

Le principe des méthodes par croissance de germes est d'effectuer une détection de points d'intérêt ou une segmentation de l'image, de mettre en correspondance une partie des pixels puis d'augmenter progressivement le nombre de pixels mis en correspondance de manière itérative. Nous pouvons citer deux travaux significatifs :

- [Lhuillier 00, Lhuillier 04] – La mise en correspondance est effectuée avec une mesure de corrélation (ZNCC, cf. équation (3.16)) pour des pixels germes (points d'intérêt) et la contrainte de symétrie est utilisée pour obtenir des correspondances les plus fiables possibles. Au cours de l'étape de propagation, on considère deux ensembles : un ensemble des pixels appariés, \mathcal{P}_A et un ensemble des pixels qu'il reste à appairier \mathcal{P}_N . Tous les pixels germes sont dans \mathcal{P}_A . La propagation s'effectue de manière itérative : à chaque étape, on sélectionne le pixel dans \mathcal{P}_A qui a obtenu le meilleur coût local et on recherche les correspondances dans le voisinage de ce pixel. Pendant cette étape, la taille de la zone d'agrégation choisie est plus petite que celle utilisée au cours de la première étape, cela permet de limiter l'influence des occultations (plus la taille de la zone est petite et moins il y a de pixels avec une disparité différente du pixel étudié qui sont pris en compte). La méthode proposée dans [Zhang 00] s'appuie sur le même principe en ajoutant la limite du gradient de disparité comme contrainte.
- [Wei 04] – La méthode du *mean shift* est utilisée pour segmenter l'image en régions. La mise en correspondance initiale est effectuée en utilisant une mesure de corrélation (SSD ou SAD, cf. équations (3.20) et (3.19)). Au départ, chaque région est considérée comme non appariée. La

propagation s'effectue en deux étapes :

1. Une mesure de confiance est attribuée à chaque région. Elle dépend du nombre de points de contrôle de la région, sachant que les points de contrôle sont déterminés en sélectionnant les correspondances non ambiguës. Les régions sont ordonnées suivant leur mesure de confiance.
2. Pour chaque région, une mesure d'ambiguïté est calculée. Elle dépend du nombre de disparités différentes dans la région. Si cette ambiguïté est faible, alors la région est considérée comme appariée, sinon, elle est divisée en deux régions et les deux régions sont considérées comme non appariées.

La propagation s'arrête lorsque toutes les régions sont appariées.

1.12 Affinement des résultats

1.12.1 Mise en correspondance au sous-pixel

Dans la majorité des publications, les résultats obtenus sont au pixel près. Dans certains domaines, on ne peut pas se satisfaire d'une erreur d'un demi-pixel, par exemple, dans le cas d'images aériennes, si la résolution est faible, une erreur d'un pixel peut entraîner une erreur de plusieurs mètres dans la reconstruction 3D. C'est pourquoi plusieurs techniques ont été proposées pour effectuer un calcul plus fin, un calcul au sous-pixel, cf. [Tian 86, Davis 95], c'est-à-dire que les positions des correspondants ne sont plus des valeurs entières mais des nombres réels. De très nombreuses méthodes ajoutent une étape d'affinement des résultats au sous-pixel après avoir effectué la mise en correspondance au pixel près, comme dans [Szeliski 02, Hirschmüller 02, Park 03, Georgescu 04, Szeliski 04, Brockers 04b, Forstmann 04]. Deux techniques sont envisageables :

- *Interpolation quadratique des coûts locaux* – C'est la méthode la plus populaire et la plus utilisée. Il suffit d'interpoler les scores de corrélation obtenus pour les pixels voisins du correspondant trouvé. La méthode la plus utilisée est décrite par Tian et Huhns [Tian 86] et une version améliorée, beaucoup plus récente, est fournie par Shimizu et Okutomi [Shimizu 05].
- *Interpolation des niveaux de gris* – Dans le cadre de la mise en correspondance, cette méthode est beaucoup moins répandue. Elle consiste à interpoler (le plus souvent par interpolation bicubique) les niveaux de gris pour augmenter la résolution des images à effectuer la mise en correspondance avec ces images à haute résolution en prenant en compte les résultats ayant été obtenus à une plus faible résolution (cette méthode correspond aussi à une méthode à passages multiples).

Même si elles sortent du cadre de notre étude, signalons que les techniques de corrélation de phase permettent d'obtenir une mise en correspondance au sous-pixel [Tian 86, Shekarforoush 95].

1.12.2 Interpolation des disparités

Dans certains cas, lorsque la mise en correspondance n'est pas dense, les auteurs décident de calculer une carte dense en interpolant les disparités trouvées, comme proposé dans [Fua 91, Cochran 92, Fua 93, Menard 97, Tsai 99, Silva 00, Kostková 03]. Une des méthodes les plus courantes est de considérer les pixels que l'on a appariés comme des germes et de propager les résultats de ces germes sur les pixels voisins, en suivant certaines règles, de manière itérative, jusqu'à obtenir l'ensemble des disparités pour tous les pixels de l'image ; c'est ce qui est réalisé dans [Cochran 92] (cette méthode peut être vue comme une méthode à passages multiples). D'autres, comme dans [Tsai 99], utilisent une interpolation linéaire.

1.12.3 Prise en compte d'erreurs d'appariement

Dans de très nombreux articles, l'ensemble des correspondances finales subit quelques corrections dans le but d'éliminer des erreurs. Nous pouvons citer deux buts à ces corrections :

- Supprimer des correspondances et ainsi, augmenter la fiabilité des résultats – Nous pouvons distinguer de nombreuses techniques :
 - utilisation de la contrainte de symétrie (c'est la méthode la plus populaire) [Falkenhagen 97, Fusiello 97b, Belli 00, Hirschmüller 02, Zhang 02, Maier 03, Lhuillier 04, Gutiérrez 04] ou de symétrie faible [Crouzil 97, Belli 00] (cf. paragraphe 1.6.7) ;
 - utilisation de la contrainte d'ordre [Dhond 95, Szeliski 02, Gong 05] (cf. paragraphe 1.6.6) ;
 - utilisation de la contrainte d'unicité [Koschan 93, Yang 93, Koschan 96, Mühlmann 01, Kim 05] (cf. paragraphe 1.6.5) ;
 - application d'un seuil sur le coût local (dans le cadre des méthodes locales par corrélation) [Garcia 01c, Di Stefano 01] ;
 - utilisation d'une mesure de confiance d'un appariement s'appuyant sur l'étude de la variation du coût local suivant différentes configurations (coins, contours, occultations) [Anandan 89] ;
 - étude de la cohérence avec le voisinage [Xu 97, Dufournaud 00].
- Corriger les erreurs de disparités – Nous pouvons citer [Yang 93, Koschan 97, Mühlmann 01, Egnal 02, Eklund 03] où les auteurs utilisent un filtre médian sur la carte de disparités qui permet de corriger des disparités très différentes des disparités voisines.

1.13 Approche multirésolution



FIG. 1.9 – *Multirésolution pour la mise en correspondance* – Cette figure illustre, d'une part la construction d'une pyramide d'images de I^0 à I^3 et, d'autre part, les difficultés rencontrées si nous utilisons l'image de plus faible résolution, I^3 , où nous pouvons observer la disparition de détails importants, comme la caméra, et le lissage des contours des objets.

Le but de la multirésolution est d'utiliser différents niveaux de résolution pour diminuer les temps de calcul. Le principe est d'obtenir une pyramide d'images dont le niveau le plus grossier n'a conservé que les grandes structures et le niveau le plus fin contient tous les détails. Dans le cadre de la mise en correspondance, le but est d'apparier au niveau le plus grossier, puis de répercuter les résultats sur les niveaux plus élevés, c'est-à-dire, en supposant que le niveau le plus grossier de la pyramide est $I_l^{N_p-1}$ et le niveau le plus fin I_l^0 , au niveau n , on utilise les résultats du niveau $n + 1$ pour effectuer l'appariement. Les points importants de cette technique sont donc :

- la méthode de construction des différents niveaux de résolution (pyramides d'images) ;
- la technique de propagation entre deux niveaux successifs.

La figure 1.9 illustre une pyramide d'images. Nous pouvons constater que l'image de résolution grossière ne contient que les grandes structures : la tête, le carton, la table, la tête de la lampe et le fond. On s'aperçoit alors que la caméra ou le bras de la lampe ont disparu, ce qui met en évidence les inconvénients de la multirésolution :

- Si le niveau de résolution le plus bas est trop faible, l'effet de lissage risque d'éliminer des détails pertinents de la scène (comme le bras de la lampe devant le panneau).
- La difficulté d'appariement aux discontinuités de profondeur est aggravée par l'effet de lissage, il risque donc d'y avoir des erreurs d'appariement au niveau le plus grossier qui vont se propager.

1.13.1 Construction de pyramides d'images

Le problème consiste donc à produire à partir d'une image, une autre image de plus faible résolution. Plusieurs techniques ont été développées pour calculer des pyramides d'images, principalement, dans le domaine de la compression d'images [Burt 83, Adelson 84]. Il existe différents types de pyramides :

- *Les pyramides régulières* – Ce sont les plus couramment employées, notamment dans [Burt 83, Adelson 84, Muñoz Barrutia 00]. Le même voisinage est considéré dans la phase de réduction de la résolution des images.
- *Les pyramides irrégulières* – Elles sont utilisées de manière beaucoup plus marginale [Montanvert 91, Bischof 94, Kropatsch 95]. Nous pouvons distinguer les pyramides stochastiques exposées dans [Haxhimusa 02, Jolion 03], des pyramides centrées proposées dans [Brigger 99].

Les pyramides irrégulières ne sont pas utilisées dans le cadre de la mise en correspondance. Elles ont souvent été proposées pour effectuer une segmentation en régions. Ces pyramides semblent, en effet, mieux respecter les différentes structures de l'image et ne pas supprimer les détails pertinents. Dans la suite de ce paragraphe, nous ne détaillerons pas ces méthodes, mais seulement les pyramides régulières puisque ce sont les plus couramment utilisées pour la mise en correspondance.

La construction d'une pyramide régulière d'images peut être définie par :

$$(I_l^n)^{i,j} = \operatorname{red}_{(i',j') \in ZV(i,j)} f_r((I_l^{n-1})^{i',j'}), \quad (1.39)$$

sachant que $ZV(i,j)$ correspond au voisinage de réduction (c'est une fonction qui permet de retourner l'ensemble des pixels considérés au niveau $n - 1$ pour évaluer la valeur d'un pixel au niveau n), f_r est la fonction qu'il faut appliquer aux pixels pris en compte (le plus souvent il s'agit d'une pondération) et l'opérateur red est une fusion des valeurs des pixels voisins du niveau $n + 1$ pour obtenir le niveau n .

Les pyramides régulières les plus courantes et les plus utilisées dans le cadre de la mise en correspondance sont :

- *Les pyramides gaussiennes* – Dans [Barnard 89, Luo 95, Pritchett 98, Sankar Kishore 00], les auteurs effectuent un lissage par un filtre gaussien, puis un sous-échantillonnage, en récupérant un pixel sur quatre.
- *Les pyramides moyennes* – Dans [Crouzil 97, p. 75–82], une moyenne sur un voisinage 2×2 est calculée.
- *Les pyramides non linéaires* – C'est une technique utilisée en *shape from shading* [Peleg 90, Crouzil 03], qui nécessite dans un premier temps le calcul de l'image des pentes en utilisant les niveaux de gris et l'équation de l'eikonale, au niveau n , puis d'effectuer un filtrage gaussien et un sous-échantillonnage et enfin, de calculer l'image au niveau $n + 1$, en utilisant de nouveau l'équation de l'eikonale.

1.13.2 Propagation des résultats

Au cours de la propagation des résultats, lorsqu'on effectue la mise en correspondance au niveau n , la méthode employée est contrainte par les résultats obtenus au niveau $n + 1$. Par exemple, dans le cadre de la mise en correspondance locale [Crouzil 97, p. 79–82], les disparités fournies au niveau $n + 1$ permettent de réduire la zone de recherche au niveau n . C'est au cours de cette deuxième étape que les erreurs peuvent se propager.

1.14 Synthèse des méthodes de mise en correspondance

Nous avons choisi de classer toutes les méthodes de mise en correspondance stéréoscopique de pixels en quatre catégories (cf. tableau 1.3) :

- les méthodes locales ;
- les méthodes globales ;
- les méthodes mixtes ;
- les méthodes à passages multiples (pour cette catégorie nous distinguons les méthodes locales des méthodes globales).

Dans chacune de ces catégories, ce qui nous semble le plus discriminant est le coût de mise en correspondance globale. Mais si nous choisissons ce critère, nous obtenons quasiment autant de sous-catégories que d'articles cités. Nous avons donc préféré une classification en fonction de la méthode d'optimisation choisie.

À partir de ce tableau récapitulatif, nous pouvons faire les remarques suivantes :

- Il y a autant de méthodes locales que de méthodes globales.
- Les méthodes mixtes et les méthodes à passages multiples sont plus récentes et tendent à se développer plus que les méthodes locales ou globales.
- La programmation dynamique est la méthode d'optimisation la plus populaire.

1.15 Conclusion

Tout d'abord, ce premier chapitre nous a permis de caractériser et de définir les différents éléments constituant de la mise en correspondance stéréoscopique de pixels. Puis, nous avons distingué quatre grandes catégories de méthodes. Enfin, nous avons montré qu'il y a un nombre important de méthodes locales, ce qui porte notre intérêt à étudier ce type de méthodes que nous détaillerons au chapitre 3.

Vu le grand nombre de méthodes proposées, nous pouvons nous demander comment elles sont évaluées et comparées. Si nous nous attardons sur les évaluations effectuées dans les articles cités, nous réalisons que ces évaluations sont souvent peu précises (ce sont souvent des évaluations visuelles) et les comparaisons sont partielles. Il est effectivement difficile d'établir une évaluation et une comparaison précise et exhaustive, notamment lorsque nous étudions des méthodes qui utilisent des techniques très différentes. Il nous semble important d'aborder à présent les protocoles d'évaluation et de comparaison proposés dans la littérature et de définir le protocole que nous avons retenu dans notre travail.

TYPE		MÉTHODE D'OPTIMISATION	RÉFÉRENCES
<i>Méthode locale</i> (1975–2004, 40)		Recherche exhaustive	[Nack 75, Nishihara 88, Seitz 89, Fua 91, Jones 92, Weng 92, Koschan 93, Fua 93, Lotti 94, Zabih 94, Schmid 95, Crouzil 96, Kanade 96, Koschan 96, Fusiello 97b, Falkenhagen 97, Koschan 97, Menard 97, Xu 97, Paparoditis 98, Smith 98, Arcara 00, Belli 00, Tuytelaars 00, Banks 01, Garcia 01c, Mühlmann 01, Ullah 01, Egnal 02, Hirschmüller 02, Kaneko 02, Matas 02, Ghafoor 03, Pajares 03, Rogelj 03, Stock 03, Benhimane 04, Carneiro 04, Georgescu 04, Kaplan 04]
<i>Méthode globale</i> (1976–2005, 51)		Recherche exhaustive	[Tao 00, Tao 01, Bleyer 04]
		Programmation dynamique	[Cox 92, Intille 94, Cox 95, Dhond 95, Geiger 95, Cox 96, Birchfield 98, Bobick 99, Tsai 99, Fielding 00, Park 01, Zhang 02, Bovyryn 03, Gong 03, Veksler 03, Zickler 03, Forstmann 04, Gong 04, Gong 05, Veksler 05]
		Recuit simulé	[Barnard 89, Starink 95, Lee 02]
		Relaxation	[Nasrabadi 92a, Zhang 93, Gouet 98, Gouet 00]
		Coupure de graphe	[Ishikawa 98, Birchfield 99, Veksler 99, Boykov 01, Kang 01, Kolmogorov 01, Kolmogorov 02, Lin 02, Hong 04, Deng 05]
		Réseaux de neurones	[Marr 76, Nasrabadi 92b, Hu 93, Wang 99, Jahn 01]
		Propagation de croyance	[Sun 03, Tappen 03, Gutiérrez 04, Sun 05]
		Algorithme génétique	[Saito 95, Han 01]
<i>Méthode mixte</i> (1989–2004, 16)		Programmation dynamique	[Rziza 00, Sun 02, Criminisi 03, Leung 04, Torr 04]
		Recuit simulé	[Le Besnerais 98]
		Relaxation	[Anandan 89, Black 93, Black 96, Shao 02, Brockers 04a, Brockers 04b]
		Coupure de graphe	[Roy 98, Agrawal 04]
		Algorithme génétique	[Gong 02, Goulermas 03]
<i>Méthode à passages multiples</i> (1992–2005, 17)	<i>Locale</i>	Recherche exhaustive	[Cochran 92, Zoghiami 96, Manduchi 99, Kostková 02, Eklund 03, Kostková 03]
		Croissance de germes	[Lhuillier 00, Zhang 00, Lhuillier 04, Wei 04]
		Méthode coopérative	[Zitnick 00, Mayer 03, Park 03]
	<i>Globale</i>	Programmation dynamique	[Ohta 85, Kim 05]
		Programmation dynamique et recuit simulé	[Belhumeur 92]
		Relaxation	[Luo 95]

TAB. 1.3 – Classement des méthodes de mise en correspondance – Nous précisons, entre parenthèse, les années et le nombre de publications pour chaque catégorie de méthodes.

Chapitre 2

Protocole d'évaluation des méthodes de mise en correspondance stéréoscopique

2.1 Introduction

2.1.1 Évaluation des performances en vision par ordinateur

Le chapitre précédent résume les différentes méthodes de mise en correspondance existantes. Leur multitude montre à quel point le nombre de publications dans ce domaine est important. De plus, il s'avère difficile d'établir des comparaisons pertinentes et de trouver quel type de méthodes utiliser suivant le contexte dans lequel on se situe (type d'images, type d'applications). Depuis plusieurs années, les chercheurs tentent de trouver des protocoles pertinents pour établir des évaluations et des comparaisons précises et exhaustives des méthodes qu'ils proposent. D'une part, nous pouvons observer cet effort d'évaluation à travers le programme national, TechnoVision¹, lancé en 2004 par le ministère délégué à la recherche, proposant de financer des campagnes d'évaluation et de création de ressources associées permettant l'évaluation de diverses technologies de la vision par ordinateur et du traitement d'images. D'autre part, plusieurs publications significatives illustrent ce phénomène :

- Dans [Unnikrishnan 05], des critères d'évaluation et de comparaison de résultats de segmentation ont été proposés.
- Dans [Schlögl 04], une méthodologie pour établir une plate-forme d'évaluation d'algorithmes de détection et de suivi de mouvement (génération de séquences vidéo de synthèse et critères d'évaluation des résultats) a été mise en place.
- Dans [Chabrier 04], une étude sur les critères d'évaluation dans le cadre de la segmentation d'images a été proposée.
- Dans [Messer 04], la plus grande base de données pour évaluer les performances des méthodes de reconnaissance des visages² a été développée.
- Dans [Gribkov 03], un système d'évaluation d'algorithmes de détection de contours³ a été proposé.
- Dans [Scharstein 02], un protocole d'évaluation et de comparaison de méthodes de mise en correspondance⁴ a été mis à la disposition de la communauté.
- Dans [Klette 00], des techniques d'évaluation de méthodes en vision par ordinateur et en traitement d'images ont été exposées dans de nombreux domaines tels que l'estimation de mouvement

1. <http://www.recherche.gouv.fr/appel/2004/technovision.htm>

2. <http://www.ee.surrey.ac.uk/banca/>

3. <http://www.ensi-bourges.fr/LVR/SIV/interpretation/evaluation/>

4. <http://cat.middlebury.edu/stereo/>

ou la segmentation d'images. Pour compléter les aspects traités dans ce livre, le site *Performance Characterization in Computer Vision* met à disposition des cours et des exemples d'études comparatives⁵.

2.1.2 Caractéristiques d'un protocole d'évaluation

Lorsqu'on souhaite établir un protocole d'évaluation et de comparaison dans le cadre de la mise en correspondance, plusieurs aspects sont à considérer :

- *Les images utilisées* – Il faut déterminer quel est le type des images testées (images réelles, de synthèse, images bruitées, images avec des occultations) et si ce sont des images avec vérité terrain ou non (nous aborderons cette notion dans le paragraphe 2.2).
- *Les critères évalués* – Ils dépendent, bien sûr, du type d'images utilisées (résistance au bruit, aux occultations) et de la possibilité ou non d'avoir une vérité terrain.
- *Les méthodes comparées* – Afin d'obtenir des résultats comparables et significatifs, il convient de choisir minutieusement le type de méthodes comparées. Par exemple, si on souhaite apporter une amélioration dans le choix du coût local d'une méthode de mise en correspondance mixte (cf. paragraphe 1.9.4), il paraît judicieux de comparer les résultats obtenus à ceux des méthodes du même type. En revanche, il n'est pas justifié de comparer les résultats obtenus avec ceux d'une méthode globale ou d'une méthode locale.
- *La synthèse des résultats* – Si plusieurs critères sont évalués, il est plus facile de caractériser le comportement des méthodes, en revanche, il est beaucoup plus difficile de trouver la méthode la plus performante. Pour cela, il faut donc mettre en place une stratégie de synthèse des résultats obtenus.

2.1.3 Objectifs

Dans le cadre de ce mémoire, nous souhaitons proposer plusieurs méthodes de mise en correspondance stéréoscopique de pixels à base de corrélation : des méthodes prenant en compte le problème des occultations et des méthodes utilisant les composantes couleur de l'image. De manière à valider nos approches, nous souhaitons évaluer et comparer nos méthodes de la façon la plus rigoureuse possible. Actuellement, à notre connaissance, il existe peu de publications faisant l'inventaire de techniques d'évaluation. Le protocole d'évaluation qui nous semble le plus important est celui de Scharstein et Szeliski [Scharstein 02]. Il est très utilisé par la communauté mais toutes les images proposées par les auteurs ne sont pas prises en compte, les critères évalués sont assez généraux et les résultats obtenus ont été finalement peu commentés et peu exploités. Les objectifs de ce chapitre sont donc les suivants :

- établir un état de l'art sur les protocoles d'évaluation et de comparaison ;
- approfondir le travail proposé dans [Scharstein 02] en proposant une analyse des résultats qu'ils ont obtenus ;
- proposer notre propre protocole d'évaluation et de comparaison pour valider nos approches.

Pour atteindre ce dernier objectif, nous souhaitons notamment :

- utiliser les images proposées dans [Scharstein 02] et non exploitées dans le protocole proposé par les auteurs ;
- établir notre propre technique de construction de données avec vérité terrain et mettre à disposition de la communauté l'outil nécessaire ainsi que les nouvelles données ;
- prendre en compte des critères plus « adaptés » aux méthodes à base de corrélation et aux aspects auxquels nous nous intéressons : les images couleur et la prise en compte des occultations.

5. <http://peipa.essex.ac.uk/benchmark/>

2.1.4 Plan

Dans un premier temps, nous allons exposer un bref état de l'art sur les protocoles d'évaluation et de comparaison, en exposant le type d'images utilisées, les critères évalués et les algorithmes comparés. Ensuite, nous décrirons le protocole de Scharstein et Szeliski [Scharstein 02] qui est, à notre connaissance, celui qui a permis l'évaluation et la comparaison du plus grand nombre de méthodes de mise en correspondance stéréoscopique. Nous donnerons également une analyse des résultats qu'ils ont obtenus. Enfin, nous exposerons le protocole que nous avons utilisé pour établir des évaluations et des comparaisons des méthodes existantes et valider nos approches, c'est-à-dire, le type d'images utilisées (construction de données avec vérité terrain), les différentes zones d'évaluation considérées, les différents critères évalués ainsi que la technique de synthèse des résultats associée.

2.2 Définitions

2.2.1 Carte de disparités

Dans nos études comparatives, une des manières de représenter visuellement les résultats de la mise en correspondance consiste à utiliser une image appelée « carte de disparités » :

Chaque pixel d'une carte de disparités représente l'amplitude de la disparité, c'est-à-dire, la distance entre la position du pixel de l'image gauche et celle de son correspondant dans l'image droite.

Dans la plupart des cartes de disparités, plus le pixel est clair et plus la disparité est importante et les pixels noirs représentent les pixels sans correspondant. Dans tout ce mémoire, nous avons choisi d'utiliser des cartes de disparités avec ces conventions.

2.2.2 Vérité terrain

La vérité terrain correspond à toutes les connaissances que nous pouvons avoir sur la scène et la formation des images : la localisation tridimensionnelle exacte de certains points de l'image et la distance exacte entre deux points de l'image. Mainmone et Shafer [Mainmone 96] distinguent même deux catégories de vérité terrain : le cas où l'on connaît toute la vérité, c'est-à-dire le type d'éclairage, le type et le taux de bruit, le taux de dégradation des images et dans ce cas, ils parlent d'environnement contrôlé et le cas où une partie de la vérité terrain est connue grâce à des objets dans la scène dont on connaît les caractéristiques et les propriétés et ils parlent d'environnement mesuré.

Dans le cadre de ce travail, la vérité terrain correspond à la connaissance de l'ensemble des correspondances entre les pixels homologues des deux images. Cela se traduit par la connaissance de toutes les disparités par rapport à une image de référence et la représentation visuelle de cette vérité terrain est une carte de disparités.

2.3 Images testées

Une des premières choses à déterminer est le type d'images à tester. Nous considérons deux catégories d'images : les images de synthèse et les images réelles. Dans le cadre des images réelles, nous distinguons les images avec vérité terrain et les images sans vérité terrain.

2.3.1 Images de synthèse

2.3.1.1 Stéréogramme aléatoire

Très souvent, un simple stéréogramme aléatoire est utilisé [Marr 76, Gennert 88, Cochran 92, Cox 96, Fusiello 97b, Scharstein 98, Jawahar 02, Eklund 03]. L'image gauche est générée aléatoirement. Dans le cas des images en niveaux de gris, nous affectons à chaque pixel de l'image une valeur aléatoire comprise entre 0 et 255. L'image droite est déduite de l'image gauche et du relief de la scène souhaitée. Nous illustrons la construction de ce type de stéréogrammes sur la figure 2.1. Nous choisissons de représenter, dans les deux images, la projection du premier plan de la scène par un rectangle dont les côtés sont parallèles aux bords de l'image. De plus, nous choisissons d'effectuer une faible translation parallèle aux lignes entre les deux images et nous ne modélisons pas les déformations perspectives. Nous posons $N_g^{\text{lig}} = N_d^{\text{lig}} = N^{\text{lig}}$ et $N_g^{\text{col}} = N_d^{\text{col}} = N^{\text{col}}$. Tout d'abord, l'image droite est générée en décalant les pixels de l'image gauche de d_1 colonnes, c'est-à-dire, $I_d^{i,j-d_1} = I_g^{i,j}$ avec $(i,j) \in [0; N^{\text{lig}} - 1] \times [d_1; N^{\text{col}} - 1]$. La projection du premier plan, dans l'image gauche, a respectivement pour coin supérieur gauche et coin inférieur droit les pixels de coordonnées (i_1, j_1) et (i_2, j_2) . La projection du premier plan dans l'image droite est générée en décalant les pixels de ce premier plan de d_2 colonnes, c'est-à-dire, $I_d^{i,j-d_2} = I_g^{i,j}$, avec $i \in [i_1; i_2]$ et $j \in [j_1; j_2]$. Enfin pour tous les pixels $\mathbf{p}_d^{i,j}$, tels que $(i,j) \in ([0; N^{\text{lig}} - 1] \times [N^{\text{col}} - d_1 - 1; N^{\text{col}} - 1]) \cup ([i_1; i_2] \times [j_2 - d_2; j_2])$, on attribue un nouveau niveau de gris aléatoire.

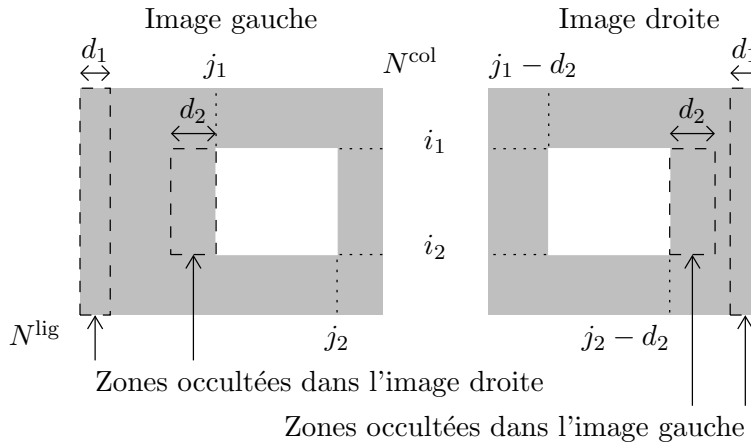


FIG. 2.1 – Génération d'un stéréogramme aléatoire – Cette figure illustre la technique que nous avons utilisée dans ce mémoire. Les zones encadrées en pointillés dans l'image droite représentent les zones où il faut effectuer un deuxième tirage aléatoire (ces zones ne sont pas visibles de l'image gauche). Les zones encadrées en pointillés dans l'image gauche sont les zones invisibles dans l'image droite. Pour le stéréogramme que nous avons utilisé, nous avons choisi $N^{\text{lig}} = N^{\text{col}} = 256$, $(i_1, j_1) = (78, 78)$, $(i_2, j_2) = (178, 178)$, $d_1 = 0$ et $d_2 = 10$.

2.3.1.2 Modélisation de scènes en 3D

Les couples d'images de synthèse peuvent aussi être obtenus en modélisant de manière plus complexe des scènes 3D. Pour cela, il faut distinguer trois étapes :

- modéliser la scène composée d'objets géométriques ;
- « texturer » les objets ;
- faire un rendu de la scène du point de vue des deux caméras.

La vérité terrain de ces images est alors calculée exactement puisque nous connaissons les matrices de projection perspective \mathbf{M}_l associées aux deux caméras. Le principe utilisé est le suivant :

Si on considère \mathbf{p}_g , nous pouvons calculer son rayon de projection grâce à \mathbf{M}_g . Ce rayon traverse les surfaces des objets de la scène en un ou plusieurs points. Il faut retenir le point \mathbf{P} le plus proche du plan image gauche et calculer sa projection \mathbf{p}_d sur l'image droite grâce à \mathbf{M}_d .

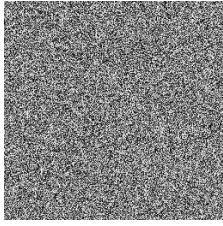
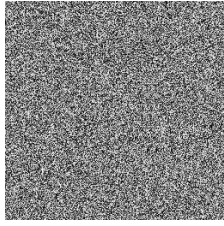
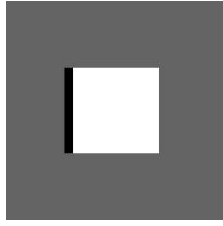
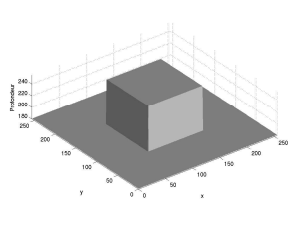


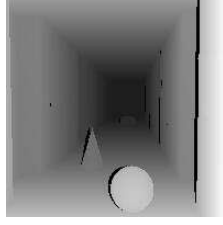
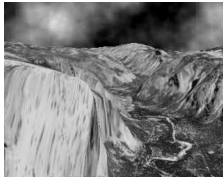
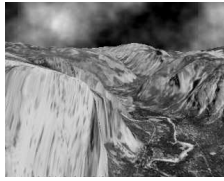
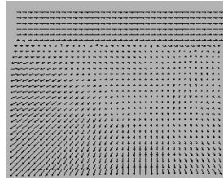
NOM	IMAGES			
	(a)	(b)	(c)	(d)
Stéréogramme aléatoire				
Corridor				
Yosemite				

FIG. 2.2 – *Couples de synthèse* – Pour le stéréogramme aléatoire, les niveaux de gris des images gauche (a) et droite (b) sont tirés aléatoirement suivant une loi uniforme entre 0 et 255. Elles représentent un cube posé sur un plan parallèle au plan image et nous pouvons voir, en (c), la carte de disparités associée et, en (d), une visualisation 3D de la scène. Pour les images corridor, elles ont été générées en couleur à l'Université de Bonn. Elles ont été proposées pour faire de l'évaluation de méthodes de mise en correspondance et la vérité terrain est disponible sous forme d'une carte de disparités et d'une carte des occultations que nous avons fusionnées pour adopter notre représentation en (c). Les images yosemite correspondent à une séquence d'images de synthèse représentant le survol d'un paysage montagneux, avec ou sans nuage. Elles ont été proposées pour effectuer des tests d'estimation de flux optique et ainsi la vérité terrain est fournie sous la forme de champs discrets de vecteurs déplacements (c).

Les images de synthèse les plus utilisées sont les images (cf. figure 2.2) :

- *Corridor* – Elles proviennent de l'Université de Bonn et ont été proposées par Gerdes⁶ et testées, notamment, dans [Frohlinghaus 96, Cyganek 03]. Gerdes a mis à disposition sur son site un programme⁷ qui génère les cartes de disparités ainsi que les cartes des occultations avec une précision au sous-pixel. Il est possible de régler la position des caméras. Il n'y a aucune publication relative à ces travaux.

6. http://www-dbv.cs.uni-bonn.de/stereo_data/

7. <http://www-student.informatik.uni-bonn.de/~gerdes/MRTStereo/>

- *Yosemite* – Ces images ont été proposées par Quam⁸ et apparaissent dans [Black 94]. Elles sont utilisées, notamment, dans [Szeliski 99a].

2.3.1.3 Résumé sur les images de synthèse

Elles sont populaires et sont notamment utilisées dans [Puzicha 97, Fusiello 99, Leclerc 00, Garcia 01a, McCane 01, Shafique 03, Xiao 03] car elles ont les avantages suivants :

- La vérité terrain est connue, ce qui permet d'utiliser des critères quantitatifs d'évaluation.
- Tous les types de bruits, de changements de luminosité peuvent être ajoutés. De plus, de nombreuses occultations peuvent être synthétisées.

Même si le réalisme des images de synthèse a fait récemment des progrès significatifs, les images de synthèse utilisées pour l'évaluation de méthodes de mise en correspondance ne sont pas récentes. Ainsi, elles ne prennent pas en compte les déformations induites par le capteur et l'éclairage de la scène et les textures des objets ne sont pas réalistes. Toutefois, Pugeault et Krügler [Pugeault 03] ont proposé de plaquer des images réelles sur les surfaces des objets de synthèse. D'autres auteurs ont proposé d'appliquer un ensemble d'homographies à l'image de gauche pour générer l'image de droite (ainsi la vérité terrain est connue), mais ces images ne sont pas très représentatives de scènes rencontrées dans les applications réelles, comme on peut le voir dans [Georgescu 04].

Pour évaluer les méthodes de mise en correspondance, les couples d'images de synthèse ne sont actuellement utilisés que dans une première étape qui doit être suivie de tests sur des images réelles.

2.3.2 Images réelles sans vérité terrain

Il existe deux très grandes bases de couples stéréoscopiques sans vérité terrain :

- la base de données JISCT (JPL, INRIA, SRI, CMU, et Teleos)⁹ [Bolles 93] ;
- la base de données de l'université de Carnegie Mellon, CMU¹⁰.

Ces deux bases sont mises à disposition par le laboratoire VASC, *Vision and Autonomous Systems Center*¹¹. Les couples stéréoscopiques les plus célèbres sont (cf. figure 2.3) :

- *pentagon* utilisés dans [Collins 96, Torr 04] ;
- *epi* utilisés dans [Silva 00, Cyganek 01] ;
- *pm* utilisés dans [Cox 96, Fusiello 97b].

Les images réelles sans vérité terrain ont l'avantage d'être très diverses (tous les types de scènes peuvent être envisagés : scènes d'intérieur, d'extérieur, complexes). Cependant, seules des évaluations visuelles peuvent être faites, ce qui limite la pertinence des interprétations.

2.3.3 Images réelles avec vérité terrain

Dans le cas des images réelles, on ne peut pas avoir accès aux correspondances exactes : on ne peut donc pas disposer de la vérité terrain au sens strict du terme. En revanche, on utilise des correspondances que l'on suppose suffisamment fiables et précises pour servir de référence et donc de constituer, par abus de langage, la vérité terrain. Nous réalisons alors à quel point établir ces correspondances qui vont servir de vérité terrain est un problème délicat à résoudre.

8. <http://www.cs.brown.edu/people/black/images.html>

9. <http://vasc.ri.cmu.edu/idb/html/jisct/>

10. <http://vasc.ri.cmu.edu/idb/html/stereo/>

11. <http://vasc.ri.cmu.edu/>



FIG. 2.3 – *Couples sans vérité terrain* – Toutes ces images sont mises à disposition par le laboratoire VASC, Vision and Autonomous Systems Center. Les images pentagon sont une vue aérienne du Pentagone. Ces images sont très texturées et présentent de très nombreuses occultations. Il y a peu de variations d'intensité entre les deux images. Les images epi sont extraites d'une séquence vidéo. Il y a un faible mouvement entre deux images successives. Ce sont des images d'arbres qui présentent de nombreuses occultations. Les images pm sont extraites d'une séquence vidéo. Il y a un faible mouvement entre deux images successives. Il s'agit d'un parc devant un immeuble.

2.3.3.1 Obtention de la vérité terrain

Nous distinguons trois catégories d'obtention de la vérité terrain :

- *Obtention manuelle* – Dans certaines publications, les auteurs ont effectué la mise en correspondance manuellement, comme dans [Belli 00, Egnal 02].
- *Vision active* – Les systèmes d'acquisition possèdent une source de lumière propre, en plus des sources ambiantes de la scène. Ces techniques sont souvent très précises mais nécessitent un équipement lourd.
- *Utilisation de contraintes géométriques sur la scène* – Il faut se placer dans un cas particulier de scène qui simplifie l'obtention des correspondances.

Un grand nombre de critères de qualité d'appariement peuvent être évalués sur les images réelles avec vérité terrain. Cet avantage majeur nous semble très important lorsque l'on souhaite établir un protocole d'évaluation pertinent.

Cependant, utiliser des images réelles avec vérité terrain pose deux problèmes :

- Il faut choisir quelle technique mettre en œuvre pour obtenir la vérité terrain.
- Le type de scènes envisageables est souvent limité par la technique choisie.

Les références que nous fournissons dans cette section ne sont pas forcément des travaux dont le but initial était d'obtenir des images avec vérité terrain, mais les approches proposées nous semblent appropriées dans le cadre de la construction de données de référence.

2.3.3.2 Obtention manuelle

Dans certains couples d'images, il n'est pas trop difficile d'obtenir manuellement une partie de la vérité terrain. En effet, lorsque les images ont été rectifiées et qu'il y a une faible translation, on peut visuellement établir des correspondances pour des pixels qui se repèrent facilement à l'œil, c'est-à-dire des pixels dans des zones contrastées et bien souvent des pixels particuliers comme des coins. Cette technique est peu précise et donne des cartes de disparités non denses ; c'est pourquoi, ces dix dernières années, elle a été très peu utilisée. Dans les bases de données du JIST et du CMU que nous avons présentées au paragraphe 2.3.2, certains couples sont fournis avec quelques correspondances obtenues manuellement. Le couple le plus utilisé est *castle*, notamment dans [Roy 98, Gutiérrez 04] (cf. figure 2.4). Un autre exemple de couple dont la vérité terrain a été obtenue manuellement est *tsukuba*, cf. [Nakamura 96]. Ce couple d'images est très utilisé notamment dans [Scharstein 02] (cf. paragraphe 2.8.1 et figure 2.7).



FIG. 2.4 – Couple avec vérité terrain obtenue manuellement – Les images (a) et (b) correspondent à deux images successives dans une séquence de onze images. Dans cette séquence, la caméra se déplace suivant une translation puis s’approche de la scène. La scène est une maquette de château et une mire de calibrage est plaquée sur le fond de la scène (points noirs). La vérité terrain est fournie pour vingt-huit points répartis sur les bâtiments.

2.3.3.3 Vision active

Scanners laser 3D – Ce sont des dispositifs qui peuvent acquérir la géométrie tridimensionnelle d’une scène placée dans un volume donné. Le système envoie un rayon laser sur les objets de la scène qui est réfléchi par ces objets et récupéré par des capteurs. La structure 3D des objets se déduit du temps de parcours ou de la différence de phase du rayon. Le capteur peut aussi faire le rapport de l’intensité reçue sur celle émise et ainsi déterminer la réflectance de la surface des objets. Le rayon est déplacé sur toute la surface à acquérir. Un tel outil d’acquisition devrait permettre d’obtenir la vérité terrain avec une très bonne précision. Mulligan *et al.* [Mulligan 01] utilisent ce dispositif car l’objectif est de comparer les performances d’algorithmes de mise en correspondance stéréoscopique dans le cadre particulier de la télé-présence. La vérité terrain d’une scène utile en télé-présence (un buste humain¹², cf. figure 2.5) a été obtenue à partir d’un scanner laser 3D reconstruisant le relief dans un volume cylindrique. Le scanner et les caméras ont préalablement été calibrés en utilisant une mire composée de plans avec des motifs. Les correspondances entre les points dans les images acquises par la caméra et les points dans les données du scanner ont été faites à la main (nous pouvons remarquer que la précision de l’acquisition peut être remise en question). Ensuite le buste a été acquis par la caméra et par le scanner.

Lumière structurée – Le système d’acquisition possède souvent une source de lumière, typiquement un laser ou un vidéo-projecteur, permettant de projeter sur la scène des motifs de structure connue. Selon le traitement effectué sur les images acquises, on peut classer les techniques à base de lumière structurée en deux catégories.

Dans la première catégorie, la lumière structurée est une aide à la mise en correspondance. En projetant sur la scène un point lumineux ou une ligne visible par les deux caméras, il devient plus facile d’apparier les pixels qui correspondent aux points de la scène ainsi éclairés. La résolution de l’appariement dépend de la qualité de la lumière, une lumière laser offrant une meilleure qualité. Dans [Kang 95], Kang *et al.* utilisent un système actif avec quatre caméras permettant une acquisition en temps réel et en projetant des motifs sinusoidaux, et dans [Koschan 97], les motifs employés par Koschan et Rodehorst sont des bandes colorées. Devernay *et al.* [Devernay 02] projettent un motif aléatoire en niveau de gris sur la scène et s’appuient, entre autres, sur une mise en correspondance par corrélation.

¹². <http://www.cis.upenn.edu/~janem/rsrch/dataset/ReadMe.html>

Dans la seconde catégorie, la surface de la scène est déduite de la déformation des motifs sur cette surface. Avec plusieurs motifs et l'utilisation d'outils de la géométrie différentielle, il est ainsi possible de reconstituer le relief de la scène. Plus la distance entre la source lumineuse et la scène est grande et plus la lumière projetée diffuse. Dans [Guisser 00, Devernay 02], les auteurs utilisent cette technique dans le but de reconstruire un modèle 3D de la scène. Dans [Guisser 00], Guisser *et al.* projettent une grille sur l'image, extraient les courbes (qui correspondent aux projections, dans l'image, de droites formant la grille) puis déterminent les correspondances entre les courbes et les droites formant la grille.

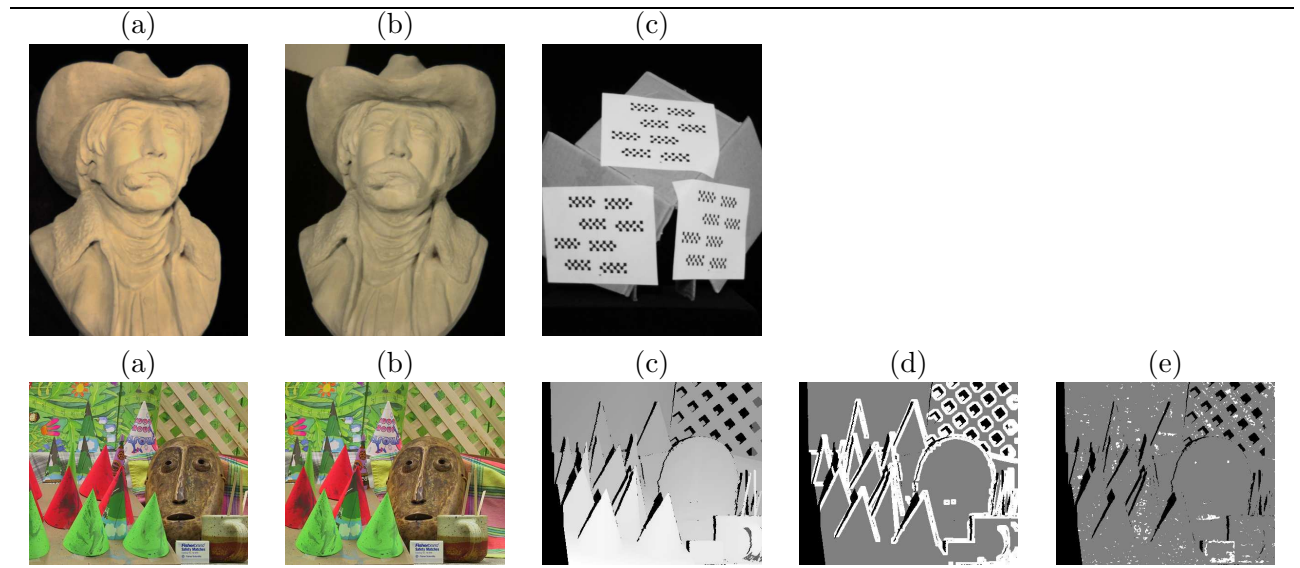


FIG. 2.5 – Couples avec vérité terrain obtenue par vision active – Pour les images *Buffalo Bill*, Mulligan *et al.* [Mulligan 01] ont acquis quatre images qui correspondent à quatre positions de la caméra. Ici, nous présentons la position gauche (a) et la position droite (b) ainsi que la mire de calibrage (c). Les auteurs décrivent une méthode qui permet d'obtenir la vérité terrain à l'aide d'un scanner 3D, mais ils ont fourni uniquement les paramètres de calibrage (il n'y a pas de carte de disparités disponible). Pour les images *cones*, Scharstein et Szeliski [Scharstein 03] ont acquis neuf images. Ici, nous présentons les images gauche (a) et droite (b) pour lesquelles les auteurs ont donné la carte de disparités et la carte des occultations, obtenues en utilisant une technique de lumière structurée, que nous avons fusionnées en (c). Les auteurs fournissent aussi une carte des zones non texturées et une carte des discontinuités de profondeur. Pour les cartes (d) et (e), les pixels blancs correspondent respectivement aux pixels proches d'une discontinuité et aux pixels dans une zone non texturée. Les pixels noirs correspondent aux pixels occultés.

Scharstein et Szeliski [Scharstein 03] utilisent une technique qui fait partie des deux catégories ci-dessus puisque la lumière structurée permet de reconstruire le relief en estimant la déformation sur la surface de la scène et, en plus, elle aide à la mise en correspondance. La lumière projetée ici est composée de bandes horizontales ou de bandes verticales dont le but est d'identifier tous les pixels de la scène, c'est-à-dire de leur affecter une signature unique. Plusieurs motifs sont utilisés et, pour chaque motif, les pixels correspondent à des points de la scène éclairés ou non. Avec un nombre suffisant de motifs, on peut associer un code binaire unique à chaque pixel. L'appariement est sans difficulté et sans ambiguïté. D'autre part, les disparités pour chaque couple caméra-projecteur sont estimées en utilisant la géométrie différentielle. Les disparités de mise en correspondance et celles d'illumination sont fusionnées. Deux difficultés subsistent :

- Il est délicat de déterminer si un pixel est illuminé ou non.

- Certains points de la scène sont toujours situés dans une zone d'ombre quel que soit le motif. Plusieurs sources de lumière ambiante situées à des positions différentes peuvent alors s'avérer nécessaires pour identifier tous les pixels en fusionnant l'information provenant des appariements pour chaque source. Malgré tout, il reste des pixels correspondant à une zone d'ombre et, par conséquent, les cartes de disparités obtenues possèdent des pixels dont la disparité est inconnue.

Pour les images acquises (cf. figure 2.5), une grande translation a été effectuée entre les deux prises de vue. Les images présentent des objets complexes, des occultations et des violations de la contrainte d'ordre¹³.

2.3.3.4 Contraintes géométriques sur la scène

Une autre technique consiste à faire des hypothèses sur la scène et s'appuie sur les points suivants :

- faire des hypothèses sur la scène ou le mouvement des caméras ;
- élaborer un modèle géométrique à partir de ces hypothèses ;
- trouver les paramètres de ce modèle ;
- calculer les correspondances en utilisant ce modèle.

Nous pouvons distinguer deux types d'hypothèses :

- *Hypothèses sur le mouvement des caméras* – Dans [Russ 99], les images sont acquises depuis des positions équidistantes de la scène le long d'un cercle et Russ et Reeves font l'hypothèse que les projections de tous les objets de la scène suivent un mouvement circulaire dans la séquence d'images. La profondeur d'un objet est directement proportionnelle au diamètre du cercle que suit la trajectoire de l'objet dans l'image.
- *Hypothèses sur la scène* – Une technique beaucoup plus populaire est de faire l'hypothèse que la scène est plane par morceaux. Dans ce cas, il existe une transformation homographique entre chaque « région plane » (en réalité projection d'une surface plane) de l'image gauche et la région plane correspondante dans l'image droite. Ces techniques ne sont pas particulièrement dédiées à l'obtention de vérité terrain mais plutôt à la reconstruction 3D de scènes [Schindler 03] et, entre autres, d'immeubles [Schmid 97]. Ces techniques nécessitent une segmentation préalable, des images en plans, c'est-à-dire de déterminer, dans l'image, les différentes régions qui « sont les projections des plans de la scène ». Cette segmentation se fait le plus souvent manuellement [Szeliski 99b]. Certains auteurs ont proposé une segmentation automatique :
 - Schindler [Schindler 03] s'appuie sur l'estimation des homographies et la géométrie épipolaire (évoquée au paragraphe 1.6.2) pour déterminer les plans, en considérant des scènes sans occultation.
 - Koch [Koch 96] effectue une mise en correspondance initiale à base de corrélation et de programmation dynamique, interpole la carte des disparités puis effectue la segmentation en plans grâce à cette carte.
 - Boufama et O'Connell [Boufama 02] qui calculent les homographies à partir des appariements de trois points d'intérêt (qui correspondent à des points contour) et des épipôles (cf. paragraphe 1.6.2). Puis, une croissance de région qui s'appuie, entre autres, sur une mesure de corrélation est effectuée.

Ensuite, pour chaque plan, il faut estimer l'homographie [Kanatani 99, Irani 02, Pollefeys 02], calculer les disparités et déduire les occultations. Les travaux qui s'appuient sur des hypothèses sur la scène pour établir des données de référence pour l'évaluation de méthodes de mise en correspondance sont (cf. figure 2.6) :

- Szeliski et Zabih [Szeliski 99b] qui proposent le couple d'images *map*.

13. <http://cat.middlebury.edu/stereo/newdata.html>

- Scharstein et Szeliski [Scharstein 02] qui proposent sur le même modèle que dans [Szeliski 99b] de nouvelles images¹⁴, dont *sawtooth* et *venus*, qui sont intégrées dans leur protocole d'évaluation et de comparaison.
- Šára et Bajcsy [Šára 97] et Kostková et Šára [Kostková 03] fournissent la vérité terrain pour une scène, *stairs*, constituée de cinq plans longs et fins qui s'occulent les uns les autres¹⁵. Ils utilisent une source de lumière secondaire (utilisation des ombres) et une segmentation manuelle.

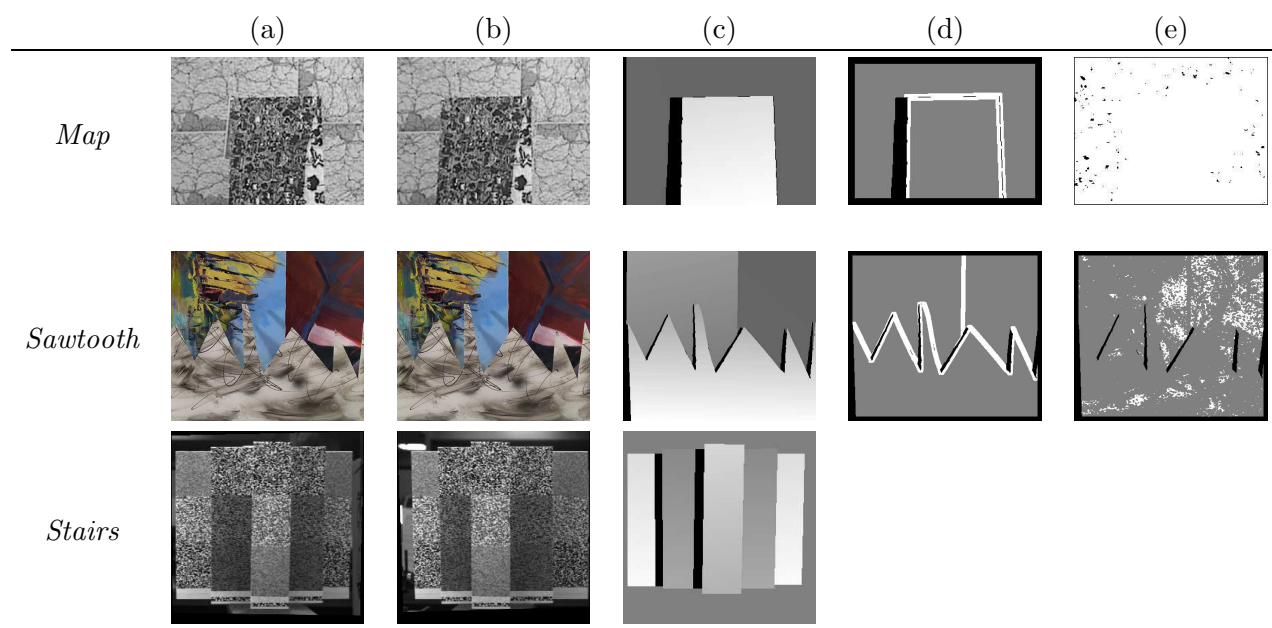


FIG. 2.6 – Couples avec vérité terrain obtenue à partir de contraintes géométriques – Pour tous ces exemples, nous donnons, en (a) et (b), les images gauche et droite, en (c), la carte de disparités qui respecte notre règle de représentation. Pour les cartes (d) et (e), les pixels blancs correspondent respectivement aux pixels proches d'une discontinuité et aux pixels dans une zone non texturée et les pixels noirs correspondent aux pixels occultés, excepté pour l'image map, en (e) où les pixels noirs correspondent aux pixels dans une zone non texturée (la détection de ce type de zone est abordée au paragraphe 2.8.2). Pour map, Szeliski [Szeliski 99a] a acquis un couple d'images. Pour sawtooth, Scharstein et Szeliski [Scharstein 02] ont acquis neuf images. Dans toutes ces images, le mouvement entre chaque image est une faible translation. Pour stairs, les auteurs de [Šára 97, Kostková 03] ont acquis quatre images.

2.3.4 Résumé sur les images testées dans la littérature

Ainsi trois possibilités sont envisageables pour le choix des images à tester :

- *Les images de synthèse* – Elles permettent une évaluation fiable puisqu'on possède la vérité terrain mais, pour l'instant, les images disponibles manquent de réalisme.
- *Les images réelles sans vérité terrain* – N'importe quel type de scène peut être testé mais il est impossible de réaliser une évaluation quantitative.

14. <http://cat.middlebury.edu/stereo/data.html>

15. <http://cmp.felk.cvut.cz/~sara/Stairs/home.html>

- *Les images avec vérité terrain* – Elles permettent une évaluation quantitative mais elles sont souvent restreintes à des scènes particulières et la vérité terrain n'est pas fiable et difficile à obtenir.

Même si les données de la dernière catégorie nous semblent les plus nécessaires, les trois catégories sont complémentaires et doivent, à ce titre, apparaître dans un protocole d'évaluation des performances.

2.4 Zones d'évaluation

Pour évaluer le comportement des méthodes face à différentes difficultés, certains protocoles ont distingué plusieurs zones :

- *Zones des occultations* – Elles correspondent toujours aux zones pour lesquelles les pixels n'ont pas de correspondant, c'est-à-dire les discontinuités de profondeur avec une partie visible dans une seule image. C'est la technique d'obtention de la vérité terrain qui permet de les déterminer [Sato 96, Scharstein 02].
- *Zones non texturées* – Ce sont généralement des zones difficiles à apparier, c'est pourquoi de nombreux auteurs ont pris en compte ce type de zones. Le plus souvent, elles sont détectées en utilisant le gradient de l'image, cf. [Scharstein 02], ou manuellement, cf. [Birchfield 98, Scharstein 02].
- *Zones de surfaces très inclinées* – Elles correspondent aux zones où il peut y avoir des raccourcissements importants. Elles sont détectées manuellement [Birchfield 98].
- *Zones des discontinuités* – Ce sont les zones où les pixels ont des disparités très différentes de celles de leurs voisins (mais il n'y a pas d'occultation). Elles correspondent aux discontinuités de profondeur sans partie cachée. Elles sont détectées en examinant la carte de disparités, cf. [Birchfield 98, Scharstein 02].

2.5 Critères d'évaluation

2.5.1 Critères sans vérité terrain

Les critères d'évaluation les plus populaires sont l'aspect visuel des cartes de disparités et le temps d'exécution.

La plupart des publications qui utilisent des tests sans vérité terrain se contentent d'effectuer une validation visuelle. Il y a très peu de critères proposés pour les images sans vérité terrain et ceux-ci sont évalués à partir du couple d'images et des matrices de projection des caméras :

- [Leclerc 00] – Il s'agit d'un critère d'auto-consistance que l'on peut appliquer avec au moins trois images. Si on prend le cas particulier de trois images I_1 , I_2 et I_3 , le principe du critère est le suivant :

Si $\mathbf{p}_1^{i,j}$, dans I_1 , a pour correspondant $\mathbf{p}_2^{i,v}$, dans I_2 et $\mathbf{p}_3^{i,r}$, dans I_3 alors les points \mathbf{P} et \mathbf{P}' reconstruits respectivement à partir des correspondances $(\mathbf{p}_1^{i,j}, \mathbf{p}_2^{i,v})$ et $(\mathbf{p}_1^{i,j}, \mathbf{p}_3^{i,r})$ doivent être identiques.

Ainsi, les auteurs calculent, pour tous les pixels de chaque image, la distance entre le point de la scène retrouvé respectivement à partir des correspondances entre I_1 et I_2 et I_1 et I_3 . L'histogramme des valeurs de ces distances constitue le critère d'évaluation.

- [Szeliski 99a, Carneiro 04] – De même, ce critère nécessite plus de deux images. L'ensemble des images est divisé en deux sous-ensembles, les images de prédiction et les images de validation. Les images de prédiction sont apparées et utilisées pour générer des nouvelles vues. Ces nouvelles vues sont comparées aux images de validation (on calcule une différence d'images).

Ainsi, nous voyons que d'une part, ces critères d'évaluation s'appuient sur l'utilisation de plus de deux images, et d'autre part, ils nous semblent difficiles à interpréter simplement.

2.5.2 Critères avec vérité terrain

Si l'on dispose de la vérité terrain, il existe de très nombreux critères. Nous pouvons distinguer les critères s'appliquant sur toute l'image des critères s'appliquant dans des zones particulières. Nous notons d_{ref} la fonction de disparité théorique et \mathbf{occ} la disparité symbolique attribuée à un pixel occulté. De plus, nous posons :

$$\text{Err}_g^{i,j} = \|d(\mathbf{p}_g^{i,j}) - d_{\text{ref}}(\mathbf{p}_g^{i,j})\|. \quad (2.1)$$

Le calcul de $\text{Err}_g^{i,j}$ n'a de signification que si $d(\mathbf{p}_g^{i,j}) \neq \mathbf{occ}$ et $d_{\text{ref}}(\mathbf{p}_g^{i,j}) \neq \mathbf{occ}$. Les critères que l'on rencontre habituellement sont les suivants :

1. *Pourcentage d'appariements corrects* – Un appariement est correct si $\text{Err}_g^{i,j} = 0$.
2. *Pourcentage d'appariements précis* – Un appariement est précis si $\text{Err}_g^{i,j} < 1$. Ce critère est utilisé notamment dans [Faugeras 92, Satoh 96, Lan 97, Belli 00, McCane 01, Eklund 03, Tesser 03].
3. *Pourcentage d'appariements acceptés* – Un appariement est accepté si $1 \leq \text{Err}_g^{i,j} < 2$, dans le cas où la disparité est réelle, ou si $\text{Err}_g^{i,j} = 1$, dans le cas d'une disparité entière, cf. [Lan 97, McCane 01] .
4. *Pourcentage d'appariements mauvais* – L'appariement est mauvais si $2 \leq \text{Err}_g^{i,j} < 3$, dans le cas où la disparité est réelle, ou si $\text{Err}_g^{i,j} = 2$, dans le cas d'une disparité entière, cf. [Lan 97, McCane 01].
5. *Pourcentage d'appariements erronés* – L'appariement est erroné si $\text{Err}_g^{i,j} \geq 3$, dans le cas où la disparité est réelle, ou si $\text{Err}_g^{i,j} \neq 0$, dans le cas d'une disparité entière, cf. [Lan 97, McCane 01].
6. *Erreur moyenne* – La moyenne des erreurs, utilisée par [Fua 93, Lundqvist 03], est donnée par :

$$\overline{\text{Err}}_g = \frac{1}{N_g^{\text{lig}} N_g^{\text{col}}} \sum_{i=0}^{N_g^{\text{lig}}-1} \sum_{j=0}^{N_g^{\text{col}}-1} \text{Err}_g^{i,j}. \quad (2.2)$$

7. *Erreur quadratique moyenne* – C'est la racine de la moyenne des carrés des erreurs, *Root mean square error*. Elle est très utilisée, cf. [Sankar Kishore 00, Fielding 00, Scharstein 02, Eklund 03, Reddy 04], et elle est donnée par :

$$\text{RMS}_g = \sqrt{\frac{1}{N_g^{\text{lig}} N_g^{\text{col}}} \sum_{i=0}^{N_g^{\text{lig}}-1} \sum_{j=0}^{N_g^{\text{col}}-1} (\text{Err}_g^{i,j})^2}. \quad (2.3)$$

8. *Faux positif* – Un appariement est considéré comme un faux positif lorsqu'une correspondance a été établie pour un pixel de gauche alors qu'il y a une occultation, c'est-à-dire, si $d(\mathbf{p}_g^{i,j}) \neq d_{\text{ref}}(\mathbf{p}_g^{i,j})$ et $d_{\text{ref}}(\mathbf{p}_g^{i,j}) = \mathbf{occ}$ [Fielding 00, Egnal 02, Tesser 03].
9. *Faux négatif* – Un appariement est considéré comme un faux négatif lorsqu'un pixel de gauche est considéré comme occulté alors qu'il ne l'est pas, c'est-à-dire, si $d(\mathbf{p}_g^{i,j}) \neq d_{\text{ref}}(\mathbf{p}_g^{i,j})$ et $d(\mathbf{p}_g^{i,j}) = \mathbf{occ}$ [Egnal 02, Tesser 03].

Les critères 1, 2 et 5 permettent d'évaluer les performances générales de la méthode et les critères 3 et 4 affinent cette évaluation au niveau des appariements erronés. Le critère 3 permet de voir si la méthode, lorsqu'elle obtient un faible pourcentage d'appariements corrects, donne tout de même une approximation acceptable des disparités. Les critères 6 et 7 permettent d'affiner encore plus l'évaluation des types d'erreurs et ainsi la précision de la méthode. Les critères 8 et 9 sont représentatifs des erreurs

induites par les discontinuités et les occultations. On peut aussi associer à ces deux derniers critères le calcul du pourcentage de pixels correctement détectés comme occultés, que nous appelons les vrais négatifs, comme dans [Sato 96].

2.6 Méthodes évaluées

Nous pouvons distinguer trois cas :

- les protocoles où seules les méthodes proposées par les auteurs sont évaluées et comparées ;
- les protocoles où les méthodes proposées par les auteurs sont évaluées et comparées avec une méthode choisie arbitrairement, le plus souvent une méthode à base de corrélation, comme dans [Sato 96, Di Stefano 03, Trujillo 04] ;
- les publications qui exploitent le protocole de Scharstein et Szeliski [Scharstein 02].

Parmi les travaux les plus complets, nous pouvons citer :

- Aschwanden et Guggenbül [Aschwanden 92] qui effectuent une évaluation des mesures de corrélation en comparant dix-neuf mesures différentes ;
- Falkenhagen [Falkenhagen 97] qui compare une méthode à base de corrélation et une méthode utilisant une approche multirésolution avec sa méthode (mise en correspondance multirésolution à base de corrélation) ;
- Belli *et al.* [Belli 00] qui testent une méthode en utilisant des images couleur et en utilisant des images en niveaux de gris.
- Egnal et Wildes [Egnal 02] qui comparent cinq algorithmes à base de corrélation (sans contrainte, avec contrainte de symétrie, en ajoutant au coût des contraintes une mesure d'ambiguïté, une mesure d'imprécision ou une mesure d'entropie).
- Scharstein et Szeliski [Scharstein 02] qui effectuent la comparaison de méthodes de mise en correspondance de pixels la plus exhaustive à notre connaissance (nous la détaillerons dans le paragraphe 2.8).
- Tappen et Freeman [Tappen 03] qui établissent une comparaison de différentes méthodes de coupure de graphe.
- Škerl *et al.* [Škerl 04] qui effectuent une évaluation des mesures de corrélation en comparant neuf mesures différentes.

2.7 Synthèse des critères et classement des méthodes

La plupart des protocoles d'évaluation cités jusqu'à présent n'utilisent que très peu de critères et il n'est donc pas difficile de classer les méthodes. Dans [Aschwanden 92], les auteurs comparent dix-neuf méthodes par rapport à leur résistance au bruit (gaussien et impulsionnel) et au zoom. Trois critères sont évalués, les résultats sont fournis, mais aucun classement n'est proposé. Il est donc peu évident de faire émerger une méthode ou un groupe de méthodes. Scharstein et Szeliski [Scharstein 02] sont les seuls à proposer un protocole qui classe les méthodes à partir de trois critères. Nous détaillerons celui-ci dans le paragraphe suivant.

2.8 Protocole d'évaluation et de comparaison de Scharstein et Szeliski

Ce protocole est le plus utilisé. Les auteurs ont mis dix couples d'images avec vérité terrain à disposition sur le web¹⁶. Leur évaluation s'appuie sur trois critères. Depuis leur site web, l'utilisateur peut soumettre ses résultats (sous la forme d'une carte de disparités); il peut alors connaître son classement par rapport aux autres méthodes et laisser ses résultats apparaître sur le site, s'il le désire. Ce site a permis de récolter les résultats de quarante méthodes, ce qui constitue, à notre avis, la plus grande comparaison qui ait été réalisée. Certains articles utilisent les images et les critères de ce protocole, comme dans [Jagmohan 04], sans laisser apparaître sur le site les résultats obtenus.

Ce site est toujours en évolution. En août 2005, une nouvelle évaluation a débuté et, en octobre 2005, elle est toujours en évolution. Aussi, dans chaque paragraphe, nous aborderons cette nouvelle évaluation mais celle-ci est susceptible d'avoir été modifiée au moment de la lecture de ce mémoire.

2.8.1 Images testées

Dans l'ordre chronologique (cf. figure 2.7), les images proposées sont :

- *Tsukuba* – Au départ, ces images (cf. figure 2.6) proviennent de Nakamura *et al.* [Nakamura 96] qui ont acquis vingt-cinq images et ont fourni une vérité terrain extraite manuellement. Scharstein et Szeliski [Scharstein 02] ont repris ces images et fournissent une carte de disparités avec une précision au quart de pixel¹⁷. Il y a uniquement sept disparités différentes. Il y a aussi une localisation des occultations. Il ne faut pas prendre en compte le bord de l'image de dix-huit pixels d'épaisseur pour lequel la vérité terrain n'est pas fournie. Ces images sont les plus utilisées dans la littérature, pourtant, il nous semble que la technique d'obtention de la vérité terrain est la moins fiable.
- *Map* – Ces images proposées dans [Szeliski 99b] ont été acquises avec un système trinoculaire. Les plans de la scène ont été extraits manuellement puis Szeliski et Zabih ont estimé l'homographie pour chaque plan et la carte de disparités a été calculée. La précision est au huitième de pixel (cf. figure 2.6).
- *Sawtooth, venus, bull, poster, barn 1 et barn 2* – Scharstein et Szeliski [Scharstein 02] ont repris le même principe que celui utilisé dans [Szeliski 99b]. Cette fois, neuf images sont acquises, avec une faible translation entre deux images successives (cf. figures 2.6 et 2.7).
- *Cones et teddy* – Ce sont les dernières images acquises par Scharstein et Szeliski [Scharstein 03]. La technique utilisée est totalement différente puisque les auteurs ont fait appel à la vision active. Il y a toujours neuf images disponibles. Les cartes de disparités présentent des pixels pour lesquels la disparité n'est pas connue (cf. figures 2.5 et 2.7).

Parmi toutes ces images, seules *tsukuba*, *map*, *sawtooth* et *venus* ont été utilisées pour effectuer leurs premières évaluations.

Dans le but d'être plus pertinente, la nouvelle évaluation en cours prend en compte les images *tsukuba*, *venus*, *cones* et *teddy*. En effet, les trois images *map*, *sawtooth* et *venus* sont assez similaires, c'est-à-dire qu'elles représentent les mêmes difficultés, et les images *cones* et *teddy* sont de plus en plus prises en compte dans les nouvelles publications.

16. <http://cat.middlebury.edu/stereo/>

17. <http://cat.middlebury.edu/stereo/data/imagehtml/tsukuba.html>

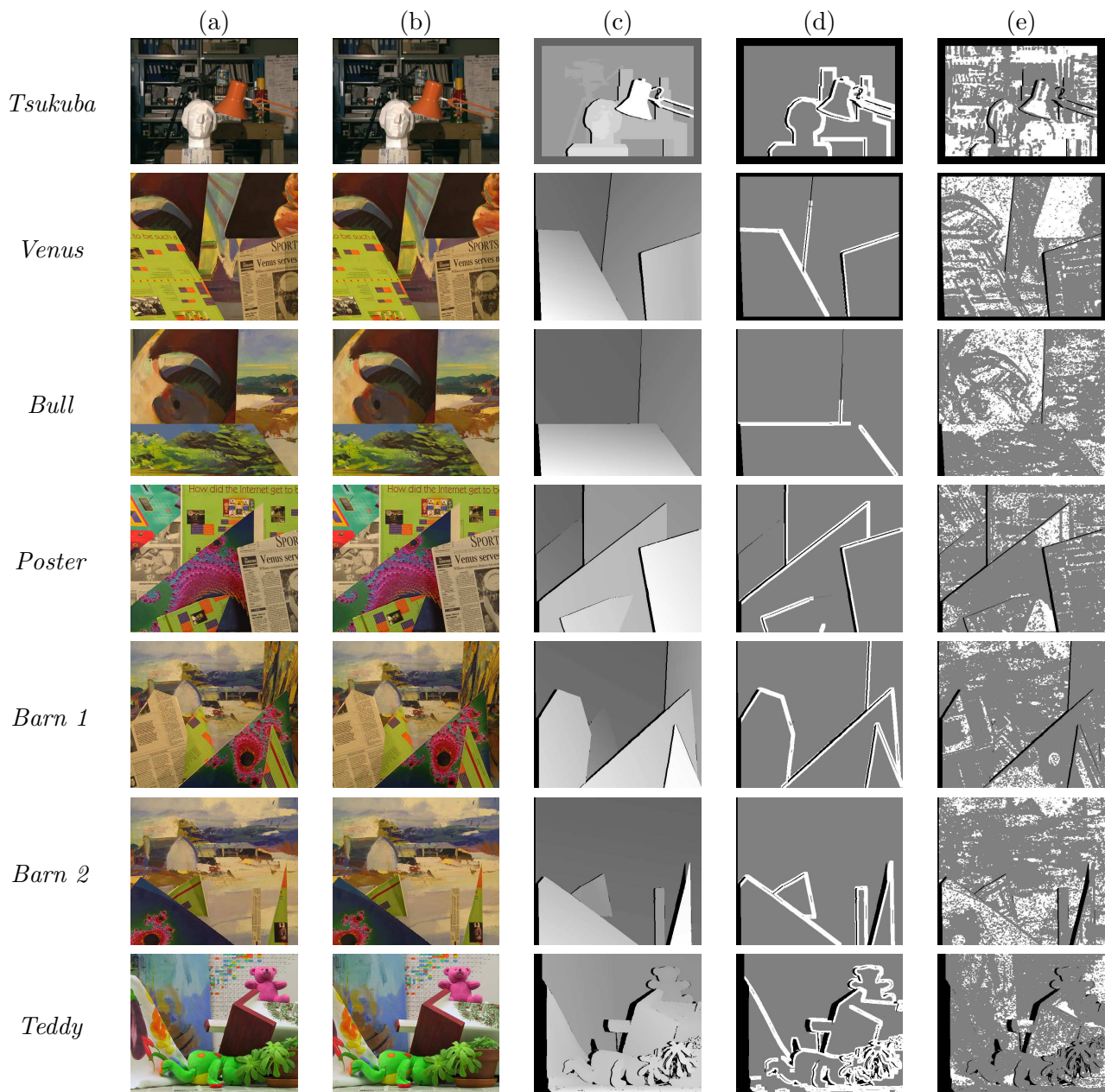


FIG. 2.7 – Exemples de couples proposés par Scharstein et Szeliski – Les images utilisées pour classer les méthodes, dans la première évaluation proposée, sont *tsukuba*, *map*, *sawtooth* (cf. figure 2.6) et *venus* alors que dans la seconde, il s’agit de *tsukuba*, *venus*, *cones* (cf. figure 2.5) et *teddy*. Pour chaque couple (a) et (b), les auteurs donnent la carte de disparités et la carte des occultations que nous avons fusionnées (d), la carte des discontinuités (e) et la carte des zones non texturées (f). Dans (e) et (f), les pixels noirs correspondent aux pixels occultés et les pixels blancs correspondent respectivement aux pixels proches d’une discontinuité et aux pixels dans une zone non texturée.

2.8.2 Zones d’évaluation

Scharstein et Szeliski [Scharstein 02] proposent de distinguer six catégories de pixels dans l’image (on peut distinguer ces catégories sur la figure 2.7) : tous les pixels, les pixels non occultés, les pixels occultés, les pixels dans une zone texturée, les pixels dans une zone non texturée, les pixels proches

d'une discontinuité. Les pixels occultés et les pixels proches d'une discontinuité sont détectés à partir de la carte de disparités théoriques. Dans la première version de leur évaluation, la zone des pixels proches des discontinuités ne contenaient pas les pixels occultés ou proches des occultations.

La seconde version définit comme zone des discontinuités tous les pixels proches d'une occultation ou d'une discontinuité (les auteurs ne définissent pas précisément ce qu'ils entendent par « proches »). Les pixels dans une zone non texturée sont déterminés en appliquant un seuil sur la composante horizontale du gradient horizontal (les auteurs ne justifient pas cette technique).

2.8.3 Critères d'évaluation

Scharstein et Szeliski [Scharstein 02] proposent seize critères d'évaluation :

- *Calcul de l'erreur quadratique moyenne* – Ils proposent d'évaluer le RMS pour les six catégories de pixels dans l'image.
- *Appariements erronés* – Comme dans [Lan 97], ils considèrent qu'un appariement est erroné quand l'erreur dépasse un seuil T_e :

$$B = \sum_{i=0}^{N_g^{\text{lig}}-1} \sum_{j=0}^{N_g^{\text{col}}-1} J(\text{Err}_g^{i,j} > T_e), \text{ avec } J(E) = 0 \text{ si } E \text{ est faux ou } 1 \text{ sinon.} \quad (2.4)$$

Ce critère est aussi évalué dans les six types de zones considérées.

- *Erreurs de prédiction* – Le principe est de construire une des sept autres vues en utilisant une image comme image de référence et la carte de disparités associée. Puis on calcule la différence des images (entre l'image construite et celle de départ).

Sur leur site, seuls trois critères sont évalués :

- Pour la première évaluation proposée, il s'agit des appariements erronés dans toute l'image, dans les zones des discontinuités et dans les zones non texturées. Ils ne prennent en compte ni les pixels occultés, ni les pixels de la bordure (dix pixels pour toutes les images, sauf *tsukuba* avec dix-huit). Pour *map*, ils ne calculent pas les erreurs d'appariement dans les zones non texturées car leur surface est très petite par rapport à la taille des images (elles représentent moins de 1% de l'image).
- Pour la seconde évaluation, les appariements erronés dans toute l'image, dans la zone des discontinuités et dans la zone des occultations sont évalués. En perspective, ils désirent distinguer les méthodes qui fournissent des disparités dans les zones des occultations des méthodes qui donnent des cartes des occultations (et ainsi pas de disparité dans les zones des occultations).

2.8.4 Synthèse des critères et classement des méthodes

Dans la première évaluation, pour chaque critère, Scharstein et Szeliski [Scharstein 02] classent les méthodes. Puis la somme de tous les rangs est effectuée pour chaque méthode. Les méthodes sont finalement classées grâce à cette valeur. Ce classement est statique.

Leur seconde évaluation fait intervenir un classement dynamique, c'est-à-dire que l'utilisateur peut choisir de classer la méthode suivant : son rang global, ses résultats pour une image ou ses résultats pour un critère particulier. L'utilisateur peut aussi changer les seuils qui interviennent dans les critères.

Quarante méthodes ont concouru lors de la première évaluation. Pour la seconde évaluation qui est en cours, il y a déjà huit classements possibles, c'est-à-dire que l'utilisateur peut classer les résultats de huit manières différentes, en fonction : d'un critère en particulier (il y a trois critères), des résultats obtenus pour une image en particulier (il y a quatre images testées) ou du rang global obtenu par la méthode. Pour l'instant, seule une partie des quarante premières méthodes testées apparaissent

dans cette nouvelle évaluation. C'est pourquoi, nous ne présenterons ici que le classement issu de la première évaluation.

Sur leur site, Scharstein et Szeliski [Scharstein 02] fournissent le classement des méthodes, en précisant les résultats obtenus sur chaque critère, pour chaque image et en donnant les rangs pour chaque critère. De plus, une carte de disparités, une carte du signe des erreurs de disparités et une carte de visualisation des appariements erronés (qui présentent une erreur de plus d'un pixel) sont ajoutées. Dans le tableau 2.1, nous donnons, d'une part, le classement fourni par Scharstein et Szeliski (cela correspond à la colonne R) et, d'autre part, nous complétons ce classement en précisant les différents éléments constitutifs pris en compte par la méthode (colonnes TYP à MUL) et en indiquant si la méthode prend en compte les aspects qui nous intéressent, à savoir : l'utilisation de la couleur, la prise en compte d'une mesure de corrélation et la prise en compte explicite des occultations (colonnes COU à COR). Nous utilisons les notations suivantes :

- TYP – Le type de méthode est : locale (L) ou globale (G).
- OPT – La méthode d'optimisation. Lorsqu'il s'agit d'une méthode globale, nous utilisons les abréviations suivantes :
 - AG – Algorithme génétique ;
 - CG – Coupure de graphe ;
 - PM – Passages multiples avec croissance de germes ;
 - PC – Propagation de croyance ;
 - PD – Programmation dynamique ;
 - RE – Relaxation ;
 - RG – Recherche gloutonne ;
 - RN – Réseau de neurones ;
 - RS – Recuit simulé.

Lorsqu'il s'agit d'une méthode locale, nous précisons :

- AC – Algorithme coopératif ;
- WTA – *Winner take all*.
- AFF – L'étape d'affinement :
 - SP – Sous-pixel ;
 - Sym – Symétrie.
- PRÉ – L'étape de prétraitement :
 - Fil – Filtrage ;
 - Seg – Segmentation.
- MUL – L'utilisation de la multirésolution.
- Dans ce travail, nous nous intéressons particulièrement aux méthodes de mise en correspondance par corrélation, à la prise en compte des occultations et à l'utilisation de la couleur, c'est pourquoi, nous précisons pour chaque méthode :
 - COU – Utilisation de la couleur ;
 - OCC – Prise en compte des occultations ;
 - COR – Utilisation d'une mesure de corrélation.
- R – Le rang de la méthode dans la comparaison exposée sur le site de Scharstein et Szeliski.

2.8.5 Méthodes évaluées

Grâce au tableau récapitulatif 2.1, nous pouvons faire différentes remarques :

- La plupart des méthodes sont globales (trente-deux méthodes).

RÉFÉRENCE	TYP	OPT	PRÉ	AFF	MUL	COU	OCC	COR	R
[Sun 05]	G	PC	Seg	-	-	oui	oui	-	1
[Deng 05]	G	CG	Seg	-	-	oui	oui	-	2
[Hong 04]	G	CG	Seg	-	-	oui	oui	SAD	3
[Bleyer 05]	G	CG	Seg	-	-	oui	oui	SSD	4
[Bleyer 04]	G	RG	Seg	-	-	oui	oui	SAD	6
[Sun 03]	G	PC	-	-	-	oui	oui	-	7
[Lin 02]	G	CG	Seg	SP	oui	oui	oui	-	8
[Kim 05]	G	PD	Fil	SP	-	-	oui	SAD	9
[Kolmogorov 01]	G	CG	-	-	-	-	oui	-	11
[Wei 04]	G	PM	Seg	-	-	oui	oui	SAD, SSD	10
[Kolmogorov 02]	G	CG	-	-	-	oui	-	SSD	12
[Kolmogorov 01]	G	CG	-	-	-	-	oui	-	13
[Mayer 03]	L	WTA	Fil	SP	-	oui	oui	SAD, NCC	14
[Yoon 05]	L	WTA	-	-	oui	oui	-	SAD, SSD	15
[Goulermas 03]	G	AG	-	SP	-	-	-	ZNCC	16
[Agrawal 04]	G	CG	-	-	-	oui	oui	SAD	17
[Veksler 03]	G	PD	-	-	-	-	oui	-	18
[Scharstein 02]	G	CG	-	-	-	oui	-	-	19
[Gong 05]	G	PD	-	-	-	oui	oui	-	20
[Birchfield 99]	G	CG	Fil	-	-	-	oui	-	21
[Boykov 01]	G	CG	-	-	-	-	oui	-	22
[Veksler 05]	G	PD	-	-	-	-	-	-	23
[Criminisi 05]	G	PD	-	-	-	-	oui	ZSSD	24
[Veksler 01]	L	WTA	-	-	-	-	-	-	25
[Hirschmüller 02]	L	WTA	-	Fil	-	-	oui	SAD	26
[Zitnick 00]	L	AC	-	-	-	-	oui	NCC, SAD	27
[Brockers 04a]	G	RE	-	SP, Sym	-	-	oui	ZNCC	28
[Scharstein 98]	G	RN	-	SP	-	oui	-	SSD	29
[Lee 02]	G	RS	-	-	oui	oui	oui	-	30
[Gong 02]	G	AG	-	-	oui	oui	oui	SAD, SSD	31
[Scharstein 02]	L	WTA	-	SP	-	oui	oui	SSD	32
[Birchfield 98]	G	PD	-	-	-	-	oui	-	33
[Roy 98]	G	CG	-	-	-	-	oui	SSD	34
[Scharstein 02]	G	PD	-	SP	-	oui	-	-	35/36
[Forstmann 04]	G	PD	-	oui	oui	oui	oui	-	37
[Mühlmann 01]	L	WTA	Fil	SP, Sym	-	oui	oui	SAD	38
[Shao 02]	G	RE	-	-	oui	-	-	NCC	39
[Sun 02]	G	PD	-	SP	-	-	-	ZNCC	40

TAB. 2.1 – Tableau récapitulatif des méthodes ayant participé au test de Scharstein et Szeliski – Le caractère « - » signifie que l'élément n'existe pas dans la méthode considérée. La méthode de rang 5 est une soumission anonyme, nous n'avons aucune information sur cet article. Dans [Kolmogorov 02] le choix des paramètres est automatique, contrairement au choix effectué dans [Kolmogorov 01]. Dans [Scharstein 02], les méthodes classées 35 et 36 n'ont pas le même coût global.

- Les méthodes d'optimisation les plus utilisées sont celles à base de coupures de graphes (douze méthodes) et la programmation dynamique (dix méthodes). Les méthodes utilisant les coupures de graphe sont en général mieux classées. Ces méthodes sont les plus récentes et nous pouvons penser que ce comportement plus performant est en grande partie dû au choix d'un coût de correspondance plus élaboré plutôt qu'au choix de la méthode d'optimisation elle-même.
- Des méthodes récentes et bien classées (c'est le cas des quatre premières méthodes) utilisent une segmentation (sept méthodes).
- Peu de méthodes effectuent un calcul au sous-pixel (dix méthodes).
- Peu de méthodes utilisent une approche multirésolution (six méthodes).
- Plus de la moitié des méthodes utilisent des images couleur.
- La plupart des méthodes prennent en compte explicitement le problème des occultations.
- Plus de la moitié des méthodes utilisent une mesure de corrélation (vingt et une méthodes). Dans la majorité des cas, cette mesure intervient dans la méthode d'initialisation mais certaines méthodes l'emploient aussi dans le coût local de mise en correspondance.

Ce tableau comparatif nous montre que les méthodes les plus utilisées actuellement sont les méthodes globales associées à une segmentation couleur où les occultations sont prises en compte et où une mesure de corrélation est aussi utilisée.

2.9 Protocole d'évaluation et de comparaison proposé

Bien que le protocole de Scharstein et Szeliski soit le plus pertinent à notre connaissance, nous pensons qu'il n'est pas parfaitement adapté aux études comparatives que nous souhaitons réaliser. En effet, nous voulons comparer et évaluer des méthodes à base de corrélation et nous voulons surtout étudier le comportement de ces méthodes face aux occultations. De plus, nous souhaitons évaluer le gain de performance apporté par l'utilisation de la couleur. Nous allons donc proposer notre propre protocole. Pour cela, nous commençons par décrire la technique d'obtention de la vérité terrain que nous avons choisie et les nouvelles images que nous avons proposées. Ensuite, nous exposons les zones d'évaluation considérées et les critères utilisés. Enfin, nous décrivons la technique de synthèse des critères d'évaluation permettant d'obtenir le classement final des méthodes.

2.9.1 Obtention de la vérité terrain

2.9.1.1 Objectifs

Dans le but de proposer de nouvelles images de référence, un travail de DEA, cf. [Bocquillon 05a, Bocquillon 05b], a été effectué sur les techniques d'obtention de la vérité terrain. Nous avons choisi d'utiliser une technique à base de contraintes géométriques sur la scène car c'est celle qui nous semble la plus simple à mettre en place. Les différentes publications dans ce domaine montrent que les cartes de disparités obtenues sont suffisamment fiables pour servir de référence. Nous souhaitons compléter les travaux existants.

Tous les travaux utilisant une approche par contraintes géométriques fournissent des disparités qui nous paraissent fiables, mais ne localisent pas précisément les occultations. Les plans sont segmentés manuellement et nous aimerions augmenter la précision de cette segmentation dans le but d'améliorer la localisation des occultations. Dans les travaux cités, aucune publication ne donne une démarche suffisamment précise pour être facilement reproduite. Nous espérons pallier ce manque en fournissant un algorithme simple pour obtenir des images avec vérité terrain. Le but de ce travail est de développer un outil pour obtenir la vérité terrain à partir d'un couple d'images respectant l'hypothèse suivante : les scènes considérées sont polyédriques, c'est-à-dire composées uniquement de facettes planes. En

effet, deux images d'un plan sont liées par une homographie selon l'équation suivante :

$$\lambda \mathbf{m}_d = \mathbf{H} \mathbf{m}_g, \quad (2.5)$$

où le vecteur $\mathbf{m}_g = (u_g \ v_g \ 1)^T$ contient les coordonnées d'un point de l'image gauche, le vecteur $\mathbf{m}_d = (u_d \ v_d \ 1)^T$ contient les coordonnées de son correspondant dans l'image droite, \mathbf{H} est la matrice 3×3 contenant les paramètres de l'homographie et λ est un scalaire non nul. Connaissant \mathbf{H} , l'équation (2.5) permet d'obtenir les coordonnées du correspondant de n'importe quel point. Chaque polygone constitué par la projection d'une facette plane de la scène dans l'image gauche étant lié à son correspondant dans l'image droite par une homographie, l'estimation des matrices de toutes les homographies permet de calculer les disparités associées aux pixels de toute l'image.

2.9.1.2 Approche proposée

L'approche que nous proposons se déroule en cinq étapes :

- *Segmentation des images* – Il s'agit d'identifier les régions polygonales qui sont les projections des facettes planes de la scène. Cette étape est détaillée dans le paragraphe 2.9.1.3.
- *Détection des points d'intérêt dans l'image gauche* – Nous utilisons le détecteur de Harris et Stephens [Harris 88].
- *Mise en correspondance de points d'intérêt* – Nous utilisons une technique de mise en correspondance locale par corrélation (cf. algorithme 3.1) avec la mesure ZNCC (cf. équation (3.16)), la contrainte de symétrie (cf. paragraphe 1.6.7) et un affinement au sous-pixel par l'interpolation des niveaux de gris (les images sont en couleur mais l'obtention de la vérité terrain est effectuée à partir des niveaux de gris).
- *Estimation des homographies* – Nous avons testé trois méthodes d'estimation des paramètres des homographies : la méthode linéaire faisant intervenir un critère algébrique, la méthode non linéaire itérative de Levenberg-Marquardt avec un critère géométrique et la méthode robuste RANSAC (*Random Simple Consensus*) avec un critère géométrique. Cette dernière méthode permet d'avoir une estimation de qualité même en présence d'erreurs provenant de l'étape de mise en correspondance. Ces trois méthodes classiques sont décrites en détails dans [Hartley 04, chap. 3].
- *Calcul des disparités* – À partir des matrices des homographies estimées pour chaque polygone à l'étape précédente, les disparités sont calculées et les occultations sont localisées. Plus précisément, pour chaque pixel de chaque polygone de l'image gauche, les coordonnées du point qui est son image par l'homographie associée au polygone sont calculées. Si ce point appartient au polygone correspondant dans l'image droite alors il est considéré comme le correspondant, sinon il est considéré comme occulté.

2.9.1.3 Segmentation des images

Différentes techniques – Pour identifier les régions correspondant à des facettes planes de la scène, plusieurs approches ont été envisagées :

- utilisation d'une technique par croissance de régions [Wang 93] ;
- utilisation de techniques de détection du mouvement [Irani 95, Heisele 00] ;
- utilisation d'une mise en correspondance initiale [Koch 96, Mulligan 00] ;
- segmentation manuelle [Szeliski 99b, Scharstein 02] ;
- utilisation d'une segmentation couleur [Wei 04, Bleyer 04, Hong 04, Sun 05].

Dans le cas où la segmentation est manuelle, le bon nombre de plans est détecté (ce qui est beaucoup plus délicat avec les techniques automatiques), mais les frontières des régions sont souvent peu précises. C'est pourquoi, nous proposons la démarche semi-automatique suivante :

- segmentation manuelle grossière des images (il peut y avoir plusieurs pixels d'erreur) ;
- affinement automatique de la segmentation initiale.

Polygones actifs – Nous détaillons à présent l'affinement de cette segmentation initiale. D'après les hypothèses, nous savons que les facettes planes se projettent en des polygones dans l'image. Soit $Poly$ le polygone considéré et notons les N_s sommets de ce polygone s^1, \dots, s^{N_s} . Ces sommets sont reliés par les cotés $C^{(1,2)}, \dots, C^{(N_s-1, N_s)}$ et $C^{(N_s, 1)}$. Nous appliquons le modèle des contours actifs exposé dans [Blake 98] à des polygones, c'est pourquoi nous employons le terme de « polygones actifs ». Les contours actifs sont utilisés pour la détection de contours mais aussi pour le suivi de contours mobiles dans des séquences d'images. Ils correspondent à un contour élastique qui s'adapte à la forme de l'objet suivi et à son mouvement. Une énergie, $E(\mathcal{C})$, est associée à ce contour \mathcal{C} :

$$E(\mathcal{C}) = E_{\text{interne}}(\mathcal{C}) + E_{\text{externe}}(\mathcal{C}) + E_{\text{image}}(\mathcal{C}). \quad (2.6)$$

Le premier terme, $E_{\text{interne}}(\mathcal{C})$, modélise les contraintes de continuité de la courbe (terme de régularisation) et il prend en compte notamment sa longueur et sa courbure. Le second terme, $E_{\text{externe}}(\mathcal{C})$, prend en compte les informations *a priori*, comme les positions des points de contrôle. Le dernier terme, $E_{\text{image}}(\mathcal{C})$, correspond au terme d'attache aux données et est en rapport avec les caractéristiques photométriques de l'image à prendre en compte (on peut utiliser par exemple le contraste de l'image). Dans le cadre de notre utilisation, nous pouvons supprimer les termes suivants :

- $E_{\text{interne}}(\mathcal{C})$ car le polygone est formé de segments ;
- $E_{\text{externe}}(\mathcal{C})$ car nous n'avons pas de contraintes sur les positions des sommets.

Pour localiser le polygone, nous faisons les hypothèses suivantes :

- Le contour de départ est proche du contour final.
- Les contours des polygones correspondent à des contours de l'image.

Grâce à ces deux hypothèses, nous pouvons conclure que plus la valeur de l'amplitude du gradient de l'image, évaluée aux points appartenant aux cotés du polygone dans la direction orthogonale à celle du côté, est élevée, plus le contour est proche du contour recherché. Nous décidons donc d'utiliser un gradient directionnel pour le calcul de l'énergie du polygone. Ainsi, en s'inspirant du modèle fourni dans l'équation (2.6) et en prenant en compte les hypothèses énoncées précédemment, nous choisissons l'énergie suivante :

$$E(Poly) = E_{\text{image}}(Poly) = \sum_{C \in Poly} \int_a^b (\nabla I^s \cdot \mathbf{n}_C)^2 ds, \quad (2.7)$$

où a et b sont les extrémités du côté C , s est l'abscisse curviligne du point \mathbf{p}^s dont le niveau de gris associé est I^s , le vecteur gradient associé ∇I^s et qui appartient au côté C . La normale au côté C est \mathbf{n}_C . L'utilisation du produit scalaire permet de projeter le gradient dans la direction recherchée et l'utilisation du carré permet de s'affranchir du signe de cette projection. Nous aurions pu choisir la valeur absolue mais dans ce cas, l'expression ne serait pas différentiable. Nous avons besoin d'une forme discrète de l'équation (2.7) pour pouvoir l'utiliser, ce qui donne :

$$E(Poly) = \sum_{\mathbf{p}^{i,j} \in Poly} (\nabla I^{i,j} \cdot \mathbf{n}_C)^2, \quad (2.8)$$

où $\mathbf{p}^{i,j}$ est un point appartenant au polygone $Poly$. Le polygone doit donc être échantillonné en un nombre N_n de points. Pour comparer les différentes énergies, nous normalisons l'équation (2.8), qui

devient :

$$E(Poly) = \frac{1}{N_n} \sum_{\mathbf{p}^{i,j} \in Poly} (\nabla I^{i,j} \cdot \mathbf{n}_C)^2. \quad (2.9)$$

En pratique, nous utilisons l'opérateur de Sobel pour calculer les composantes du vecteur gradient. Les points d'échantillonnage n'étant pas des pixels entiers, nous calculons les valeurs du gradient par interpolation bicubique.

Nous avons envisagé deux techniques pour résoudre ce problème de minimisation, une recherche gloutonne et une recherche par recuit simulé.

Recherche gloutonne – Nous définissons Ω , l'espace des polygones constitué de tous les polygones possibles obtenus en déplaçant les sommets autour de leurs positions initiales dans une fenêtre carrée de taille $N_p \times N_p$. Pour trouver le meilleur polygone, nous devons parcourir totalement cet espace, ce qui devient très coûteux lorsque N_s et N_p sont élevés. Nous allons donc considérer des sous-espaces Ω^i . Le processus itératif utilisé est récapitulé dans l'algorithme 2.1.

1. Lire $Poly^0$, polygone initial donné par l'utilisateur
2. Répéter
 - 2.1. Construire le sous-espace Ω^i qui contient l'ensemble des polygones dont les sommets sont dans le voisinage $N_p \times N_p$ des sommets de $Poly^i$.
Ce voisinage est relatif à la précision $prec$ choisie.
Si $prec < 1$ alors calculer les niveaux de gris de la fenêtre de recherche par interpolation des niveaux de gris
 - 2.2. Explorer Ω^i pour trouver le polygone $Poly^{i+1}$ avec l'énergie minimale $E(Poly^{i+1})$
Jusqu'à $E(Poly^{i+1}) > E(Poly^i)$
3. $Poly^i$ est le polygone retenu

ALGO. 2.1 – Exploration de l'espace des polygones.

Dans le cas général, il y a N_p^2 sommets possibles pour chaque sommet et ainsi il y a pour chaque polygone $N_p^{2N_s}$ possibilités. Si nous recalculons, à chaque étape, le gradient pour chaque segment, nous effectuons donc $N_s N_p^{2N_s}$ calculs. Afin de réduire le nombre d'interpolations du gradient calculé en chaque point d'échantillonnage, nous calculons et nous stockons en mémoire les énergies de tous les segments possibles, c'est-à-dire $N_s N_p^4$ énergies. Nous passons ainsi d'une complexité exponentielle en N_s à une complexité linéaire en N_s .

Pour obtenir une précision supérieure, après avoir effectué un calcul au pixel près, nous affinons successivement le calcul en divisant la précision par deux, c'est-à-dire que nous affinons le résultat en utilisant une précision à 0.5 pixel puis à 0.25 ... Jusqu'à obtenir la précision souhaitée.

Jusqu'à présent, nous nous sommes uniquement intéressés au cas d'un polygone. Il convient maintenant de décider comment gérer tous les polygones de l'image. Nous ne pouvons pas considérer simultanément tous les sommets de tous les polygones car l'espace des polygones serait trop grand. Nous choisissons plutôt d'estimer les polygones successivement et de fixer les positions des sommets des polygones déjà estimés. Cette démarche est critiquable car *a priori* rien ne nous assure que les positions finales des sommets soient les mêmes quel que soit l'ordre d'estimation des polygones. Mais, en pratique, cet ordre influe peu sur les positions des sommets.

Recherche par recuit simulé – Dès que le nombre de sommets des polygones est supérieur ou égal à 6 ou 7, les temps de calcul deviennent trop importants. Nous devons alors abandonner l'exploration complète des sous-espaces Ω^i et choisir une méthode d'optimisation. La manière d'explorer les sous-espaces Ω^i est donc différente de celle utilisée pour la recherche gloutonne.

Nous avons choisi la méthode du recuit simulé car la seule hypothèse nécessaire pour utiliser cette méthode est d'avoir une fonction d'énergie qui possède un minimum global.

Nous avons déjà décrit le principe du recuit simulé dans le paragraphe 1.10.3. La méthode utilisée dans ce paragraphe est donnée par l'algorithme 2.2.

1. Fixer une température initiale T^0 et générer le polygone $Poly^0$
2. Répéter
 - 2.1. Effectuer un changement aléatoire sur $Poly^i$ pour obtenir $Poly^{i+1}$ et baisser la température T^i
 - 2.2. Calculer la variation d'énergie

$$\Delta E = E(Poly^{i+1}) - E(Poly^i) \quad (2.10)$$

2.3. Si $\Delta E < 0$ alors accepter le changement

2.4. Si $\Delta E \geq 0$ alors accepter le changement avec la probabilité $e^{\left(\frac{\Delta E}{-kT_i}\right)}$

Jusqu'à ($t > N_i$) et $Poly^i$ ne change pas

ALGO. 2.2 – *Recuit simulé pour l'estimation d'un polygone* – Le terme N_i correspond au nombre maximal d'itérations et t est l'itération courante.

Les remarques suivantes peuvent être faites sur l'utilisation de cet algorithme dans ce cadre précis :

- La température initiale T^0 doit être suffisamment élevée pour qu'un grand nombre de polygones d'énergies plus basses soient acceptés au début de la simulation. Cela est nécessaire pour éviter les minima locaux.
- La fonction de décroissance de la température est $T^{i+1} = qT^i$ (nous avons pris $q = 0.9$). De même pour ne pas se retrouver dans un minimum local, il est préférable de ne baisser la température qu'après un certain nombre d'itérations ou de prendre une valeur de q proche de 1.

2.9.1.4 Résultats expérimentaux

Nous allons présenter les résultats à la fois sous forme qualitative et quantitative. Nous avons décidé de valider notre approche sur deux types d'images (images de synthèse et images réelles avec vérité terrain) et sur trois niveaux différents (nous évaluons la précision de la segmentation, du calcul des disparités et de l'estimation des homographies).

Images testées – Nous avons tout d'abord réalisé des couples stéréoscopiques de synthèse photo-réalistes : nous avons modélisé des scènes constituées de facettes polygonales planes et nous avons plaqué une image réelle sur chaque plan. Le rendu des images a été réalisé grâce au logiciel libre Blender 6¹⁸. Les cartes de disparités ont été calculées grâce à un lancer de rayon classique [Whitted 80]. Les couples que nous avons produits (« journaux » et « murs ») sont présentés sur la figure 2.8 et les résultats que nous avons obtenus sur la figure 2.9.

18. <http://www.upperfold.com/blender.shtml>

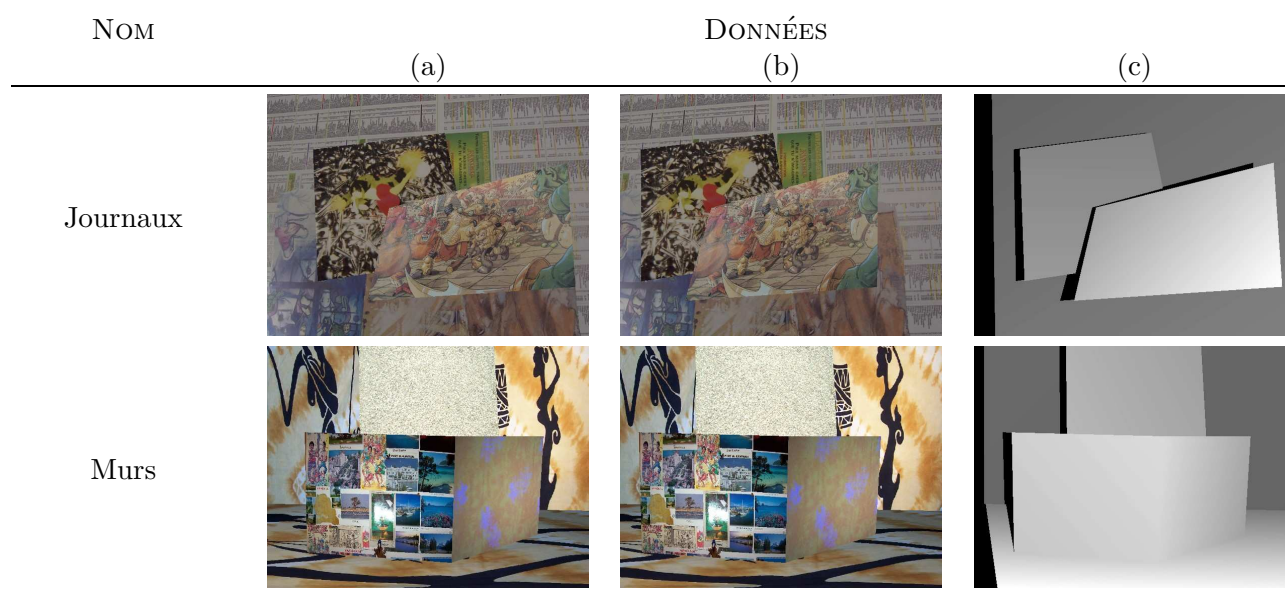


FIG. 2.8 – Couples stéréoscopiques de synthèse proposés – Les images gauche (a) et droite (b) sont présentées ainsi que la carte de disparités (c). Les pixels noirs correspondent aux pixels occultés.

Nous avons ensuite testé la technique sur sept couples d'images réelles. Nous présentons ici les résultats pour trois de ces couples, *map*, *poster* et *stairs*. Pour ces couples, les plans ont été identifiés manuellement par leurs auteurs [Šára 97, Szeliski 99b, Scharstein 02]. Nous avons segmenté grossièrement les images en plans manuellement puis nous avons calculé automatiquement les polygones. Les cartes de disparités, les cartes de différences entre les cartes de disparités théoriques (celles données par les auteurs) et les cartes de disparités calculées sont présentées sur la figure 2.10.

Critères d'évaluation – La segmentation est évaluée en mesurant sur les cartes de disparités les erreurs commises aux frontières des plans. Ces erreurs sont affichées sous la forme de cartes de différences (cf. figures 2.10 et 2.9). Nous considérons qu'il y a une erreur sur un pixel dans les deux cas suivants :

- Un pixel est détecté comme occulté dans la première carte alors qu'il est apparié dans la seconde (pixel noir dans la carte des différences).
- Un pixel est affecté à un plan dans la première carte et affecté à un autre plan dans la seconde carte (pixel gris dans la carte des différences).

Par souci de lisibilité, nous avons ajouté un cadre noir de un pixel d'épaisseur autour des cartes de différences.

Nous évaluons les disparités en calculant le pourcentage d'appariements corrects, noté COR, le pourcentage d'appariements acceptés, noté ACC, le pourcentage d'appariements erronés, noté ERR, le pourcentage de faux positifs et de faux négatifs, notés respectivement FPO et FNE, l'erreur moyenne et l'erreur maximale, notée respectivement EMO et EMA, et le pourcentage de pixels occultés correctement détectés, noté ZO (l'utilisation de ces critères sera détaillée au paragraphe 2.9.5). Les résultats pour ces différents critères sont donnés dans le tableau 2.2. Comme les erreurs sur les disparités aux frontières des plans sont déjà prises en compte dans l'évaluation de la segmentation, nous excluons les pixels des frontières (les contours et leurs 4-voisins).

Dans le but d'apprécier l'estimation des homographies, nous visualisons des iso-disparités à partir des cartes de disparités. Les iso-disparités sont affichées sous la forme de bandes colorées : dans une

même bande, les disparités sont constantes. En effet, les iso-disparités d'un polygone plan doivent être des droites parallèles entre elles (l'explication de ce phénomène est fourni dans [Bocquillon 04, p. 67]). Les homographies sont d'autant mieux estimées que les iso-disparités sont linéaires.

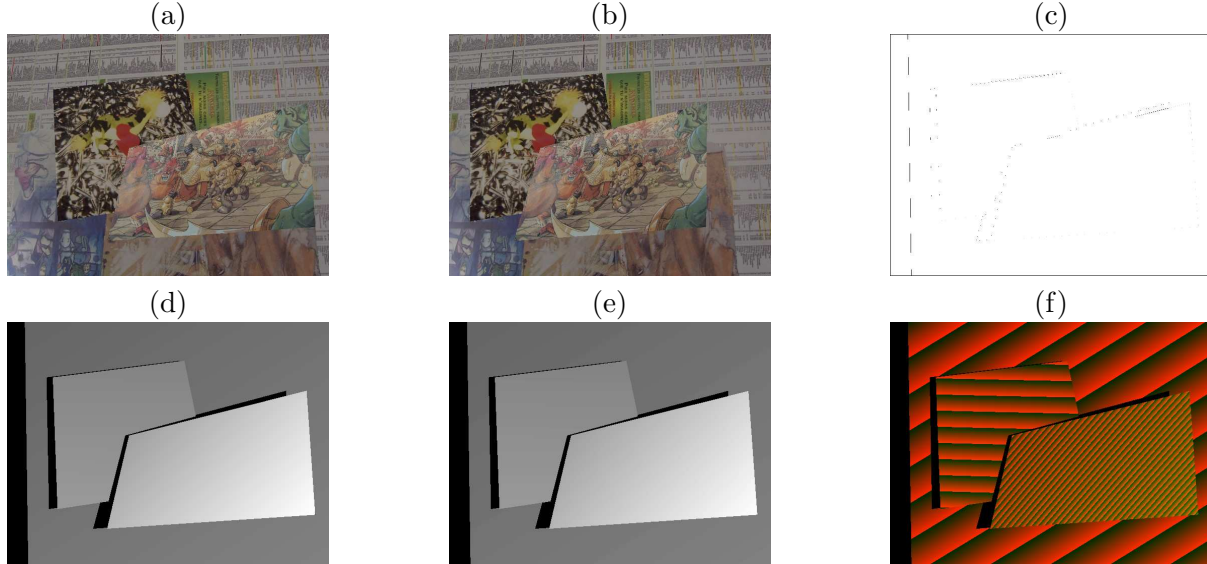


FIG. 2.9 – *Vérité terrain théorique et vérité terrain calculée pour un couple d'images de synthèse – Les cartes de disparités entre l'image gauche (a) et l'image droite (b) du couple « journaux », celle que nous avons fournie (d) et celle que nous avons calculée (e) sont comparées grâce à la carte des différences (c). Les iso-disparités pour la carte (e) sont présentées en (f). Nous ne présentons pas les résultats obtenus avec le couples « murs » car ils sont équivalents.*

NOM	COR	ACC	ERR	FPO	FNE	EMO	EMA	ZO
<i>Map</i>	99.2	0	0.27	0.35	0.18	20.98	23.5	95.11
<i>Poster</i>	98.3	0.06	0.13	0.72	0.85	2.51	7.75	82.71
<i>Stairs</i>	95.76	0.38	3.05	0.89	0.31	9.04	16.6	94.97

TAB. 2.2 – *Résultats obtenus pour les couples d'images testées – Évaluation quantitative des disparités obtenues.*

Analyse des résultats – Les erreurs au niveau de la segmentation entraînent des faux positifs, des faux négatifs et des appariements erronés près des frontières des plans. Pour tous les couples d'images réelles testés, les auteurs ont effectué une segmentation manuelle, l'erreur commise est donc de plus ou moins un pixel. Grâce aux cartes de différences que nous proposons, nous pouvons voir que la distance entre les frontières des plans de nos cartes et celles des auteurs, est d'au plus un pixel. Pour les images *map* et *poster*, les valeurs calculées des pixels sont proches de celles de Scharstein et Szeliski [Scharstein 02] qui fournissent ces cartes de disparités avec une erreur sur les disparités de l'ordre du quart de pixel. Pour les cartes que nous fournissons, nous avons choisi une précision de 0.1 pixel, nous espérons donc avoir une erreur de disparité du même ordre. Il est donc raisonnable de comparer nos résultats à ceux obtenus dans [Scharstein 02]. Pour les images *stairs*, il y a des erreurs en haut et en bas de chaque plan. Visiblement, dans [Šára 97], la totalité des plans n'ont pas été pris en compte au

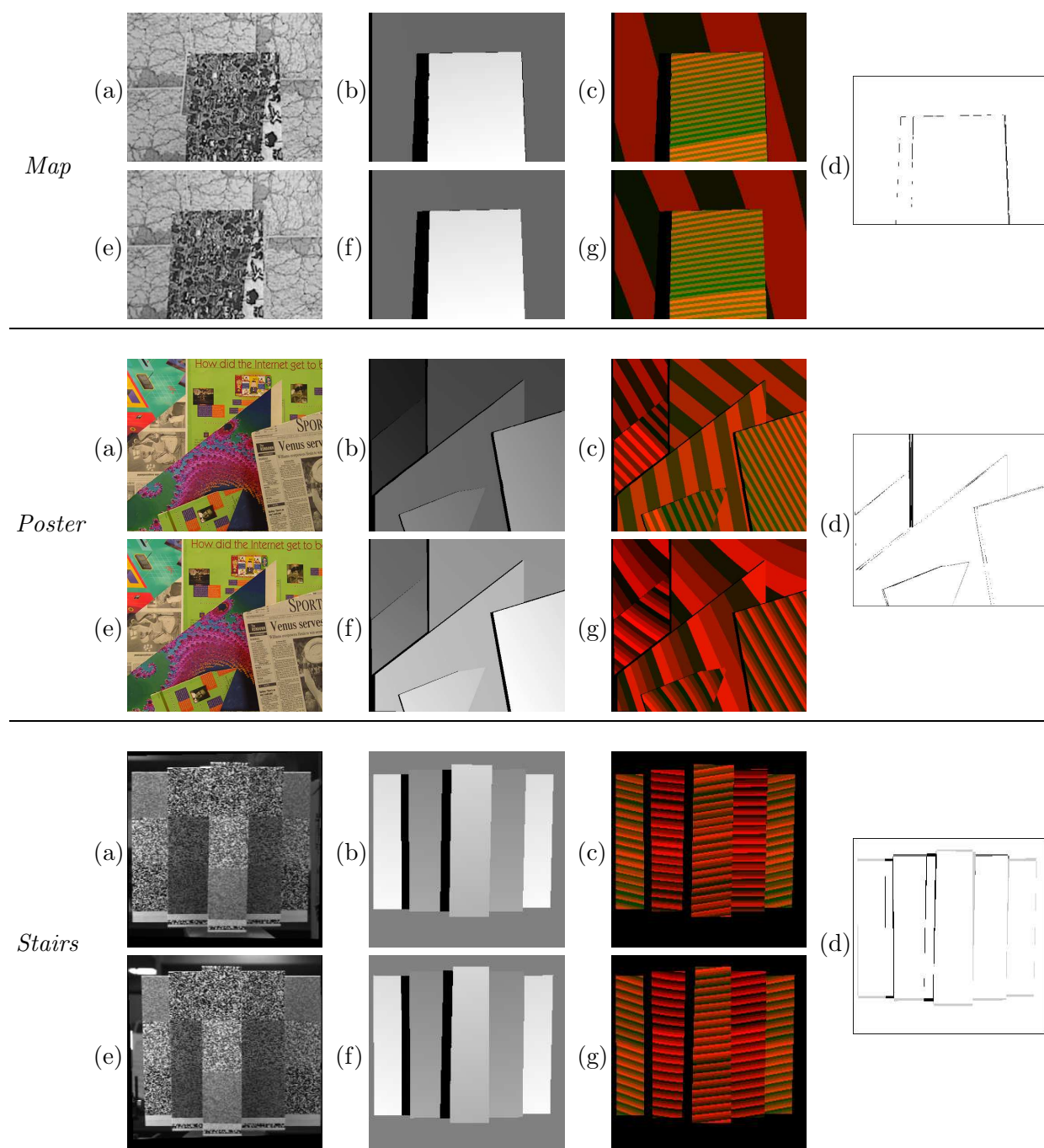


FIG. 2.10 – Vérité terrain théorique et vérité terrain calculée pour des couples d'images réelles – Les cartes de disparités entre l'image gauche (a) et l'image droite (e), celle fournie par les auteurs (b) et celle que nous avons calculée (f) sont comparées grâce à la carte des différences (d). Les iso-disparités pour les cartes (b) et (f) sont présentées respectivement en (c) et (g).

cours de la segmentation, ce qui explique ces différences. En dehors de ces zones, les résultats que nous obtenons sont proches de la carte de disparités proposée dans [Šára 97]. Si nous observons les cartes d'iso-disparités, elles sont assez proches de celles obtenues pour les cartes théoriques. Cependant, pour *map* et *poster*, les iso-disparités que nous avons calculées montrent qu'il y a quelques imperfections

dans l'estimation des homographies. Nous attribuons ces différences au fait que dans [Scharstein 02], neuf images sont utilisées alors que nous n'en utilisons que deux. Les erreurs que nous commettons restent assez faibles (de l'ordre d'un huitième de pixel). Enfin, certains paramètres ont une très grande influence sur les résultats obtenus. Le nombre de points d'intérêt doit être suffisamment grand pour obtenir une bonne estimation de l'homographie. La taille de la zone d'agrégation, la précision au sous-pixel et le seuil sur le coût local influent sur la qualité de l'appariement des points d'intérêt. Il est préférable de choisir une estimation robuste pour l'estimation des homographies.

Enfin, afin de valider notre approche sur les données de référence proposée dans [Scharstein 02], nous avons testé l'image *venus* (car c'est la seule qui satisfait nos hypothèses de départ) sur leur site. Nous avons donc classé les méthodes uniquement suivant les résultats obtenus sur cette image. Nos résultats nous classent en première position sur les trois critères évalués dans [Scharstein 02]. Ce qui confirme la fiabilité de notre estimation. Nous n'avons pas soumis définitivement ces résultats car ils ne nous semblent pas comparables aux autres méthodes. En effet, notre méthode est semi-automatique alors que les autres sont automatiques.

2.9.2 Obtention de nouvelles images avec vérité terrain

2.9.2.1 Acquisition des images

Nous avons suivi les étapes suivantes :

- choix de la scène et acquisition des images ;
- rectification épipolaire des images (évoquée au paragraphe 1.6.2).

Scène et acquisition – La scène doit respecter les contraintes suivantes :

- Elle doit être polyédrique.
- Les polygones ne doivent pas être trop petits, sous peine d'avoir trop peu de points d'intérêt détectés dans un polygone et donc d'avoir une mauvaise estimation de l'homographie associée à ce polygone.
- Les polygones doivent être suffisamment texturés, dans le but d'apparier un maximum de points d'intérêt.
- Les frontières des polygones doivent se distinguer pour que la segmentation automatique soit efficace.

Nous avons acquis les images avec un un appareil photographique numérique à visée réflex (Canon EOS 300D, 6.1Mp).

Rectification des images – Nous avons utilisé une technique qui s'appuie sur l'estimation de la matrice fondamentale. Devernay a mis à disposition des outils pour mettre en œuvre cette technique ¹⁹.

2.9.2.2 Calcul de la vérité terrain

Description de l'outil – Nous avons développé un outil muni d'une interface graphique, afin de rendre interactive la segmentation en plans des images et l'obtention de la vérité terrain. L'interface possède les fonctionnalités nécessaires pour initialiser la segmentation des images en polygones plans :

- création de sommets ;
- déplacement de sommets ;
- suppression de sommets.

19. <http://devernay.free.fr/cours/vision/ensg2003.html>

Il est possible d'agrandir une partie de l'image afin d'affiner la segmentation initiale. L'appréciation qualitative correspond à la visualisation de la segmentation sur les images, tandis que l'appréciation quantitative concerne l'évolution des énergies des polygones (affichées par l'outil pendant l'estimation de la segmentation) au cours des itérations. Lorsque la segmentation au pixel près est satisfaisante, nous lançons la segmentation au sous-pixel. Il est possible de segmenter partiellement les images, c'est-à-dire de ne traiter qu'une partie des polygones, et ainsi segmenter l'image en plusieurs fois. L'intérêt de segmenter un seul plan peut être, par exemple, la recherche des plans dont la segmentation est meilleure que pour d'autres. Nous commençons alors par estimer ces plans, afin de contraindre les autres plans.

Une fois les images segmentées, le calcul de la vérité terrain peut être fait en suivant les étapes suivantes décrites au paragraphe 2.9.1.2. À chaque étape, il est possible de revenir en arrière si le résultat n'est pas satisfaisant (nombre de points d'intérêt trop petit, nombre de pixels appariés trop faible, résidu trop élevé dans l'estimation d'une homographie, iso-disparités non linéaires, ...). Il est possible d'ajuster les paramètres pour certains plans seulement, en fixant les plans dont la vérité terrain nous satisfait. Ceci devient nécessaire lorsque les plans ont des textures très différentes.

Résultats – Nous avons utilisé cet outil pour produire la vérité terrain d'un couple d'images réelles que nous appelons « livres ». Les polygones ont été segmentés avec une précision de 0.1 pixel. La figure 2.11 nous montre ce nouveau couple stéréoscopique. La carte de disparités ne présente pas de fausses occultations et la linéarité des iso-disparités est satisfaisante.



FIG. 2.11 – *Couple stéréoscopique d'images réelles proposé, « livres »* – Les images gauche (a) et droite (b) sont présentées ainsi que la carte de disparités (c) et la carte des iso-disparités (d).

2.9.3 Images testées

Les différentes images testées dans notre protocole sont donc :

- Le stéréogramme présenté au paragraphe 2.3.1.1, dans le but d'établir une première évaluation de la méthode : si la méthode obtient de mauvais résultats sur ce type d'images, on peut en conclure qu'elle ne fonctionne pas correctement, en revanche, si elle obtient de bons résultats, il faut les affiner avec d'autres tests.
- Les images de synthèse proposées pour l'évaluation de la technique d'obtention de la vérité terrain « journaux » et « murs » (cf. figure 2.8) car les résultats obtenus avec ce type d'images sont assez proches de ceux obtenus avec les images réelles, donc elles permettent une bonne estimation du comportement général de la méthode.
- Les images réelles proposées dans [Scharstein 02] (cf. paragraphe 2.8.1) car ces images permettent une évaluation suffisamment précise avec des images réelles.

- Le couple d'images réelles proposé « livres » car ce dernier couple dont nous avons produit la vérité terrain permet aussi de valider le comportement des méthodes étudiées. Cependant, ces images possèdent peu de plans et peu d'occultations.

Les couples d'images que nous proposons ainsi que l'outil d'obtention de la vérité terrain sont mis à disposition sur le site <http://www.irit.fr/~Benoit.Bocquillon/MYCVR/research.php>. Toutes les images utilisées sont présentées en annexe, page 187.

2.9.4 Zones d'évaluation

Dans nos évaluations, nous considérons les zones suivantes :

- zone des occultations (cf. paragraphe 2.4) ;
- zone d'influence des occultations ;
- zone totale des occultations ;
- zone des discontinuités (cf. paragraphe 2.4).

Ces différentes zones sont caractérisées par les définitions suivantes (cf. figure 2.12) :

- *Les pixels occultés* – Ils ne possèdent pas de correspondant dans l'autre image :

$$O(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} \text{ est un pixel occulté} \\ 0 & \text{sinon.} \end{cases} \quad (2.11)$$

- *La zone des occultations* – Elle est notée $ZO(I_g)$ et contient tous les pixels occultés dans l'image I_g :

$$ZO(I_g) = \{\mathbf{p}_g^{i,j} \mid O(\mathbf{p}_g^{i,j}) = 1\}. \quad (2.12)$$

- *Les pixels proches des pixels occultés* – Ce sont les pixels qui se trouvent dans le voisinage, défini par rapport à la zone d'agrégation, d'un pixel occulté. Cela correspond à la dilatation morphologique de la zone des occultations en utilisant comme élément structurant la zone d'agrégation :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } O(\mathbf{p}_g^{i,j}) = 0 \text{ et } V(\mathbf{p}_g^{i,j}) > 0 \\ 0 & \text{sinon,} \end{cases} \quad (2.13)$$

$$\text{avec } V(\mathbf{p}_g^{i,j}) = \sum_{\mathbf{p}_g^{i',j'} \in ZA(\mathbf{p}_g^{i,j})} O(\mathbf{p}_g^{i',j'}). \quad (2.14)$$

- *La zone d'influence des occultations* – Elle est notée $ZI(I_g)$ et contient tous les pixels proches des pixels occultés dans l'image I_g :

$$ZI(I_g) = \{\mathbf{p}_g^{i,j} \mid PO(\mathbf{p}_g^{i,j}) = 1\}. \quad (2.15)$$

- *La zone totale des occultations* – Elle est notée $ZT(I_g)$ et elle est l'union de la zone des occultations et de la zone d'influence des occultations dans l'image I_g :

$$ZT(I_g) = ZO(I_g) \cup ZI(I_g) = \{\mathbf{p}_g^{i,j} \mid O(\mathbf{p}_g^{i,j}) = 1 \text{ ou } PO(\mathbf{p}_g^{i,j}) = 1\}. \quad (2.16)$$

- *Les pixels proches d'une discontinuité* – Ce sont les pixels qui se trouvent dans le voisinage, défini par rapport à la zone d'agrégation, d'un pixel qui possède une disparité suffisamment différente :

$$PD(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \exists \mathbf{p}_g^{i,v} \in ZA(\mathbf{p}_g^{i,j}) \mid \|d(\mathbf{p}_g^{i,v}) - d(\mathbf{p}_g^{i,j})\| > T_D \\ 0 & \text{sinon.} \end{cases} \quad (2.17)$$

$$(2.18)$$

Le seuil T_D a été déterminé de manière empirique pour chaque image.

- *La zone des discontinuités* – Elle est notée $ZD(I_g)$ et contient tous les pixels proches d'une discontinuité dans l'image I_g :

$$ZD(I_g) = \{ \mathbf{p}_g^{i,j} \mid PD(\mathbf{p}_g^{i,j}) = 1 \}. \quad (2.19)$$

Ainsi, nous distinguons les zones proches des occultations et les zones des discontinuités. En effet, nous considérons le cas où il y a disparition de pixels (qui s'accompagne toujours d'une discontinuité de profondeur) et le cas où il y a une discontinuité de profondeur sans disparition de pixels.

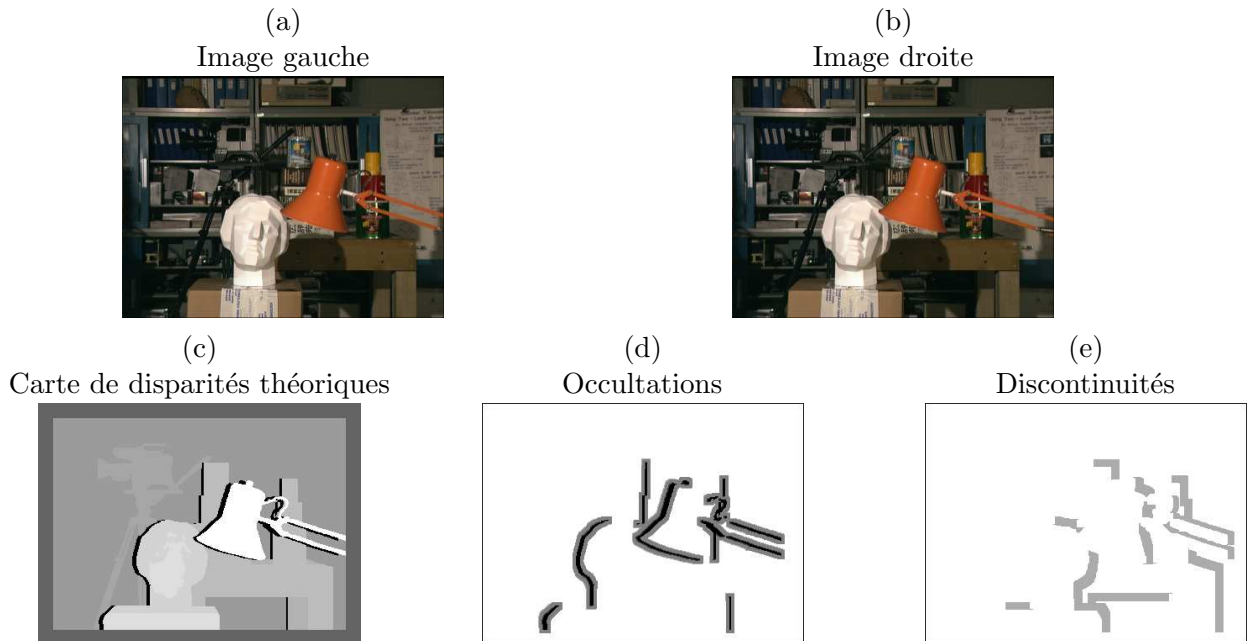


FIG. 2.12 – Zones des occultations et zone des discontinuités considérées – Dans cet exemple, les images gauche et droite sont données en (a) et (b). Nous avons calculé la zone totale des occultations, ZT, en (d), grâce à la vérité terrain fournie dans [Scharstein 02]. En (e), nous représentons la zone des discontinuités (il n'y a pas d'intersection entre cette zone et les différentes zones des occultations).

2.9.5 Critères d'évaluation

Nous donnons pour chaque résultat, la carte de disparités correspondante. Les critères évalués, dans le cadre de tous nos tests, sont les suivants (la description de ces critères est donnée dans le paragraphe 2.5.2) :

1. pourcentage d'appariements corrects, noté COR ;
2. pourcentage d'appariements acceptés, noté ACC ;
3. pourcentage d'appariements mauvais, noté MAU ;
4. pourcentage d'appariements erronés, noté ERR ;
5. pourcentage de faux positifs, noté FPO ;
6. pourcentage de faux négatifs, noté FNE.

Le critère 1 est évalué sur chaque type de zones que nous notons respectivement ZI, ZO, ZT et ZD pour zone d'influence, zone des occultations, zone totale des occultations et zone des discontinuités.

Afin d'obtenir un résultat plus visuel, nous fournissons aussi des cartes d'appariements corrects : Chaque pixel de la carte indique si l'appariement est correct ou non, c'est-à-dire :

- Si le pixel est blanc alors l'appariement est incorrect.

- Si le pixel est noir alors c'est un vrai négatif.
- Si le pixel est gris alors l'appariement est correct.

2.9.6 Méthodes évaluées

Nous allons évaluer des méthodes à base de corrélation qui utilisent des images en niveau de gris, puis des méthodes qui prennent en compte des images couleur et enfin des méthodes qui tentent de prendre en compte les occultations. Ainsi, les méthodes évaluées seront précisées dans chaque chapitre.

2.9.7 Synthèse des critères et classement des méthodes

Pour effectuer le classement des méthodes, nous attribuons un rang pour chaque critère, puis nous effectuons le classement des méthodes en fonction des moyennes de ces rangs. Par la suite, dans les tableaux de résultats, nous utilisons la notation RT pour indiquer le rang de chaque méthode parmi toutes les méthodes évaluées. Comme dans [Scharstein 02], nous classons les méthodes suivant la moyenne des rangs sur chaque critère.

2.10 Conclusion

Dans ce chapitre, nous avons établi, dans un premier temps, un état de l'art sur les protocoles d'évaluation des méthodes de mise en correspondance. En particulier, nous avons identifié les éléments suivants :

- le type d'images utilisées : images de synthèse, réelles avec ou sans vérité terrain ;
- les critères évalués ;
- le choix des méthodes comparées ;
- la manière de synthétiser les résultats et de classer les méthodes.

En conclusion de cet état de l'art, il s'avère que le protocole d'évaluation le plus complet est celui de Scharstein et Szeliski [Scharstein 02]. Ce protocole a l'avantage de proposer des tests avec vérité terrain et il a été utilisé pour comparer quarante algorithmes. Les auteurs ne fournissent cependant pas d'outil permettant d'obtenir la vérité terrain pour d'autres images et la masse importante de résultats obtenus n'est pas analysée.

Le dernier aspect abordé dans ce chapitre concerne l'élaboration de notre propre protocole d'évaluation. Pour cela, nous avons décidé de construire de nouvelles données de référence, ce qui a donné naissance à un outil disponible sur le web²⁰ et permettant de produire la vérité terrain pour des images de scènes polyédriques. Ensuite, nous avons défini les différentes zones de l'image que nous considérons et les critères que nous souhaitons évaluer. La synthèse des résultats s'effectue en moyennant le rang de chaque méthode pour chaque critère.

20. <http://www.irit.fr/~Benoit.Bocquillon/MYCVR/research.php>

Chapitre 3

Mise en correspondance par corrélation des niveaux de gris – Corrélation 1D

3.1 Introduction

Dans le cadre de nos travaux, nous nous intéressons particulièrement aux méthodes de mise en correspondance par corrélation des niveaux de gris. Nous utilisons aussi le terme corrélation 1D, car il n’y a qu’une seule composante à considérer : le niveau de gris, par opposition à la corrélation 3D où il y a trois composantes à considérer : la couleur (nous verrons la corrélation 3D dans le chapitre 4). Dans tout ce chapitre, pour faciliter la lecture, nous parlerons simplement de corrélation pour désigner la corrélation 1D.

Comme nous avons pu le voir dans les chapitres 1 et 2, les méthodes locales ou globales à base de corrélation sont très populaires, cf. tableaux 1.3 et 2.1. Notons aussi que ces méthodes de mise en correspondance par corrélation servent, en vision par ordinateur, pour :

- le recalage d’images [Gottesfeld Brown 92] ;
- la reconnaissance d’objets [Puzicha 97, Kawanishi 04] ;
- l’estimation du mouvement [Mulligan 00, Reddy 04] ;
- l’étude de la déformation 3D des objets [Garcia 01b] ;
- le suivi d’objets [Garcia 02, Lantagne 03, Matthews 03, Stock 03] ;
- l’indexation vidéo [Porter 03] ;
- l’interprétation de scènes urbaines [Chehata 05].

Ici, nous allons uniquement nous intéresser aux méthodes locales à base de corrélation.

3.1.1 Avantages et inconvénients

Les méthodes à base de corrélation possèdent les avantages suivants :

- Leur mise en œuvre est simple et rapide.
- Elles ne sont pas très coûteuses en temps de calcul (elles sont moins coûteuses que les méthodes globales).
- Dans de nombreux travaux, elles ont déjà prouvé leur efficacité : dans l’évaluation de Scharstein et Szeliski [Scharstein 02], la méthode proposée par [Yoon 05] se situe dans la première moitié du classement (cf. tableau 2.1).

Les difficultés qui peuvent être liées à ces méthodes sont les suivantes :

- Elles ne prennent pas toujours en compte certaines situations qui rendent difficile la mise en correspondance : les bruits, les raccourcissements, les zones non texturées, les textures répétitives,

les changements de luminosité et les occultations.

- Elles ne fonctionnent pas dans le cas de déformation des objets ou de changement d'échelle (nous parlons ici des méthodes classiques qui n'utilisent pas des zones d'agrégation qui s'adaptent à la forme des objets).
- Elles n'autorisent pas des points de vue trop éloignés.
- Comme elles sont locales, les résultats peuvent être erronés car certaines caractéristiques globales de l'image ne sont pas prises en compte. Par exemple, certaines méthodes globales segmentent l'image en régions qui correspondent à des projections d'un même objet de la scène. Cette segmentation permet de réduire l'espace des disparités envisageables, et donc de réduire les erreurs.

3.1.2 Objectifs du chapitre

De nombreuses mesures de corrélation ont été proposées pour prendre directement en compte les difficultés évoquées ci-dessus. Cela a donné lieu à un nombre conséquent de nouvelles mesures proposées durant ces deux dernières décennies. Aussi, il est difficile de choisir la mesure la plus adaptée au contexte dans lequel on se situe. Il existe des travaux dans lesquels une taxonomie des mesures de corrélation a été effectuée, la plus célèbre étant celle d'Aschwanden et Guggenbül [Aschwanden 92]. Des taxonomies plus partielles ont aussi été proposées, comme dans [Zabih 94, Bhat 98, Giachetti 00, Cyganek 01] avec des mesures ordinales, dans [Zhang 95, Lan 97] avec des mesures robustes et pour des applications autres que la mise en correspondance, par exemple dans [Gottesfeld Brown 92] pour le recalage d'images et dans [Rubner 01] pour la classification. La taxonomie d'Aschwanden et Guggenbül [Aschwanden 92] regroupe une vingtaine de mesures. Depuis 1992, d'autres mesures ont été proposées dans la littérature. Il nous est donc apparu important d'établir une nouvelle taxonomie englobant et complétant ces travaux. De plus, comme nous nous intéressons particulièrement aux problèmes induits par les occultations, nous proposons de nouvelles mesures qui prennent en compte ces difficultés.

3.1.3 Plan

Après avoir défini les notations que nous allons utiliser et la notion de mise en correspondance par corrélation, nous allons présenter et caractériser la plupart des mesures de corrélation existantes. Pour cela nous proposons une classification en cinq familles :

- la famille CROISÉE qui regroupe les mesures utilisant la corrélation croisée ;
- la famille CLASSIQUE qui regroupe les mesures utilisant des outils des statistiques classiques de la distribution des différences des niveaux de gris ;
- la famille DÉRIVÉE qui regroupe les mesures utilisant les dérivées des niveaux de gris ;
- la famille NON PARAMÉTRIQUE qui regroupe les mesures utilisant des outils des statistiques non paramétriques ;
- la famille ROBUSTE qui regroupe les mesures utilisant des outils des statistiques robustes.

Les nouvelles mesures que nous proposons font partie de la dernière famille.

3.2 Présentation de la mise en correspondance par corrélation 1D

3.2.1 Notations

Les notations suivantes sont utilisées :

- La taille des fenêtres de corrélation est $(2N_v + 1) \times (2N_h + 1)$ et le nombre de pixels dans la fenêtre de corrélation est noté $N_f = (2N_v + 1)(2N_h + 1)$.

- Les niveaux de gris des pixels des fenêtres de corrélation (matrices) sont stockés dans les vecteurs \mathbf{f}_l :

$$\mathbf{f}_l = (\dots I_l^{i+p,j+q} \dots)^T = (\dots f_l^k \dots)^T$$

où f_l^k est l'élément k du vecteur \mathbf{f}_l , $p \in [-N_v; N_v]$, $q \in [-N_h; N_h]$ et $k \in [0; N_f - 1]$.

- Le niveau de gris le plus grand de l'image est I_{\max} .
- Le vecteur gradient au pixel $(i, j)^T$ de l'image I_l est représenté par $\nabla I_l^{i,j}$. La norme et l'orientation du gradient sont notées respectivement $\|\nabla I_l^{i,j}\|$ et $\theta_l^{i,j}$.
- La partie entière de x est $\lfloor x \rfloor$.
- Les P-normes, ou distances L_P , $P \in \mathbb{N}^*$, sont notées :

$$\|\mathbf{f}_l\|_P = \left(\sum_{k=0}^{N_f-1} |f_l^k|^P \right)^{1/P}. \quad (3.1)$$

Pour la norme euclidienne, nous notons : $\|\mathbf{f}_l\| = \|\mathbf{f}_l\|_2$.

- Le produit scalaire est donné par :

$$\mathbf{f}_g \cdot \mathbf{f}_d = \sum_{k=0}^{N_f-1} f_g^k f_d^k. \quad (3.2)$$

- Les moyennes sont notées :

$$\text{moy}(\mathbf{f}_l) = \frac{1}{N_f} \sum_{k=0}^{N_f-1} f_l^k. \quad (3.3)$$

Nous utilisons aussi le vecteur des moyennes défini par :

$$\bar{\mathbf{f}}_l = \underbrace{(\text{moy}(\mathbf{f}_l) \dots \text{moy}(\mathbf{f}_l))^T}_{N_f \text{ colonnes}}. \quad (3.4)$$

- Les variances sont notées :

$$\text{var}(\mathbf{f}_l) = \overline{(\mathbf{f}_l - \bar{\mathbf{f}}_l)^2} = \bar{\mathbf{f}}_l^2 - (\bar{\mathbf{f}}_l)^2. \quad (3.5)$$

- La concaténation est notée \otimes .
- La fonction sgn est définie par :

$$\text{sgn}(x) = \begin{cases} -1 & \text{si } x < 0 \\ 0 & \text{si } x = 0 \\ 1 & \text{sinon.} \end{cases} \quad (3.6)$$

- La distance de Hamming est donnée par :

$$D_{\text{Ham}}(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} \text{sgn} |f_g^k - f_d^k|. \quad (3.7)$$

- Les valeurs ordonnées du vecteur \mathbf{f}_l sont notées :

$$(f_l)_{0:N_f-1} \leq \dots \leq (f_l)_{k:N_f-1} \leq \dots \leq (f_l)_{N_f-1:N_f-1}. \quad (3.8)$$

3.2.2 Définitions

La définition d'une mesure de corrélation que nous adoptons dans ce mémoire est la suivante :

Il s'agit d'une mesure qui évalue à quel point deux ensembles de données se ressemblent.

Par abus de langage, nous employons le terme « mesure de corrélation » aussi bien pour une mesure qui évalue le degré de similarité que pour une mesure qui évalue le degré de dissimilarité. Toutefois, pour présenter les différentes mesures de corrélation existantes, dans le paragraphe 3.3, nous préciserons toujours s'il s'agit d'une mesure de similarité ou de dissimilarité.

Le but de la mise en correspondance est de trouver pour chaque pixel $\mathbf{p}_g^{i,j}$ de l'image gauche, le pixel \mathbf{p}_d^{i,v^*} qui lui correspond dans l'image droite. La mise en correspondance par corrélation repose sur l'hypothèse suivante :

Les niveaux de gris des pixels correspondants et des pixels de leurs voisinages se ressemblent.

Les mesures de corrélation utilisent donc l'information donnée par les pixels, mais aussi celle qui est fournie par leurs voisinages respectifs. Ainsi, le score de corrélation, qui est la valeur de la mesure de corrélation, est évalué à partir de deux ensembles : le pixel de l'image gauche et son voisinage, représentés par le vecteur \mathbf{f}_g , et le pixel de l'image droite et son voisinage, représentés par le vecteur \mathbf{f}_d . Notons que la zone d'agrégation que nous avons définie au chapitre 1 correspond à ces voisinages, c'est à dire que \mathbf{f}_i contient les niveaux de gris des pixels de la zone d'agrégation $ZA(\mathbf{p}_i^{i,j})$.

Tout d'abord, il convient de déterminer (cf. figures 3.1 et 3.2) :

- la taille de la fenêtre de corrélation ;
- le sens de mise en correspondance : de l'image gauche vers l'image droite (c'est le sens choisi dans l'algorithme 3.1) ou de l'image droite vers l'image gauche ;
- la mesure à utiliser, notée Mes.

Puis, pour chaque pixel $\mathbf{p}_g^{i,j}$, il faut appliquer l'algorithme 3.1.

Les algorithmes à base de mesure de corrélation, sont souvent appelés les algorithmes « de région », *block matching* ou *area-based matching*, comme dans [Jawahar 02].

1. Déterminer la zone de recherche, *region of interest* (zone dans laquelle on pense trouver le correspondant), notée $Z_d(\mathbf{p}_g^{i,j})$, et donnée par

$$Z_d(\mathbf{p}_g^{i,j}) = \left\{ \mathbf{p}_d^{i,v} \mid j + dc_{\min} \leq v \leq j + dc_{\max} \right\} \quad (3.9)$$

où les termes dc_{\min} et dc_{\max} représentent la plus petite et la plus grande composante de la disparité selon les colonnes ($N_z = dc_{\max} - dc_{\min}$)

2. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire calculer $\text{Mes}(\mathbf{f}_g, \mathbf{f}_d)$
3. Sélectionner le pixel \mathbf{p}_d^{i,v^*} qui donne le meilleur score de corrélation

$$\mathbf{p}_d^{i,v^*} = \underset{\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})}{\text{argmin}} \text{Mes}(\mathbf{f}_g, \mathbf{f}_d) \quad (3.10)$$

ALGO. 3.1 – *Mise en correspondance par corrélation 1D (niveaux de gris) – Nous nous plaçons dans le cas des images rectifiées. Cet algorithme correspond à une recherche exhaustive dans la zone de recherche. Le terme N_z est le nombre de pixels considérés dans la zone de recherche. Les vecteurs \mathbf{f}_g et \mathbf{f}_d sont les vecteurs représentant les deux ensembles de pixels comparés.*

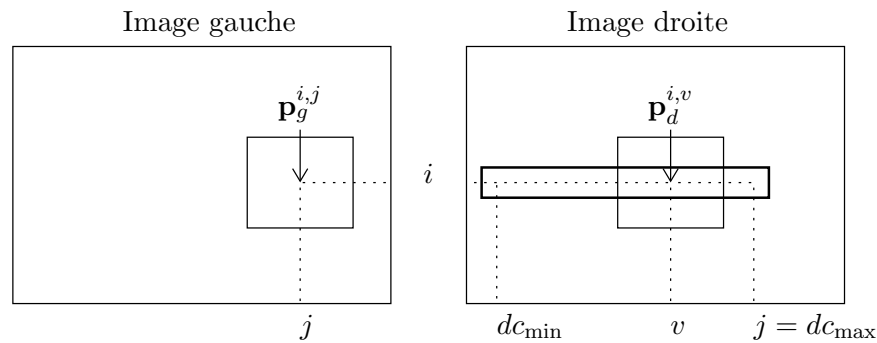


FIG. 3.1 – Zone de recherche et fenêtres de corrélation – Les rectangles correspondent aux fenêtres de corrélation, \mathbf{f}_g et \mathbf{f}_d , et le rectangle en gras correspond à la zone de recherche, $Z_d(\mathbf{p}_g^{i,j})$. Nous nous plaçons dans le cas des images rectifiées. Ainsi, cette zone de recherche est unidimensionnelle. Si les images n'étaient pas rectifiées alors la zone de recherche serait bidimensionnelle.

3.2.3 Difficultés

La principale limite de la mise en correspondance par corrélation provient de l'hypothèse de départ. En utilisant les pixels du voisinage du pixel considéré pour déterminer le score de corrélation, on suppose que ces derniers possèdent tous la même disparité. Dans la plupart des cas, cette hypothèse est quasiment respectée, mais elle ne l'est plus dans les cas suivants :

- présence de bruits ;
- changements de luminosité ;
- déformations (dus à la projection perspective) des objets de la scène ;
- occultations.

Dans le cas d'images bruitées, des méthodes, cf. [Kim 05], ont été proposées avec un préfiltrage des images pour tenter d'éliminer les bruits.

Pour résoudre le problème des changements de luminosité, des mesures centrées et normalisées sont présentées aux paragraphes 3.3.1 et 3.3.2.

Pour les déformations dues à la projection perspective, une approche de corrélation fine introduite dans [Devernay 97, chap. 3] et améliorée dans [Garcia 01a, chap. 2 et 3] a été proposée. Le principe de cette méthode est d'adapter localement la fenêtre de corrélation à la courbure et à l'orientation locale de la surface de l'objet considéré. Ainsi, en plus d'estimer la disparité associée à chaque pixel, il faut estimer l'ensemble des paramètres de la transformation locale associée à ce pixel. Alors que la méthode de Devernay [Devernay 97, chap. 3] s'appuie sur des images rectifiées, la méthode de Garcia [Garcia 01a, chap. 2 et 3] peut être utilisée avec des images non rectifiées.

Dans le cas des occultations, l'hypothèse de la mise en correspondance par corrélation n'est plus vérifiée à la frontière des objets occultants. En effet, dans le voisinage d'un pixel situé à la frontière d'un changement de profondeur, certains pixels appartiennent au premier niveau de profondeur et d'autres au deuxième niveau. Cette situation va perturber la mise en correspondance par corrélation, car d'une image à l'autre, les pixels qui se correspondent n'auront pas les mêmes niveaux de gris dans leur voisinage et leur score de corrélation ne sera pas maximal.

Plusieurs solutions ont été envisagées pour prendre en compte ce problème. Les deux solutions principales sont :

- *Les fenêtres adaptatives* – La forme de la fenêtre n'est pas fixe. Elle s'adapte en fonction de la zone de l'image qui est parcourue. La méthode de corrélation fine peut être considérée comme une méthode de fenêtres adaptatives.

- *Les mesures robustes aux occultations* – Ces mesures intègrent directement le problème des occultations dans le calcul du score de corrélation. Elles correspondent à la dernière famille que nous présentons et aux nouvelles mesures que nous proposons.

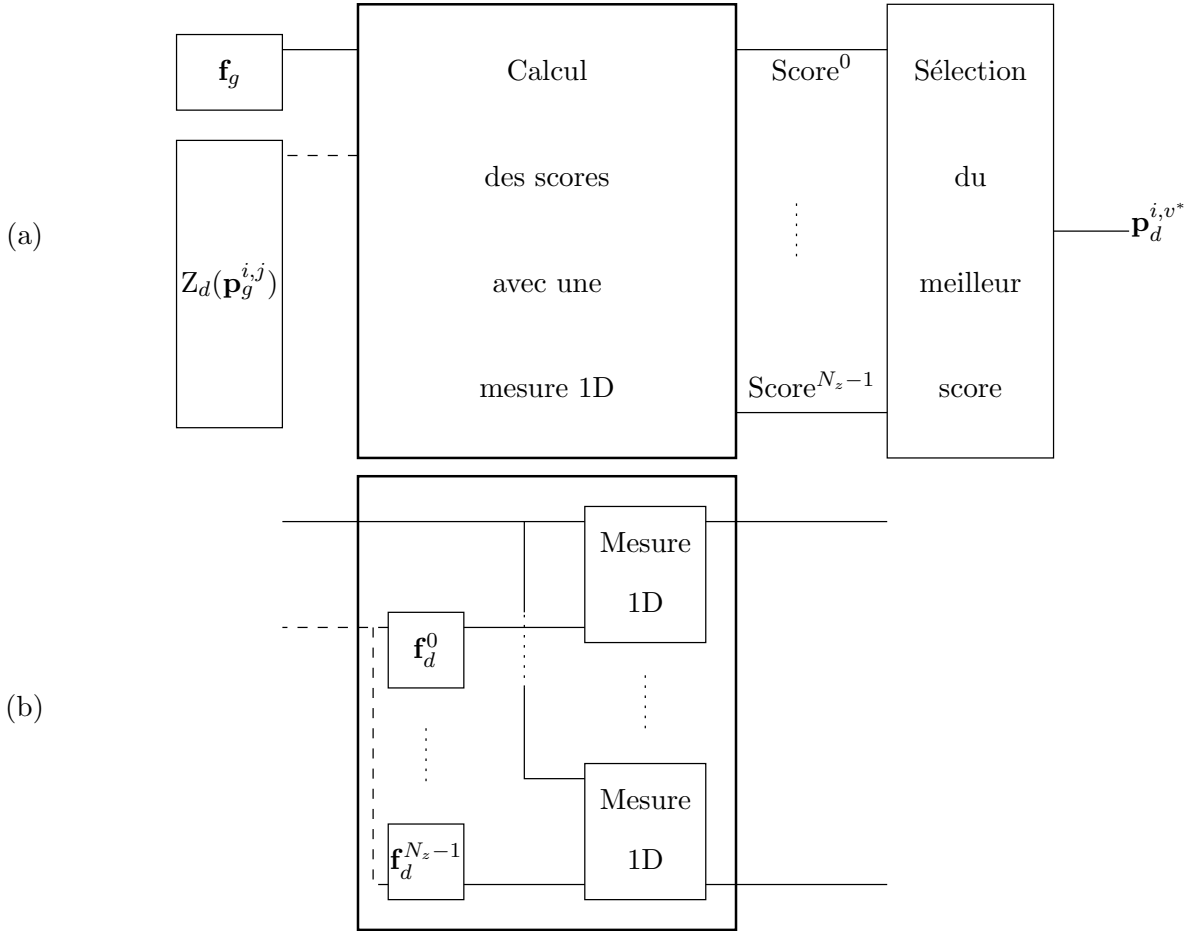


FIG. 3.2 – *Mise en correspondance par corrélation* – (a) Recherche d'un correspondant \mathbf{p}_d^{i,v^*} en utilisant une mesure 1D. (b) Détails du calcul des scores avec une mesure 1D. Il y a N_z candidats dont les fenêtres de corrélation associées sont représentées par les vecteurs \mathbf{f}_d^m avec $m = 0..N_z - 1$.

3.2.4 Propriétés

En exposant les mesures de corrélation, nous en présentons les caractéristiques :

- l'intervalle de variation des mesures formé d'un minorant et d'un majorant ;
- le type de la mesure : similarité ou dissimilarité ;
- les invariances par rapport aux transformations des niveaux de gris, c'est-à-dire par rapport aux changements de luminosité. Nous considérons trois types d'invariance, en posant $a, b, c, d \in \mathbb{R}^*$:
 - *Invariance de type gain, notée G* :

$$\text{Mes}(a\mathbf{f}_g, b\mathbf{f}_d) = \text{Mes}(\mathbf{f}_g, \mathbf{f}_d). \quad (3.11)$$

- *Invariance de type biais, notée B* :

$$\text{Mes}(\mathbf{f}_g + a, \mathbf{f}_d + b) = \text{Mes}(\mathbf{f}_g, \mathbf{f}_d). \quad (3.12)$$

◦ *Invariance de type gain et biais, notée GB :*

$$\text{Mes}(a\mathbf{f}_g + b, c\mathbf{f}_d + d) = \text{Mes}(\mathbf{f}_g, \mathbf{f}_d). \quad (3.13)$$

Cette invariance induit aussi les invariances définies par les équations (3.11) et (3.12) .

3.3 Taxonomie des mesures 1D

3.3.1 Corrélation croisée

Cette première famille (cf. tableau 3.3), notée famille CROISÉE, regroupe trois mesures qui utilisent la corrélation croisée, notée CC (*Cross-Correlation* ou produit de corrélation), donnée par :

$$\text{CC}(\mathbf{f}_g, \mathbf{f}_d) = \mathbf{f}_g \cdot \mathbf{f}_d. \quad (3.14)$$

Les valeurs de $\text{CC}(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; +\infty[$. La mesure CC ne peut pas être utilisée directement car plus les niveaux de gris sont élevés, plus les scores sont élevés. Cette mesure est exploitée sous une forme normalisée. Toutes les mesures qui utilisent la corrélation croisée sont des mesures de similarité.

3.3.1.1 Normalisation

La mesure de corrélation croisée normalisée, notée NCC (*Normalized Cross-Correlation*), est donnée par :

$$\text{NCC}(\mathbf{f}_g, \mathbf{f}_d) = \frac{\mathbf{f}_g \cdot \mathbf{f}_d}{\|\mathbf{f}_g\| \|\mathbf{f}_d\|}. \quad (3.15)$$

Les valeurs de $\text{NCC}(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; 1]$. Cette mesure est très utilisée, notamment dans [Ferrari 03, Di Stefano 03, Tsai 03], et elle possède une invariance de type gain.

3.3.1.2 Centrage et normalisation

La mesure de corrélation croisée centrée normalisée, notée ZNCC (*Zero mean Normalized Cross-Correlation*), est donnée par :

$$\text{ZNCC}(\mathbf{f}_g, \mathbf{f}_d) = \frac{(\mathbf{f}_g - \overline{\mathbf{f}_g}) \cdot (\mathbf{f}_d - \overline{\mathbf{f}_d})}{\|\mathbf{f}_g - \overline{\mathbf{f}_g}\| \|\mathbf{f}_d - \overline{\mathbf{f}_d}\|}. \quad (3.16)$$

Les valeurs de $\text{ZNCC}(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[-1; 1]$. Cette mesure correspond au coefficient de corrélation linéaire classique en statistiques. Cette mesure est l'une des plus utilisées, notamment dans [Sun 97, Zhang 00, Lhuillier 04]. Elle a l'avantage de présenter une invariance de type gain et biais. Une version robuste de cette mesure a été proposée par Trujillo et Izquierdo [Trujillo 04] (cf. paragraphe 3.3.5.5).

3.3.1.3 Corrélation de Moravec

La mesure de Moravec [Moravec 80], notée MOR, utilise une normalisation différente de celle utilisée par la mesure ZNCC :

$$\text{MOR}(\mathbf{f}_g, \mathbf{f}_d) = \frac{2(\mathbf{f}_g - \overline{\mathbf{f}_g}) \cdot (\mathbf{f}_d - \overline{\mathbf{f}_d})}{\|\mathbf{f}_g - \overline{\mathbf{f}_g}\|^2 + \|\mathbf{f}_d - \overline{\mathbf{f}_d}\|^2}. \quad (3.17)$$

Les valeurs de $\text{MOR}(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[-1; 1]$. Elle possède une invariance de type biais. D'après Aschwanden et Guggenbül [Aschwanden 92], la mesure de Moravec est robuste face à un bruit impulsif.

3.3.2 Statistiques classiques de la distribution des différences de niveaux de gris

Cette seconde famille (cf. tableau 3.4), notée famille CLASSIQUE, regroupe des mesures utilisant la distribution des différences des niveaux de gris des deux fenêtres. Elles se divisent en quatre sous-familles : les distances, les distances localement centrées (mesures proposées dans [Aschwanden 92]), les variances (mesures introduites dans [Cox 95]) et le *kurtosis* (mesure proposée dans [Rziza 01]).

3.3.2.1 Distances

Les normes L_P sont très souvent utilisées. Ces mesures de corrélation, notées D_P (*Distances*), sont données par :

$$D_P(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{f}_g - \mathbf{f}_d\|_P^P \text{ avec } P \in \mathbb{N}^*. \quad (3.18)$$

Ce sont des mesures de dissimilarité et leurs valeurs appartiennent à l'intervalle $[0; I_{\max}^P N_f]$. Les mesures les plus employées sont :

- *Somme des valeurs absolues des différences* – Il s'agit de la norme L_1 , notée SAD (*Sum of Absolute Differences*) :

$$\text{SAD}(\mathbf{f}_g, \mathbf{f}_d) = D_1(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{f}_g - \mathbf{f}_d\|_1. \quad (3.19)$$

Cette mesure est aussi une des plus populaires et elle est utilisée dans [Arcara 00, Di Stefano 02, Mayer 03, Kim 05].

- *Somme des carrés des différences* – Il s'agit de la norme L_2 , notée SSD (*Sum of Squared Differences*) :

$$\text{SSD}(\mathbf{f}_g, \mathbf{f}_d) = D_2(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{f}_g - \mathbf{f}_d\|^2. \quad (3.20)$$

Cette mesure est aussi très utilisée, cf. [Okutomi 93, Kanade 96, Manduchi 99, Ku 01].

- *Distance de Kolmogorov-Smirnov* – Dans [Puzicha 97, Smith 98, Rubner 01], les auteurs ont aussi exploité la distance D_∞ , appelée la distance de Kolmogorov-Smirnov. Elle correspond à la norme L_∞ et elle est donnée par :

$$D_\infty(\mathbf{f}_g, \mathbf{f}_d) = \max_{k=0..N_f-1} |f_g^k - f_d^k|. \quad (3.21)$$

C'est une mesure de dissimilarité et son intervalle de variation est $[0; I_{\max}]$.

Centrage – Ces mesures peuvent aussi être utilisées sous leurs versions centrées. Nous les appelons distances centrées, nous les notons ZD_P (*Zero mean Distances*) et elles sont données par :

$$ZD_P(\mathbf{f}_g, \mathbf{f}_d) = D_P(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d) = \|(\mathbf{f}_g - \bar{\mathbf{f}}_g) - (\mathbf{f}_d - \bar{\mathbf{f}}_d)\|_P^P \text{ avec } P \in \mathbb{N}^*. \quad (3.22)$$

Elles présentent une invariance de type biais. Ce sont des mesures de dissimilarité et leurs valeurs appartiennent à l'intervalle $[0; I_{\max}^P N_f]$. Les mesures les plus utilisées sont :

- *Somme des valeurs absolues des différences centrée* – Cette mesure, notée ZSAD (*Zero mean Sum of Absolute Differences*), est donnée par :

$$\text{ZSAD}(\mathbf{f}_g, \mathbf{f}_d) = ZD_1(\mathbf{f}_g, \mathbf{f}_d) = D_1(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d) = \|(\mathbf{f}_g - \bar{\mathbf{f}}_g) - (\mathbf{f}_d - \bar{\mathbf{f}}_d)\|_1. \quad (3.23)$$

- *Somme des carrés des différences centrée* – Cette mesure, notée ZSSD (*Zero mean Sum of Squared Differences*), est donnée par :

$$\text{ZSSD}(\mathbf{f}_g, \mathbf{f}_d) = ZD_2(\mathbf{f}_g, \mathbf{f}_d) = D_2(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d) = \|(\mathbf{f}_g - \bar{\mathbf{f}}_g) - (\mathbf{f}_d - \bar{\mathbf{f}}_d)\|^2. \quad (3.24)$$

Normalisation – Ces mesures peuvent aussi être utilisées sous leurs versions normalisées. Nous les appelons distances normalisées, nous les notons ND_P (*Normalised Distances*) et elles sont données par :

$$ND_P(\mathbf{f}_g, \mathbf{f}_d) = \frac{D_P(\mathbf{f}_g - \mathbf{f}_d)}{\sqrt{\|\mathbf{f}_g\|_P^P \|\mathbf{f}_d\|_P^P}} = \frac{\|\mathbf{f}_g - \mathbf{f}_d\|_P^P}{\sqrt{\|\mathbf{f}_g\|_P^P \|\mathbf{f}_d\|_P^P}} \text{ avec } P \in \mathbb{N}^*. \quad (3.25)$$

Ce sont des mesures de dissimilarité et leurs valeurs appartiennent à l'intervalle $[0; I_{\max}^P N_f]$. La mesure suivante est la plus employée :

- *Somme des carrés des différences normalisée* – Cette mesure, notée NSSD (*Normalised Sum of Squared Differences*), est donnée par :

$$NSSD(\mathbf{f}_g, \mathbf{f}_d) = ND_2(\mathbf{f}_g, \mathbf{f}_d) = \frac{D_2(\mathbf{f}_g - \mathbf{f}_d)}{\sqrt{\|\mathbf{f}_g\| \|\mathbf{f}_d\|}} = \frac{\|\mathbf{f}_g - \mathbf{f}_d\|^2}{\sqrt{\|\mathbf{f}_g\| \|\mathbf{f}_d\|}}. \quad (3.26)$$

Centrage et normalisation – Ces mesures peuvent aussi être utilisées sous leurs versions centrées et normalisées. Nous les appelons distances centrées normalisées, nous notons ZND_P (*Zero mean Normalised Distances*) et elles sont données par :

$$ZND_P(\mathbf{f}_g, \mathbf{f}_d) = ND_P(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d) = \frac{\|(\mathbf{f}_g - \bar{\mathbf{f}}_g) - (\mathbf{f}_d - \bar{\mathbf{f}}_d)\|_P^P}{\sqrt{\|\mathbf{f}_g - \bar{\mathbf{f}}_g\|_P^P \|\mathbf{f}_d - \bar{\mathbf{f}}_d\|_P^P}} \text{ avec } P \in \mathbb{N}^*. \quad (3.27)$$

Elles possèdent une invariance de type biais. Ce sont des mesures de dissimilarité et leurs valeurs appartiennent à l'intervalle $[0; I_{\max}^P N_f]$. La mesure suivante est la plus employée :

- *Somme des carrés des différences centrée normalisée* – Cette mesure, notée ZNSSD (*Zero mean Normalised Sum of Squared Differences*), est donnée par :

$$ZNSSD(\mathbf{f}_g, \mathbf{f}_d) = ZND_2(\mathbf{f}_g, \mathbf{f}_d) = ND_2(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d) = \frac{\|(\mathbf{f}_g - \bar{\mathbf{f}}_g) - (\mathbf{f}_d - \bar{\mathbf{f}}_d)\|^2}{\|\mathbf{f}_g - \bar{\mathbf{f}}_g\| \|\mathbf{f}_d - \bar{\mathbf{f}}_d\|}. \quad (3.28)$$

Une mesure de similarité utilisant ZNSSD a été proposée par Fua [Fua 93] :

$$ZNSSD_{\text{FUA}}(\mathbf{f}_g, \mathbf{f}_d) = \max(0, 1 - ZNSSD(\mathbf{f}_g, \mathbf{f}_d)). \quad (3.29)$$

Les valeurs de $ZNSSD_{\text{FUA}}(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; 1]$. Si $ZNSSD(\mathbf{f}_g, \mathbf{f}_d) > 1$, la mesure de Fua va retourner 0 quelle que soit la valeur de $ZNSSD(\mathbf{f}_g, \mathbf{f}_d)$.

3.3.2.2 Mesures localement centrées

Le principe des mesures localement centrées, proposées par Aschwanden et Guggenbül [Aschwanden 92] et notées LD_P (*Locally scaled Distances*), est de ramener la moyenne des niveaux de gris de la fenêtre de corrélation dans l'image droite à celle de la fenêtre de corrélation dans l'image gauche. Pour cela, chaque niveau de gris de l'image droite est multiplié par le rapport des moyennes des niveaux de gris des fenêtres de corrélation gauche et droite. Ces mesures sont données par :

$$LD_P(\mathbf{f}_g, \mathbf{f}_d) = D_P\left(\mathbf{f}_g, \frac{\bar{\mathbf{f}}_g}{\bar{\mathbf{f}}_d} \mathbf{f}_d\right) = \left\| \mathbf{f}_g - \frac{\bar{\mathbf{f}}_g}{\bar{\mathbf{f}}_d} \mathbf{f}_d \right\|_P^P. \quad (3.30)$$

Ce sont des mesures de dissimilarité et leurs valeurs appartiennent à l'intervalle $[0; I_{\max}^P N_f]$. Les mesures les plus utilisées sont :

- *Somme des valeurs absolues des différences localement centrée* – Cette mesure, notée LSAD (*Locally scaled Sum of Absolute Differences*), est donnée par :

$$\text{LSAD}(\mathbf{f}_g, \mathbf{f}_d) = \text{LD}_1(\mathbf{f}_g, \mathbf{f}_d) = D_1 \left(\mathbf{f}_g, \frac{\bar{\mathbf{f}}_g}{\bar{\mathbf{f}}_d} \mathbf{f}_d \right) = \left\| \mathbf{f}_g - \frac{\bar{\mathbf{f}}_g}{\bar{\mathbf{f}}_d} \mathbf{f}_d \right\|_1. \quad (3.31)$$

D'après Aschwanden et Guggenbül [Aschwanden 92], LSAD est résistante au bruit impulsionnel.

- *Somme des carrés des différences localement centrée* – Cette mesure, notée LSSD (*Locally scaled Sum of Squares Differences*), est donnée par :

$$\text{LSSD}(\mathbf{f}_g, \mathbf{f}_d) = \text{LD}_2(\mathbf{f}_g, \mathbf{f}_d) = D_2 \left(\mathbf{f}_g, \frac{\bar{\mathbf{f}}_g}{\bar{\mathbf{f}}_d} \mathbf{f}_d \right) = \left\| \mathbf{f}_g - \frac{\bar{\mathbf{f}}_g}{\bar{\mathbf{f}}_d} \mathbf{f}_d \right\|_2^2. \quad (3.32)$$

3.3.2.3 Variances

Nous pouvons distinguer la variance des différences et la variance des valeurs absolues des différences élevées à la puissance P introduites dans [Cox 95].

Variance des différences – Cette mesure, notée VD (*Variance of Differences*), est donnée par :

$$\text{VD}(\mathbf{f}_g, \mathbf{f}_d) = \text{var}(\mathbf{f}_g - \mathbf{f}_d). \quad (3.33)$$

C'est une mesure de dissimilarité. Les valeurs de $\text{VD}(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; I_{\max}^2]$.

Variance des valeurs absolues des différences élevées à la puissance P – Ces mesures, notées VAD_P (*Variance of Absolute P -powered Differences*), sont données par :

$$\text{VAD}_P(\mathbf{f}_g, \mathbf{f}_d) = \text{var} \left(|\mathbf{f}_g - \mathbf{f}_d|^P \right). \quad (3.34)$$

Ce sont des mesures de dissimilarité et leur intervalle de variation est $[0; I_{\max}^{2P}]$. Les mesures les plus utilisées sont :

- *Variance des valeurs absolues des différences* – Cette mesure, notée VOAD (*Variance Of Absolute Differences*), est donnée par :

$$\text{VOAD}(\mathbf{f}_g, \mathbf{f}_d) = \text{VAD}_1(\mathbf{f}_g, \mathbf{f}_d) = \text{var}(|\mathbf{f}_g - \mathbf{f}_d|). \quad (3.35)$$

- *Variance des carrés des différences* – Cette mesure, notée VOSD (*Variance Of Squared Differences*), est donnée par :

$$\text{VOSD}(\mathbf{f}_g, \mathbf{f}_d) = \text{VAD}_2(\mathbf{f}_g, \mathbf{f}_d) = \text{var} \left((\mathbf{f}_g - \mathbf{f}_d)^2 \right). \quad (3.36)$$

3.3.2.4 Mesure utilisant le *kurtosis*

Cette mesure, notée K_4 , proposée par Rziza *et al.* [Rziza 01] utilise un moment d'ordre 4 (coefficient d'aplatissement) :

$$K_4(\mathbf{f}_g, \mathbf{f}_d) = \left| \overline{(\mathbf{f}_g - \mathbf{f}_d)^4} - 3 \left(\overline{(\mathbf{f}_g - \mathbf{f}_d)^2} \right)^2 \right|. \quad (3.37)$$

Les valeurs de $K_4(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; I_{\max}^4]$. C'est une mesure de dissimilarité. Dans [Rziza 01], Rziza *et al.* montrent que la mesure K_4 est résistante à un bruit gaussien.

3.3.3 Mesures utilisant les dérivées des images

Les mesures présentées jusqu'à maintenant utilisaient directement les niveaux de gris. Les mesures abordées ici (cf. tableau 3.5), regroupées dans la famille DÉRIVÉE, utilisent les champs de vecteurs gradients des images. Cette troisième famille regroupe dix mesures. Elles utilisent les opérateurs de Sobel, de Roberts, de Kirsch, de Pratt, de Shen et Castan, ou encore le laplacien de gaussien.

3.3.3.1 Seitz

La forme générale des mesures de Seitz [Seitz 89], notées SEO_P (SEitz Operator), est la suivante :

$$SEO_P(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_O(\mathbf{f}_g) - \mathbf{R}_O(\mathbf{f}_d)\|_P. \quad (3.38)$$

Le terme $\mathbf{R}_O(\mathbf{f}_l)$ est le vecteur des directions des gradients de la fenêtre \mathbf{f}_l calculés avec l'opérateur choisi. Seitz a utilisé deux opérateurs : Sobel et Kirsch. Les valeurs de $SEO_P(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; I_{\max}^P N_f]$. Ce sont des mesures de dissimilarité. Elles possèdent une invariance de type gain et biais. L'étude dans [Aschwanden 92] montre que ces mesures sont assez performantes dans le cas d'un bruit impulsionnel, mais se révèlent très médiocres dans le cas d'un bruit gaussien.

Avec l'opérateur de Sobel, Seitz a proposé les mesures suivantes, notées SES_P (SEitz Sobel), avec $P = 1, 2$:

$$SES_1(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_{Sobel}(\mathbf{f}_g) - \mathbf{R}_{Sobel}(\mathbf{f}_d)\|_1 \text{ et} \quad (3.39)$$

$$SES_2(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_{Sobel}(\mathbf{f}_g) - \mathbf{R}_{Sobel}(\mathbf{f}_d)\|, \quad (3.40)$$

où $\mathbf{R}_{Sobel}(\mathbf{f}_l)$ est le vecteur des directions des gradients de la fenêtre \mathbf{f}_l calculés avec l'opérateur de Sobel. Dans le cadre de la mise en correspondance, pour calculer SES_2 , la racine carrée peut être supprimée.

De la même manière, avec l'opérateur de Kirsch, Seitz a proposé les mesures SEK_P (SEitz Kirsch), avec $P = 1, 2$:

$$SEK_1(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_{Kirsch}(\mathbf{f}_g) - \mathbf{R}_{Kirsch}(\mathbf{f}_d)\|_1 \text{ et} \quad (3.41)$$

$$SEK_2(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_{Kirsch}(\mathbf{f}_g) - \mathbf{R}_{Kirsch}(\mathbf{f}_d)\|, \quad (3.42)$$

où $\mathbf{R}_{Kirsch}(\mathbf{f}_l)$ est le vecteur des directions des gradients de la fenêtre \mathbf{f}_l calculés avec l'opérateur de Kirsch. Dans le cadre de la mise en correspondance, pour calculer SEK_2 , la racine carrée peut être supprimée.

3.3.3.2 Nishihara

La mesure de Nishihara [Nishihara 84], notée NIS, est la corrélation croisée des laplaciens binaires. Elle est donnée par :

$$NIS(\mathbf{f}_g, \mathbf{f}_d) = CC(\mathbf{R}_{LoG}(\mathbf{f}_g), \mathbf{R}_{LoG}(\mathbf{f}_d)) = \mathbf{R}_{LoG}(\mathbf{f}_g) \cdot \mathbf{R}_{LoG}(\mathbf{f}_d), \quad (3.43)$$

où $\mathbf{R}_{LoG}(\mathbf{f}_l)$ est le vecteur contenant les valeurs de l'image binaire du laplacien dans la fenêtre \mathbf{f}_l . Cette mesure utilise la corrélation croisée (cf. équation (3.14)). C'est une mesure de similarité. Les valeurs de $NIS(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; N_f]$. Dans [Aschwanden 92], la mesure de Nishihara ne s'avère pas particulièrement performante en présence de bruit ou d'occultations.

3.3.3.3 Nack

Pour les mesures de Nack [Nack 75], notées NA_m , $m = 1, 2$, les images sont convoluées avec une extension de l'opérateur de Roberts. Puis, ce résultat est binarisé grâce à un seuil adaptatif tel que seuls 15% de la zone de recherche soient au-dessus de ce seuil. Ce sont des mesures de similarité :

$$NA_1(\mathbf{f}_g, \mathbf{f}_d) = \frac{R_{Roberts}(\mathbf{f}_g) \cdot R_{Roberts}(\mathbf{f}_d)}{N_f \text{ moy}(R_{Roberts}(\mathbf{f}_d))} \text{ et} \quad (3.44)$$

$$NA_2(\mathbf{f}_g, \mathbf{f}_d) = \frac{NA_1(\mathbf{f}_g, \mathbf{f}_d)}{N_f \text{ moy}(R_{Roberts}(\mathbf{f}_g)) - R_{Roberts}(\mathbf{f}_g) \cdot R_{Roberts}(\mathbf{f}_d)}, \quad (3.45)$$

où $R_{Roberts}(\mathbf{f}_l)$ est le vecteur contenant les valeurs de l'image binaire issue de la convolution par l'opérateur de Roberts dans la fenêtre \mathbf{f}_l . Les valeurs de $NA_m(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; 1]$. Dans [Aschwanden 92], ces mesures ne s'avèrent pas robustes en présence de bruit gaussien ou impulsionnel. En revanche, leurs performances augmentent sensiblement avec la taille de la fenêtre.

3.3.3.4 Propriétés des mesures de Nishihara et Nack

Nous pouvons ajouter que ces mesures de corrélation sont assez « souples », c'est-à-dire qu'elles donnent le meilleur score aussi bien pour des pixels qui se correspondent parfaitement que pour d'autres configurations. Cette propriété peut être intéressante puisqu'elle permet aux mesures d'être plus robustes face à une occultation ou à un bruit impulsionnel, mais elle peut aussi entraîner de faux appariements. Par exemple, en prenant les vecteurs suivants :

$$\begin{aligned} R_O(\mathbf{f}_g) &= (\mathbf{0} \ 0 \ 1 \ 1 \ \mathbf{0} \ 1 \ 0 \ 1 \ 1)^T \text{ et} \\ R_O(\mathbf{f}_d) &= (\mathbf{1} \ 0 \ 1 \ 1 \ \mathbf{1} \ 1 \ 0 \ 1 \ 1)^T, \end{aligned}$$

les scores de corrélation suivants sont obtenus :

$$\begin{aligned} NIS(\mathbf{f}_g, \mathbf{f}_g) &= NIS(\mathbf{f}_g, \mathbf{f}_d) = 5, \\ NA_1(\mathbf{f}_g, \mathbf{f}_g) &= NA_1(\mathbf{f}_g, \mathbf{f}_d) = 1, \\ NA_2(\mathbf{f}_g, \mathbf{f}_g) &= NA_2(\mathbf{f}_g, \mathbf{f}_d) = 1. \end{aligned}$$

Ainsi, le meilleur score (5 pour NIS et 1 pour NA_m) est obtenu aussi bien pour deux vecteurs identiques ($R_O(\mathbf{f}_g)$ et $R_O(\mathbf{f}_g)$) que pour deux vecteurs différents ($R_O(\mathbf{f}_g)$ et $R_O(\mathbf{f}_d)$). Ainsi, l'utilisation de ces mesures peut entraîner des correspondances erronées.

3.3.3.5 Pratt

La mesure de Pratt [Pratt 78, p. 666–667], notée PRATT, est la corrélation croisée centrée normalisée (cf. équation (3.16)) appliquée aux images du laplacien. Il s'agit d'une mesure de similarité. Elle est donnée par :

$$\text{PRATT}(\mathbf{f}_g, \mathbf{f}_d) = \text{ZNCC}(R_{Pratt}(\mathbf{f}_g), R_{Pratt}(\mathbf{f}_d)), \quad (3.46)$$

où $R_{Pratt}(\mathbf{f}_l)$ est le vecteur contenant les valeurs de l'image du laplacien dans la fenêtre \mathbf{f}_l . Les valeurs de $\text{PRATT}(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[-1; 1]$.

3.3.3.6 Orientation code matching

Cette mesure, notée OCM (*Orientation Code Matching*), introduite par Ullah *et al.* [Ullah 01], s'appuie sur le codage de la direction du gradient. Le code OCM d'un pixel $\mathbf{p}_l^{i,j}$, noté $C_{OCM}(\mathbf{p}_l^{i,j})$, est obtenu de la manière suivante :

$$C_{OCM}(\mathbf{p}_l^{i,j}) = \begin{cases} \left\lfloor \frac{\theta_l^{i,j}}{\Delta\theta} \right\rfloor & \text{si } \|\nabla I_l^{i,j}\| > T_\Gamma \\ L & \text{sinon.} \end{cases} \quad (3.47)$$

Le terme $C_{OCM}(\mathbf{p}_l^{i,j})$ donne un résultat compris entre 1 et $N_{OCM} = 2\pi/\Delta\theta$, N_{OCM} étant le nombre de niveaux possibles pour le codage. Il faut fixer les constantes $\Delta\theta$, T_Γ et L . Dans [Ullah 01], seize orientations sont utilisées ; pour cela, il faut prendre $\Delta\theta = \pi/8$. De plus, Ullah *et al.* fixent les constantes telles que $L = 255$ et $T_\Gamma = 10$ et pour calculer le vecteur gradient, ils utilisent l'opérateur de Sobel.

Les vecteurs $R_{OCM}(\mathbf{f}_l)$ contenant respectivement le code OCM des pixels de la fenêtre de corrélation de l'image gauche et droite sont définis par :

$$R_{OCM}(\mathbf{f}_l) = \left(\dots C_{OCM}(\mathbf{p}_l^{i+p,j+q}) \dots \right)^T \text{ avec } p \in [-N_v; N_v] \text{ et } q \in [-N_h; N_h]. \quad (3.48)$$

La mesure de dissimilarité associée, notée OCM, est la suivante :

$$OCM(\mathbf{f}_g, \mathbf{f}_d) = \frac{1}{N_f} D_{OCM}(R_{OCM}(\mathbf{f}_g), R_{OCM}(\mathbf{f}_d)). \quad (3.49)$$

Le terme D_{OCM} est la distance définie par :

$$D_{OCM}(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} D(f_g^k, f_d^k) \text{ avec } D(a, b) = \begin{cases} \min \{|a - b|, N_{OCM} - |a - b|\} & \text{si } |a - b| < N_{OCM} \\ N_{OCM}/2 & \text{sinon.} \end{cases} \quad (3.50)$$

Les valeurs de $OCM(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; N_{OCM}/2]$. Le code OCM étant cyclique lorsqu'il y a N_{OCM} orientations, la distance maximale est $N_{OCM}/2$. C'est pour cette raison que cette valeur est affectée lorsque $|a - b| \geq N_{OCM}$. D'après Ullah *et al.*, la mesure OCM est robuste dans les cas d'occultations.

3.3.3.7 Corrélation des champs de gradients

Dans les mesures précédentes, seule la direction du gradient est utilisée. Or cette information, seule, peut entraîner des erreurs, surtout dans le cas de gradients de faible norme. Une autre mesure, appelée corrélation des champs de gradients, notée GC (*Gradient field Correlation*) et proposée par Crouzil *et al.* [Crouzil 96], utilise la similarité de vecteurs gradients et est donnée par :

$$GC(\mathbf{f}_g, \mathbf{f}_d) = \frac{\sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} \|\nabla I_g^{i+p,j+q} - \nabla I_d^{i+p,v+q}\|}{\sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} (\|\nabla I_g^{i+p,j+q}\| + \|\nabla I_d^{i+p,v+q}\|)}. \quad (3.51)$$

Il s'agit d'une mesure de dissimilarité dont les valeurs appartiennent à l'intervalle $[0; \infty[$. Cette mesure est invariante au biais. Pour calculer le gradient, deux solutions ont été proposées : le filtre de Shen et Castan [Shen 92] ou l'opérateur de Sobel. D'après Crouzil *et al.* [Crouzil 96], la mesure GC est résistante aux bruits.

3.3.4 Mesures non paramétriques

Les mesures de la famille NON PARAMÉTRIQUE (cf. tableau 3.6) sont fondées sur des transformations qui ne font aucune hypothèse sur les distributions sous-jacentes des niveaux de gris de la fenêtre de corrélation. D'une part, nous donnons les deux mesures proposées dans [Puzicha 97] et utilisées par [Smith 98, Rubner 01] : la mesure du χ^2 et la mesure de Jeffrey. Ces mesures s'appuient sur des outils des statistiques non paramétriques. Elles ont été proposées dans le cadre de la segmentation par texture mais elles nous semblent intéressantes à étudier comme mesures de corrélation dans le cadre de la mise en correspondance. D'autre part, trois grands types de mesures ont été distingués : deux mesures différentes proposées par Kaneko *et al.*, ISC (*Increment Sign Correlation*) [Kaneko 02] et SCC (*Selective Correlation Coefficient*) [Kaneko 03], deux mesures définies par Zabih et Woodfill [Zabih 94], l'une s'appuyant sur le rang (*Rank*) et l'autre sur le recensement (*Census*), et les mesures ordinales de Bhat et Nayar [Bhat 98]. Le principe des mesures ordinales est d'établir un classement des niveaux de gris des pixels au sein de la fenêtre de corrélation.

3.3.4.1 Mesure χ^2

Cette mesure, notée χ^2 , est définie par Puzicha *et al.* [Puzicha 97] :

$$\chi^2(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} \frac{(f_g^k - (f_g^k + f_d^k)/2)^2}{(f_g^k + f_d^k)/2}. \quad (3.52)$$

Cette expression peut être simplifiée :

$$\chi^2(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} \frac{(f_g^k - f_d^k)^2}{2(f_g^k + f_d^k)}. \quad (3.53)$$

C'est une mesure de dissimilarité et son intervalle de variation est $[0; I_{\max} N_f]$.

3.3.4.2 Mesure de Jeffrey

Cette mesure, notée JEFF, est définie par Puzicha *et al.* [Puzicha 97] :

$$\text{JEFF}(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} \left(f_g^k \log \frac{f_g^k}{(f_g^k + f_d^k)/2} + f_d^k \log \frac{f_d^k}{(f_g^k + f_d^k)/2} \right). \quad (3.54)$$

Il s'agit d'une mesure de dissimilarité dont l'intervalle de variation est $[0; I_{\max} N_f]$.

3.3.4.3 *Increment Sign Correlation*

La mesure ISC s'appuie sur les vecteurs \mathbf{b}_l suivants :

$$\mathbf{b}_l = \left(\dots b_l^k \dots \right)^T \text{ pour } k = 0 \cdot N_f - 2 \text{ avec } b_l^k = \begin{cases} 1 & \text{si } f_l^{k+1} \geq f_l^k \\ 0 & \text{sinon.} \end{cases} \quad (3.55)$$

Si le niveau de gris augmente entre f_l^k et f_l^{k+1} , alors b_l^k vaut 1 ou 0 dans le cas contraire. La mesure ISC compare les vecteurs \mathbf{b}_g et \mathbf{b}_d . Elle détermine si les niveaux de gris varient dans le même sens. Elle est donnée par :

$$\text{ISC}(\mathbf{f}_g, \mathbf{f}_d) = \frac{1}{N_f - 1} (\mathbf{b}_g \cdot \mathbf{b}_d + (\mathbf{1} - \mathbf{b}_g) \cdot (\mathbf{1} - \mathbf{b}_d)). \quad (3.56)$$

C'est une mesure de similarité et ses valeurs appartiennent à l'intervalle $[0; 1]$.

3.3.4.4 Selective Correlation Coefficient

La mesure SCC de « corrélation sélective » s'appuie sur le vecteur de poids \mathbf{e} défini tel que :

$$\mathbf{e} = (\dots e^k \dots)^T, k = 0 \dots N_f - 2, e^k = \begin{cases} 1 - \|b_g^k - b_d^k\| & \text{si } k = 0 \text{ ou } k \text{ pair} \\ e^{k-1} & \text{sinon.} \end{cases}$$

Si les niveaux de gris varient dans le même sens (augmentent ou diminuent) entre f_g^k et f_g^{k+1} et entre f_d^k et f_d^{k+1} alors e_l^k et e_l^{k+1} (k est pair) valent 1 ou 0 dans le cas contraire. La mesure de corrélation sélective est définie par :

$$\text{SCC}(\mathbf{f}_g, \mathbf{f}_d) = \frac{\mathbf{E}(\mathbf{f}_g - \overline{\mathbf{f}}_g) \cdot (\mathbf{f}_d - \overline{\mathbf{f}}_d)}{\|\mathbf{E}(\mathbf{f}_g - \overline{\mathbf{f}}_g)\| \|\mathbf{E}(\mathbf{f}_d - \overline{\mathbf{f}}_d)\|}. \quad (3.57)$$

La matrice \mathbf{E} contient les poids e^k sur sa diagonale et des zéros ailleurs. C'est une mesure de similarité dont les valeurs appartiennent à l'intervalle $[0; 1]$.

Dans [Kaneko 03], les auteurs remarquent qu'il y a une trop grande proportion de pixels rejetés, c'est-à-dire de pixels avec un poids nul (plus de 50%). Pour tenter de diminuer cet effet, ils proposent une modification. Des règles sur le voisinage sont ajoutées :

$$e^k = \begin{cases} 1 & \text{si } e^k = 0 \text{ et } \text{card}(\{e_v^k \mid e_v^k = 1\}) \geq \text{card}(\{e_v^k\})/2 \\ 0 & \text{si } e^k = 1 \text{ et } \text{card}(\{e_v^k \mid e_v^k = 0\}) \geq \text{card}(\{e_v^k\})/2 \\ e^k & \text{sinon,} \end{cases} \quad (3.58)$$

avec e^k un élément de \mathbf{e} et $\{e_v^k\}$ l'ensemble de ses 4-voisins.

3.3.4.5 Mesures de rang

La transformation de rang correspond au nombre de pixels dans la fenêtre de corrélation dont le niveau de gris est inférieur au niveau de gris du pixel central. Elle est donnée par :

$$\begin{aligned} \mathbf{R}_{rank}(\mathbf{f}_l) &= (\dots \text{rank}(\mathbf{p}_l^{i+p, j+q}) \dots) \text{ avec } p \in [-N_v; N_v], q \in [-N_h; N_h], \text{ et} \\ \text{rank}(\mathbf{p}_l^{i, j}) &= \text{card}(\{I_l^{i+p, j+q} \mid I_l^{i+p, j+q} < I_l^{i, j} \mid p \in [-N_v; N_v], q \in [-N_h; N_h], \}). \end{aligned} \quad (3.59)$$

La corrélation de rang, notée RANK_P , est donnée par :

$$\text{RANK}_P(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_{rank}(\mathbf{f}_g) - \mathbf{R}_{rank}(\mathbf{f}_d)\|_P^P. \quad (3.60)$$

C'est une mesure de dissimilarité et les valeurs de $\text{RANK}_P(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; N_f^{P+1}]$.

Zabih et Woodfill [Zabih 94] utilisent les normes L_1 et L_2 , ce qui donne les mesures suivantes :

$$\text{RANK}_1(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_{rank}(\mathbf{f}_g) - \mathbf{R}_{rank}(\mathbf{f}_d)\|_1 \text{ et} \quad (3.61)$$

$$\text{RANK}_2(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{R}_{rank}(\mathbf{f}_g) - \mathbf{R}_{rank}(\mathbf{f}_d)\|^2. \quad (3.62)$$

3.3.4.6 Mesure de recensement

La transformation de recensement permet d'obtenir une chaîne de bits représentative des pixels contenus dans la fenêtre de corrélation. Cette chaîne rend compte des pixels dont l'intensité est inférieure à celle du pixel central. La transformation, notée \mathbf{R}_τ , est définie par :

$$\mathbf{R}_\tau(\mathbf{f}_l) = \bigotimes_{k \in [0; N_f - 1]} \xi(f_l^{N_f/2}, f_l^k). \quad (3.63)$$

Le terme $\xi(f_l^{N_f/2}, f_l^k)$ vaut 1 si $f_l^{N_f/2} < f_l^k$ ou 0 dans le cas contraire. La mesure de corrélation associée, notée CENSUS, utilise la distance de Hamming :

$$\text{CENSUS}(\mathbf{f}_g, \mathbf{f}_d) = D_{\text{Ham}}(\mathbf{R}_\tau(\mathbf{f}_g), \mathbf{R}_\tau(\mathbf{f}_d)). \quad (3.64)$$

La mesure CENSUS est une mesure de dissimilarité et ses valeurs appartiennent à l'intervalle $[0; N_f]$.

3.3.4.7 Mesures ordinales

Un modèle de mesure α est défini par Bhat et Nayar [Bhat 98] :

$$\alpha(\mathbf{f}_g, \mathbf{f}_d) = 1 - \frac{2D_\alpha(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d))}{D_{\max}}. \quad (3.65)$$

Les termes $\mathbf{R}_\pi(\mathbf{f}_l)$ correspondent aux vecteurs contenant les rangs des pixels de \mathbf{f}_l , $D_\alpha(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d))$ est une distance à définir et D_{\max} est le maximum atteint par $D_\alpha(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d))$. Les valeurs de $\alpha(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[-1; 1]$. Il s'agit d'une mesure de similarité. Bhat et Nayar [Bhat 98] ont testé les distances suivantes : la distance de Hamming, le τ de Kendall et le ρ_s de Spearman, mais ils ont montré que ces distances ne sont pas satisfaisantes, et deux autres distances que nous allons présenter.

Mesure κ – On considère que $\mathbf{R}_\pi(\mathbf{f}_g)$ et $\mathbf{R}_\pi(\mathbf{f}_d)$ sont des permutations du vecteur $(1 \ 2 \ \dots \ N_f - 1 \ N_f)^T$. Le $i^{\text{ème}}$ élément d'une composition de permutation $\text{Comp}(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d))$ est définie par :

$$\text{Comp}^i = \mathbf{R}_\pi^k(\mathbf{f}_d), k = (\mathbf{R}_\pi^i(\mathbf{f}_g))^{-1} \text{ où } (\mathbf{R}_\pi(\mathbf{f}_g))^{-1} \text{ est l'inverse de la permutation } \mathbf{R}_\pi(\mathbf{f}_g).$$

Par exemple, si nous avons :

$$\mathbf{R}_\pi(\mathbf{f}_g) = (1 \ 3 \ 2 \ 4 \ 5 \ 6 \ 7 \ 9 \ 8)^T \text{ et } \mathbf{R}_\pi(\mathbf{f}_d) = (9 \ 3 \ 4 \ 5 \ 2 \ 6 \ 7 \ 1 \ 8)^T,$$

alors $\text{Comp}(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d)) = (9 \ 4 \ 3 \ 5 \ 2 \ 6 \ 7 \ 8 \ 1)^T$. Le $i^{\text{ème}}$ élément de la déviation $\text{Dev}(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d))$ est définie par :

$$\text{Dev}^i = \sum_{j=0}^{j=i} J(\text{Comp}^j > (i+1)) \text{ avec } J(B) = \begin{cases} 1 & \text{si B est vrai} \\ 0 & \text{sinon.} \end{cases}$$

En reprenant l'exemple précédent, nous avons :

$$\text{Dev}(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d)) = (1 \ 2 \ 2 \ 2 \ 1 \ 1 \ 1 \ 1 \ 0)^T.$$

La valeur maximale que peut prendre un élément du vecteur $\text{Dev} = (\dots \text{Dev}^k \dots)^T$ est $\lfloor N_f/2 \rfloor$. La mesure de corrélation obtenue est :

$$\kappa(\mathbf{f}_g, \mathbf{f}_d) = 1 - \frac{2 \max_{k=0..N_f-1} \text{Dev}^k}{\lfloor \frac{N_f}{2} \rfloor}. \quad (3.66)$$

Mesure χ – Une version moins coûteuse en temps de calcul que κ est la mesure χ donnée par :

$$\chi(\mathbf{f}_g, \mathbf{f}_d) = 1 - \frac{2 \text{Dev}^{N_f/2}}{\lfloor \frac{N_f}{2} \rfloor}. \quad (3.67)$$

Pour en diminuer le temps, le calcul Dev^k peut être réalisé de manière récursive, cf. [Bhat 98].

3.3.4.8 Propriétés des mesures ordinales

Toutes les mesures ordinales possèdent des invariance de type gain et biais. De plus, elles tolèrent le « factionnalisme », c'est-à-dire qu'elles sont robustes à des données très aberrantes. Cette propriété permet à ces mesures d'être robustes aux occultations. Cependant, des mesures aussi « souples » (tout comme les mesures utilisant les dérivées) posent des problèmes quand il n'y a pas d'occultation. En effet, un score de corrélation maximal peut être obtenu même quand il n'y a pas identité parfaite. À titre d'exemple, prenons les deux vecteurs de niveaux de gris suivants :

$$\mathbf{f}_g = (0 \ 1 \ 22 \ 35 \ 46 \ 58 \ 61 \ 121 \ 123)^T \text{ et}$$

$$\mathbf{f}_d = (0 \ 2 \ 42 \ 60 \ 81 \ 100 \ 123 \ 124 \ 125)^T.$$

La mesure κ de Bhat et Nayar [Bhat 98] donne les scores suivants :

$$\kappa(\mathbf{f}_g, \mathbf{f}_g) = \kappa(\mathbf{f}_g, \mathbf{f}_d) = 1.$$

Le score maximum est obtenu aussi bien avec les vecteurs $(\mathbf{f}_g, \mathbf{f}_g)$ qu'avec $(\mathbf{f}_g, \mathbf{f}_d)$. Cela peut entraîner le choix d'un correspondant erroné (cela dépend aussi du choix que l'on fait lorsque l'on a des correspondances multiples).

3.3.5 Mesures robustes

La famille des mesures robustes (cf. tableau 3.7), notée famille ROBUSTE, regroupe :

- les mesures de corrélation partielle [Zoghlami 96] et [Lan 97, chap. 3] ;
- la *quadrant correlation* [Huber 81, p. 204–205] ;
- la mesure de corrélation croisée centrée normalisée robuste [Trujillo 04] ;
- les pseudo-normes [Delon 01].

De plus, nous proposons les mesures suivantes, cf. [Chambon 03] :

- Une mesure qui utilise une variance robuste, notée MAD (*Median Absolute Deviation*).
- Trois mesures qui s'appuient sur la moindre médiane des carrés [Rousseeuw 87], notée LMP_P (*Least Median of Powers*), sur les moindres carrés tronqués [Rousseeuw 87], notée LTP_P (*Least Trimmed Powers*) et sur l'écart-type médian tronqué [Rousseeuw 92], notée SMPD_P (*Smooth Median Powered Deviation*).
- Huit mesures qui utilisent des M-estimateurs [Rousseeuw 92, Zhang 95], notées M_m, $m = 1..8$.
- Cinq mesures qui utilisent des R-estimateurs [Rousseeuw 87], notées R_m, $m = 1..5$.

Avant d'en décrire les détails, nous allons présenter le principe commun à toutes ces mesures.

3.3.5.1 Principe des mesures de corrélation robustes

Dans le cadre de ce travail, nous nous intéressons particulièrement au problème des occultations qui se traduit par le phénomène suivant :

Dans le voisinage d'un pixel situé à la frontière d'un changement de profondeur, certains pixels appartiennent au premier niveau de profondeur et d'autres au deuxième niveau de profondeur. Cette situation peut perturber la mise en correspondance et provoquer des résultats erronés.

Pour prendre en compte ce problème, les mesures robustes s'appuient sur le principe suivant :

Les pixels qui ont une profondeur différente du pixel étudié sont considérés comme des données aberrantes (cf. figure 3.3).

Ainsi, ces mesures utilisent des outils de statistiques robustes qui sont beaucoup moins sensibles aux données aberrantes que les outils des statistiques classiques.

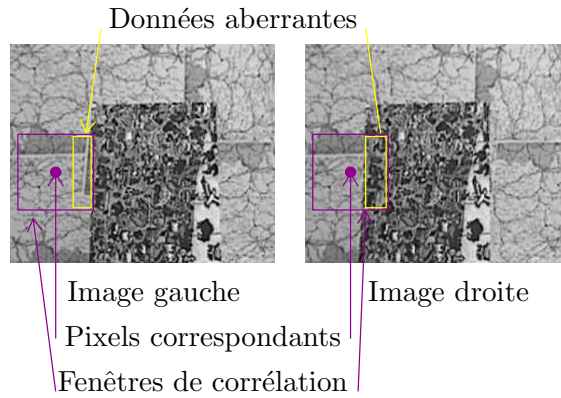


FIG. 3.3 – Principe des mesures robustes – Les deux disques violets (ou foncés) mettent en évidence deux pixels qui se correspondent parfaitement. Les rectangles jaunes (ou clairs) entourent les parties droites des fenêtres de corrélation centrées sur les pixels qui se correspondent. Ces rectangles délimitent deux zones très différentes en termes de niveaux de gris. En effet, dans l'image gauche cette zone fait partie de l'arrière plan, alors que, dans l'image droite, elle fait partie du premier plan. La comparaison de ces deux zones n'est donc pas pertinente. Le principe des mesures robustes est de considérer les niveaux de gris des pixels appartenant à cette partie de la fenêtre de corrélation comme des données aberrantes.

3.3.5.2 Corrélation partielle

Le principe des mesures de corrélation partielle est de calculer le score de corrélation en affectant des poids différents pour chaque paire de pixels. Ces poids sont déterminés en fonction du degré d'appartenance de la paire de pixels à l'ensemble des paires de pixels que l'on considère valides. Les paires qui doivent être prises en compte dans le calcul ont un poids plus élevé que les autres. Pour cela, nous définissons les matrices \mathbf{A}_l . Elles ont, sur leur diagonale, les poids w_l^k , avec $k \in [0; N_f - 1]$, appliqués aux niveaux de gris contenus dans \mathbf{f}_l et des zéros ailleurs. De plus, nous posons :

$$\mathbf{R}_m(\mathbf{f}_l) = \mathbf{A}_l \mathbf{f}_l \text{ et } \overline{\mathbf{R}_m(\mathbf{f}_l)} = \frac{1}{N_w} \sum_{k=0}^{N_f-1} \mathbf{R}_m^k(\mathbf{f}_l) \text{ avec } N_w = \sum_{k=0}^{N_f-1} w_l^k, \quad (3.68)$$

où m désigne la méthode choisie pour le calcul des poids. La forme générale des mesures de corrélation partielle, notées RM (*Reweighted Measures*), est la suivante :

$$\text{RM}_m(\mathbf{f}_g, \mathbf{f}_d) = \text{Mes}(\mathbf{R}_m(\mathbf{f}_g), \mathbf{R}_m(\mathbf{f}_d)). \quad (3.69)$$

Les mesures proposées diffèrent par le choix de la méthode pour le calcul des poids.

Mesure de Zoghلامي – Cette méthode utilise la mesure de corrélation ZNCC, cf. équation (3.16), deux fois. Elle est appliquée une première fois pour obtenir une carte de scores. Grâce à un seuil, la carte des scores est binarisée pour obtenir une carte d'occultations (on suppose que les pixels qui présentent des problèmes possèdent un score de corrélation faible). Enfin, ZNCC est appliquée, une deuxième fois, en utilisant les poids qui correspondent à la carte d'occultations calculée. Nous notons $\mathbf{R}_{F_1}(\mathbf{f}_l)$ les vecteurs obtenus après l'application des poids calculés de cette manière. La première solution proposée a été de choisir comme seuil la moyenne des scores obtenus pour chaque correspondance de l'image. Cet algorithme a été modifié par Zoghلامي *et al.* [Zoghلامي 96] en prenant comme poids

$\frac{1}{2}(\text{ZNCC}(\mathbf{f}_g, \mathbf{f}_d) + 1)$. Nous notons $\mathbf{R}_{F_2}(\mathbf{f}_1)$ les vecteurs obtenus après l'application des poids calculés de cette manière. La forme des mesures proposées est donnée par :

$$\text{ZNCC}_m(\mathbf{f}_g, \mathbf{f}_d) = \text{ZNCC}(\mathbf{R}_m(\mathbf{f}_g), \mathbf{R}_m(\mathbf{f}_d)) \text{ avec } m = F_1, F_2. \quad (3.70)$$

Méthode de Lan – Cette méthode consiste à supposer que les niveaux de gris des deux fenêtres de corrélation centrées sur deux pixels correspondants sont, en l'absence d'occultation, liés par une relation linéaire (permettant de prendre en compte les changements de luminosité entre les deux images) à laquelle vient se rajouter un bruit blanc gaussien. En d'autres termes, si l'on considère les niveaux de gris de chaque couple de pixels se trouvant à la même position dans les deux fenêtres de corrélation comme les coordonnées de points dans le plan, que l'on note $\mathbf{a}^k = (f_g^k \ f_d^k)^T$, alors on obtient un nuage de points alignés dont on aurait bruité « normalement » les positions.

Si des pixels provenant d'un objet occultant sont présents dans l'une des deux fenêtres, on peut supposer que leurs niveaux de gris sont très différents de ceux des pixels se trouvant à la même position dans l'autre fenêtre. Dans la représentation évoquée ci-dessus, les points du plan associés à ces couples de pixels ne respectent pas le modèle linéaire avec bruit « normal » suivi par les autres couples : ils constituent des données aberrantes (*outliers*). Dès lors, on peut imaginer faire appel aux techniques d'ajustement linéaire robuste permettant de détecter les points aberrants, c'est-à-dire les couples de pixels qu'il faut écarter lors du calcul du score de corrélation.

Pour cela, Lan [Lan 97, chap. 3] a proposé d'utiliser deux estimateurs robustes : la moindre médiane des carrés, LMS (*Least Median of Squares*) et l'ellipse de volume minimal, MVE (*Minimum Volume Ellipsoid*). Dans les deux cas, il utilise une technique d'échantillonnage aléatoire de type Monte-Carlo décrit dans le paragraphe suivant.

Échantillonnage aléatoire pour les estimateurs robustes – Certains estimateurs robustes de paramètres reposent sur la minimisation de critères qui n'a pas de solution directe. C'est le cas pour les estimateurs LMS et MVE.

Une première possibilité consiste à construire tous les sous-ensembles minimaux de données, à calculer les paramètres pour chacun de ces sous-ensembles et à choisir les paramètres qui minimisent le critère considéré. Le terme « minimaux » signifie ici que chaque sous-ensemble doit comporter le nombre minimum de données permettant de calculer les paramètres. Par exemple, lorsqu'on cherche les paramètres d'une droite à partir de points du plan, chaque sous-ensemble doit contenir deux points. L'inconvénient majeur de cette recherche exhaustive est le nombre de sous-ensembles, donc de calculs des paramètres, qui peut être très important : avec n données et des sous-ensembles comportant m données, on doit effectuer C_m^n calculs des paramètres.

Une alternative couramment utilisée consiste à tirer aléatoirement un nombre limité de sous-ensembles. Ce nombre dépend du risque, que l'on est prêt à prendre, de ne pas trouver le minimum global. Plus précisément, le nombre q de sous-ensembles à tirer est donné par :

$$\text{Prob} = 1 - (1 - (1 - \epsilon)^m)^q \quad (3.71)$$

où ϵ est la proportion maximale de données aberrantes dans l'ensemble initial et Prob est la probabilité qu'au moins un des sous-ensembles soit correct, c'est-à-dire qu'il ne comporte pas de donnée aberrante. En fixant *a priori* ϵ et Prob, le nombre de tirages q est donné par :

$$q = \frac{\log(1 - \text{Prob})}{\log(1 - (1 - \epsilon)^m)} \quad (3.72)$$

Nous avons pris $\epsilon = 0.5$ et Prob = 0.95, ce qui correspond à 11 tirages pour LMS ($m = 2$) et 23 tirages pour MVE ($m = 3$).

Utilisation de l'estimateur LMS – On recherche les paramètres de la droite qui minimisent le critère suivant :

$$\text{med}_{k=0..N_f-1} (r^k)^2 \quad (3.73)$$

où r^k est la distance de la droite au point \mathbf{a}^k .

Les poids indiquant l'utilisation ou le rejet de chaque couple de pixels lors du calcul du score de corrélation sont ensuite déterminés en fixant un seuil sur la distance de chaque point à la droite trouvée. Ce seuil est calculé en fonction d'une estimation robuste de l'écart-type des distances. Plus précisément, les poids sont donnés par :

$$w_{LMS}^k = \begin{cases} 1 & \text{si } \frac{|r^k|}{\sigma} \leq 2.5 \\ 0 & \text{sinon} \end{cases} \quad (3.74)$$

où l'estimation robuste de l'écart-type des distances est donnée par :

$$\sigma^k = 1.4826 \left(1 + \frac{5}{N_f - 2} \right) \sqrt{\text{med}_{k=0..N_f-1} (r^k)^2} \quad (3.75)$$

où le facteur 1.4826 est utilisé pour obtenir une estimation sans biais en présence d'un bruit gaussien et le terme $1 + 5/(N_f - 2)$ permet de corriger l'estimation lorsqu'on dispose de petits ensembles de données [Rousseeuw 87].

Enfin, le score de corrélation est donné par la mesure RZSSD (*Reweighted Zero mean Sum of Squares Differences*) [Lan 97, chap. 3] définie par :

$$\text{RZSSD}(\mathbf{f}_g, \mathbf{f}_d) = \text{RM}_{LMS}(\mathbf{f}_g, \mathbf{f}_d) = \text{ZSSD}(\mathbf{R}_{LMS}(\mathbf{f}_g), \mathbf{R}_{LMS}(\mathbf{f}_d)). \quad (3.76)$$

Utilisation de l'estimateur MVE – On recherche les paramètres de l'ellipse de volume minimal qui contient h points. Si l'on représente les paramètres d'une ellipse par le vecteur des coordonnées du centre $\bar{\mathbf{a}}$ et par la matrice de covariance \mathbf{A} et si l'on pose, comme Lan [Lan 97, chap. 3], $h = \lfloor N_f/2 \rfloor + 1$, alors les paramètres de l'ellipse de volume minimal contenant la moitié des points sont ceux qui minimisent le critère suivant :

$$\sqrt{\det(\mathbf{A})} \text{ sous la contrainte } \mathbf{A} = (\chi_{2;0.5}^2)^{-1} m^2 \mathbf{A}' \quad (3.77)$$

où $m^2 = \text{med}_{k=0..N_f-1} (\mathbf{a}^k - \bar{\mathbf{a}})^T \mathbf{A}'^{-1} (\mathbf{a}^k - \bar{\mathbf{a}})$ et la division par la médiane de la distribution du χ^2 à 2 degrés de liberté permet d'obtenir une estimation sans biais en présence d'un bruit gaussien.

Si l'on note $\hat{\bar{\mathbf{a}}}$ et $\hat{\mathbf{A}}$ les paramètres de l'ellipse recherchée, alors les poids sont calculés en fixant un seuil sur la distance de Mahalanobis entre chaque point \mathbf{a}^k et le centre $\hat{\bar{\mathbf{a}}}$, c'est-à-dire :

$$w_{MVE}^k = \begin{cases} 1 & \text{si } (\mathbf{a}^k - \hat{\bar{\mathbf{a}}})^T \hat{\mathbf{A}}^{-1} (\mathbf{a}^k - \hat{\bar{\mathbf{a}}}) \leq \chi_{2;0.975}^2 \\ 0 & \text{sinon.} \end{cases} \quad (3.78)$$

où $\chi_{2;0.975}^2$ est le seuil habituel au delà duquel un point est considéré comme aberrant.

Enfin, le score de corrélation est donné par la mesure RZNCC (*Reweighted Zero mean Normalized Cross-Correlation*) [Lan 97, chap. 3] définie par :

$$\text{RZNCC}(\mathbf{f}_g, \mathbf{f}_d) = \text{RM}_{MVE} = \text{ZNCC}(\mathbf{R}_{MVE}(\mathbf{f}_g), \mathbf{R}_{MVE}(\mathbf{f}_d)). \quad (3.79)$$

3.3.5.3 Propriétés des mesures de corrélation partielle

D'après les études faites par Zoghلامي *et al.* [Zoghلامي 96] et Lan [Lan 97, chap. 3], les mesures de corrélation partielle sont résistantes dans les zones des occultations. Cependant, nous pouvons remarquer que la mesure utilisant le MVE présente des difficultés dans le cas de parfaite correspondance. En effet, dans ce cas, quel que soit le triplet de points choisis, le MVE ne pourra pas être évalué, car la matrice de covariance ne peut pas être inversée. Nous rencontrons ce cas lorsque nous utilisons les images du stéréogramme que nous avons proposé dans notre protocole d'évaluation et de comparaison, cf. chapitre 2, paragraphe 2.9. Dans [Šára 97], la mesure RM_{MVE} est comparée avec les mesures ordinales (cf. paragraphe 3.3.4) et la mesure RM_{MVE} obtient de meilleurs résultats.

3.3.5.4 Quadrant correlation

La mesure de *quadrant correlation*, notée QUAD, est donnée par Huber [Huber 81, p. 204–205] :

$$\text{QUAD}(\mathbf{f}_g, \mathbf{f}_d) = \text{ZNCC}(\mathbf{R}_{quad}(\mathbf{f}_g), \mathbf{R}_{quad}(\mathbf{f}_d)) \quad (3.80)$$

$$\text{avec } \mathbf{R}_{quad}(\mathbf{f}_l) = \text{sgn} \left(\frac{\mathbf{f}_l - \text{med}(\mathbf{f}_l)}{\text{med} |\mathbf{f}_l - \text{med}(\mathbf{f}_l)|} \right).$$

Il s'agit d'une mesure de similarité et ses valeurs appartiennent à l'intervalle $[-1; 1]$.

3.3.5.5 Mesure de corrélation croisée centrée normalisée robuste

Cette mesure s'appuie sur ZNCC (cf. équation 3.16) en substituant la moyenne par une médiane et la norme L_2 par la norme L_1 . D'après Trujillo et Izquierdo [Trujillo 04], cette nouvelle mesure permet d'obtenir des scores dans lesquels le maximum se distingue nettement. Elle est notée ZNCC_R et elle est définie par :

$$\text{ZNCC}_R(\mathbf{f}_g, \mathbf{f}_d) = \frac{(\mathbf{f}_g - \text{med}(\mathbf{f}_g)) \cdot (\mathbf{f}_d - \text{med}(\mathbf{f}_d))}{\|\mathbf{f}_g - \text{med}(\mathbf{f}_g)\|_1 \|\mathbf{f}_d - \text{med}(\mathbf{f}_d)\|_1}. \quad (3.81)$$

Il s'agit d'une mesure de similarité dont l'intervalle de variation est $[-1; 1]$.

3.3.5.6 Pseudo-normes

Ces mesures proposées par Delon et Rougé [Delon 01] sont très proches des distances (cf. paragraphe 3.3.2.1) mais, la puissance est inférieure à 1 :

$$D_P(\mathbf{f}_g, \mathbf{f}_d) = \|\mathbf{f}_g - \mathbf{f}_d\|_P^P \text{ avec } 0 < P < 1. \quad (3.82)$$

Les valeurs $D_P(\mathbf{f}_g, \mathbf{f}_d)$ appartiennent à l'intervalle $[0; I_{\max}^P N_f]$. Ce sont des mesures de dissimilarité. À l'instar des mesures utilisant les distances, pour obtenir une invariance aux transformations affines, il est possible d'utiliser les versions centrées et normalisées de ces mesures, comme dans le paragraphe 3.3.2.1. Delon et Rougé [Delon 01] abordent le phénomène d'adhérence qui se manifeste au niveau des discontinuités provoquées par une occultation. Ce phénomène entraîne la dilatation de l'objet occultant au moment de la reconstruction. Pour les mesures de corrélation classiques, on observe ce phénomène. Delon et Rougé montrent que les pseudo-normes appliquées avec une grande fenêtre de corrélation contribuent à supprimer ce phénomène. En revanche, les pseudo-normes provoquent une érosion des coins des objets et ne sont pas indiquées pour calculer une carte de disparités dans son intégralité. Il apparaît qu'en utilisant une distance classique, plus la puissance est élevée, plus les pixels responsables d'une grande différence de niveaux de gris ont d'influence (ce qui est le cas

des pixels occultés). Au contraire, pour les pseudo-normes, plus la puissance est proche de 0, moins les pixels occultés ont d'influence. Ceci explique la robustesse des pseudo-normes aux occultations.

À présent, nous allons aborder les mesures que nous avons proposées pour tenter de prendre en compte le problème des occultations. Ces nouvelles mesures s'appuient sur le principe que nous avons déjà défini au paragraphe 3.3.5.1.

3.3.5.7 Variance robuste

Cette mesure s'appuie sur une estimation robuste de la variance des différences des niveau de gris. Il s'agit de la déviation absolue à la médiane, notée MAD (*Median Absolute Deviation*), et elle est donnée par :

$$\text{MAD}(\mathbf{f}_g, \mathbf{f}_d) = \text{med} |(\mathbf{f}_g - \mathbf{f}_d) - \text{med}(\mathbf{f}_g - \mathbf{f}_d)|. \quad (3.83)$$

Il s'agit d'une mesure de dissimilarité et ses valeurs appartiennent à l'intervalle $[0; I_{\max}]$.

3.3.5.8 Moindre médiane des puissances

Cette mesure utilise la médiane des carrés des différences des niveaux des gris, soit l'estimateur de la moindre médiane des carrés, noté LMS, défini dans [Rousseeuw 87] par :

$$\text{LMS}(\mathbf{f}_g, \mathbf{f}_d) = \text{med} ((\mathbf{f}_g - \mathbf{f}_d)^2). \quad (3.84)$$

La mesure générale utilisant LMS que nous proposons est notée LMP (*Least Median of Powers*) et elle est donnée par :

$$\text{LMP}_P(\mathbf{f}_g, \mathbf{f}_d) = \text{med} (|\mathbf{f}_g - \mathbf{f}_d|^P). \quad (3.85)$$

C'est une mesure de dissimilarité et ses valeurs appartiennent à l'intervalle $[0; I_{\max}^P]$.

3.3.5.9 Moindres puissances tronquées

Cette mesure s'appuie sur l'estimateur des moindres carrés tronqués, noté LTS (*Least Trimmed Squares*, cf. [Rousseeuw 87]) défini par :

$$\text{LTS}(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{h-1} \left((\mathbf{f}_g - \mathbf{f}_d)^2 \right)_{k:N_f-1}. \quad (3.86)$$

Son principe consiste à ranger par ordre croissant les carrés des différences des niveaux de gris, puis à faire la somme des h (nous avons choisi $h = N_f/2$) premiers carrés. La mesure générale utilisant LTS que nous proposons est LTP_P (*Least Trimmed Powers*) et elle est donnée par :

$$\text{LTP}_P(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{h-1} \left(|\mathbf{f}_g - \mathbf{f}_d|^P \right)_{k:N_f-1}. \quad (3.87)$$

C'est une mesure de dissimilarité et ses valeurs appartiennent à l'intervalle $[0; I_{\max}^P h]$.

3.3.5.10 Smooth Median Powered Deviation

Cette mesure s'appuie sur l'estimateur SMAD (*Smooth Median Absolute Deviation*), qui correspond à une estimation robuste de la variance, proposé dans [Rousseeuw 92] et défini par :

$$\text{SMAD}(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{h-1} \left(|\mathbf{f}_g - \mathbf{f}_d - \text{med}(\mathbf{f}_g - \mathbf{f}_d)|^2 \right)_{k:N_f-1}. \quad (3.88)$$

Ainsi, nous définissons la somme des h premières puissances des écarts à la médiane, notée SMPD_P (*Smooth Median Powered Deviation*), et donnée par :

$$\text{SMPD}_P(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{h-1} \left(|\mathbf{f}_g - \mathbf{f}_d - \text{med}(\mathbf{f}_g - \mathbf{f}_d)|^P \right)_{k:N_f-1}. \quad (3.89)$$

Les valeurs de cette mesure de dissimilarité appartiennent à l'intervalle $[0; \mathbf{I}_{\max}^P h]$.

3.3.5.11 M-estimateurs

Le critère des moindres carrés conduit à une estimation sensible aux données aberrantes. Le principe des M-estimateurs consiste à modifier ce critère en remplaçant le carré des résidus par une autre fonction des résidus, notée ρ et appelée *object function* ou fonction objet (pour plus de détails, cf. [Zhang 95]). La fonction ρ choisie doit être symétrique, avec un minimum unique en 0. Pour tenter de résister aux données aberrantes, elle doit croître moins rapidement que la fonction « carrée ». Les mesures s'appuyant sur des M-estimateurs que nous proposons sont notées ME_m et sont données par :

$$\text{ME}_m(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} \rho_m(f_g^k - f_d^k). \quad (3.90)$$

Les fonctions ρ_m les plus connues sont présentées dans le tableau tableau 3.1 (voir aussi la figure 3.4). L'intervalle de variation des fonctions utilisées est $[\rho_m^{\min} N_f; \rho_m^{\max} N_f]$ avec ρ_m^{\min} et ρ_m^{\max} respectivement la borne inférieure et la borne supérieure de la fonction ρ_m .

NOM	FONCTION
$L_1 - L_2$	$\rho_1(x) = \frac{\sqrt{1+x^2}}{2} - \frac{1}{2}$
<i>Fair</i>	$\rho_2(x) = x - \log(1 + x)$
Cauchy	$\rho_3(x) = \log(1 + x^2)$
Geman-McClure	$\rho_4(x) = \frac{x^2}{2(1+x^2)}$
Welsh	$\rho_5(x) = (1 - e^{-x^2})$
Tukey	$\rho_6(x) = \begin{cases} \frac{x^2}{2} & \text{si } x \leq 1.345 \\ 1.345 \left(x - \frac{1.345}{2} \right) & \text{sinon.} \end{cases}$
Huber	$\rho_7(x) = \begin{cases} \frac{x^2}{2} & \text{si } x \leq 1.345 \\ 1.345 \left(x - \frac{1.345}{2} \right) & \text{sinon.} \end{cases}$
Rousseeuw	$\rho_8(x) = 2 \log(e^x + 1) - x - 2 \log(2)$

TAB. 3.1 – Fonctions ρ_m utilisées par les M-estimateurs – Nous donnons l'expression de chaque fonction objet.

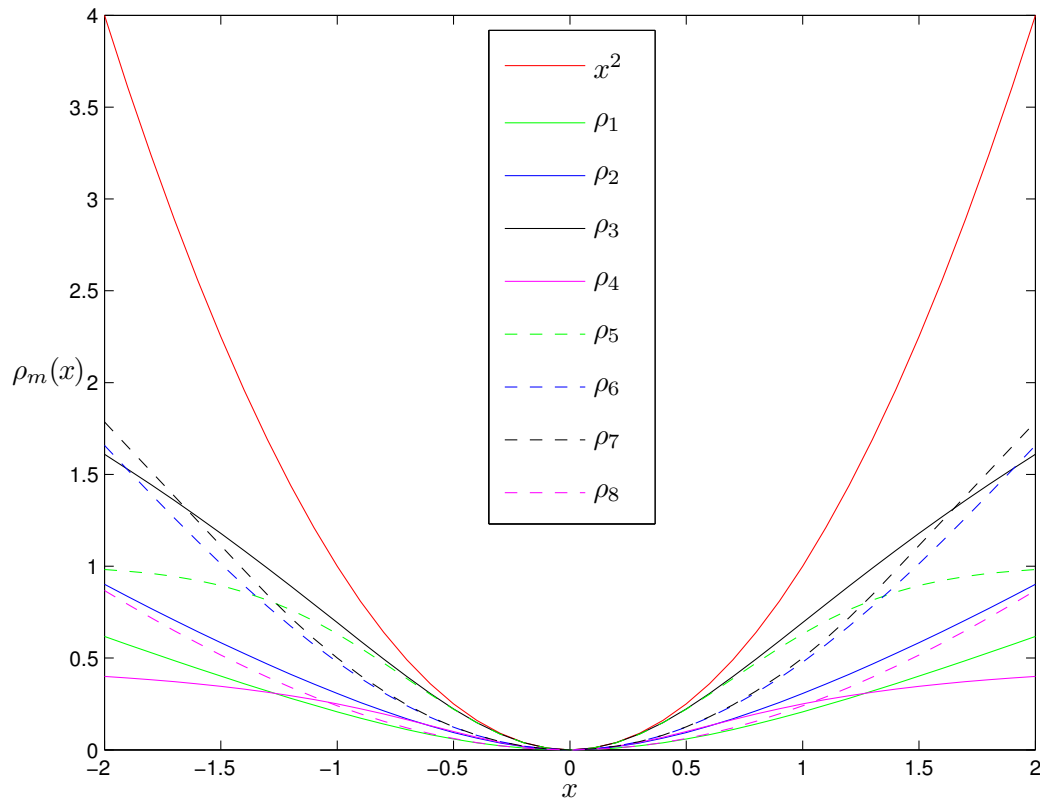


FIG. 3.4 – Graphe de visualisation des fonctions ρ_m des mesures utilisant des M-estimateurs.

3.3.5.12 R-estimateurs

Le principe des R-estimateurs est de remplacer le carré des résidus par une pondération des résidus. Les poids dépendent du rang des résidus et ils doivent être choisis de telle manière que les résidus importants, c'est-à-dire les données aberrantes, aient peu d'influence sur le calcul par rapport aux données non aberrantes. Les mesures utilisant des R-estimateurs [Rousseeuw 87] sont notées RE_m et ont la forme suivante :

$$RE_m(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} J_m \left(\frac{R_\pi(f_g^k - f_d^k)}{N_f - 1} \right) (f_g^k - f_d^k) \text{ avec } \sum_{k=0}^{N_f-1} J_m \left(\frac{R_\pi(f_g^k - f_d^k)}{N_f - 1} \right) = 0. \quad (3.91)$$

Le terme $R_\pi(f_g^k - f_d^k)$ est le rang de la différence $f_g^k - f_d^k$ dans le vecteur $\mathbf{f}_g - \mathbf{f}_d$. Les fonctions J_m les plus connues [Rousseeuw 87, Wang 92, Wiens 94] sont présentées dans le tableau 3.2 (voir aussi la figure 3.5). Ce sont des mesures de dissimilarité dont l'intervalle de variation est $[J_m^{\min} I_{\max}; J_m^{\max} I_{\max}]$ où J_m^{\min} et J_m^{\max} sont respectivement les bornes inférieures et supérieures de la fonction J_m .

NOM	FONCTION
Wilcoxon	$J_1(t) = t - \frac{1}{2}$
Médiane	$J_2(t) = \text{sgn}(t - \frac{1}{2})$
Van der Waerden	$J_3(t) = \phi^{-1}(t)$
Estimateur B-robuste optimal	$J_4(t) = \begin{cases} -1.4634 & \text{si } 0 \leq t \leq 0.39 \\ 1.47\phi^{-1}(t) & \text{si } 0.39 < t \leq 0.61 \\ 1.4634 & \text{si } 0.61 < t \leq 1 \end{cases}$
Minimax	$J_5(t) = \begin{cases} -1.14 & \text{si } 0 \leq t \leq 0.48 \\ \phi^{-1}(\frac{1}{2} + \frac{t-0.5}{t-0.1}) & \text{si } 0.48 < t \leq 0.52 \\ 1.14 & \text{si } 0.52 < t \leq 1 \end{cases}$

TAB. 3.2 – Fonctions J_m des mesures utilisant des mesures utilisant des R-estimateurs – La fonction ϕ est la fonction de densité de la distribution normale centrée réduite. Les valeurs de la fonction inverse ϕ^{-1} appartiennent à $[\phi_{\min}^{-1}; \phi_{\max}^{-1}]$, sachant que $t \in [0; 1]$.

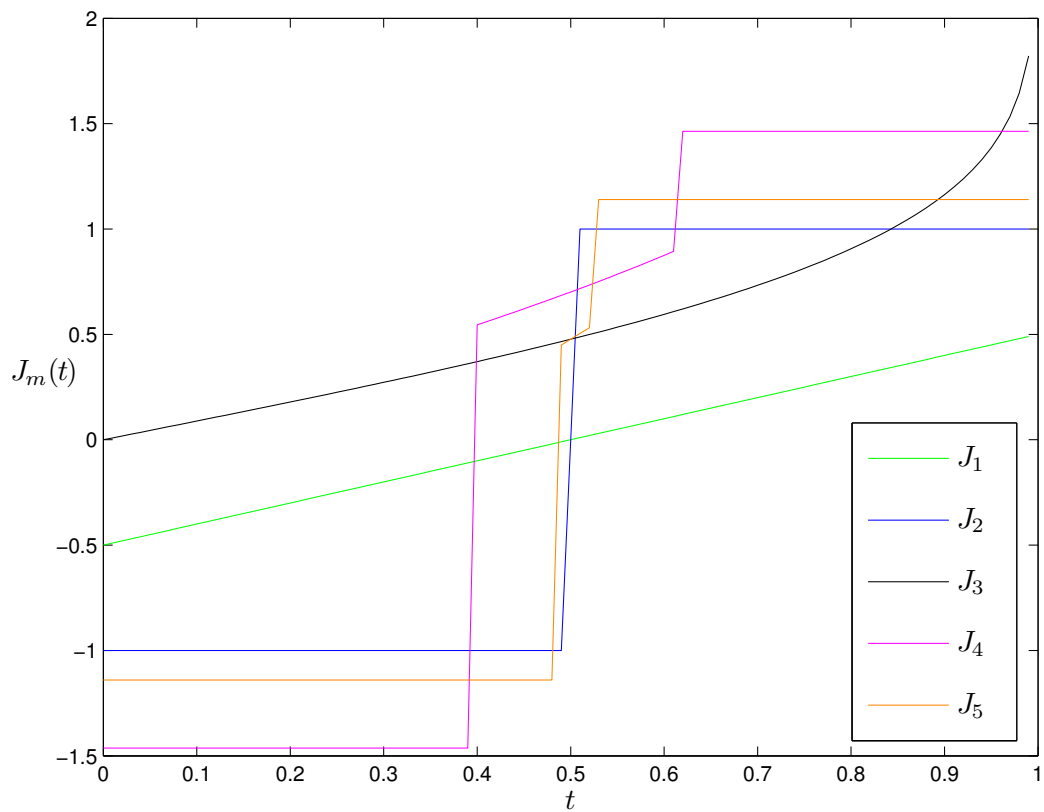


FIG. 3.5 – Graphe de visualisation des fonctions J_m des mesures utilisant des R-estimateurs.

3.4 Synthèse des mesures

Les tableaux suivants récapitulent la description des mesures de corrélation par famille en précisant leur nom, si possible, les références des articles dans lesquels elles ont été introduites, leur notation, leur définition, leur intervalle de variation dans la colonne INT., leur type dans la colonne T qui peut être similarité (S) ou dissimilarité (D) et leur invariance à des transformations affines dans la colonne I, qui peut être de type gain (G), biais (B) ou gain et biais (GB) ou ni gain ni biais (-).

NOM	NOTATION	DÉFINITION	INT.	T	I
Corrélation croisée normalisée	NCC	$\frac{\mathbf{f}_g \cdot \mathbf{f}_d}{\ \mathbf{f}_g\ \ \mathbf{f}_d\ }$	[0; 1]	S	G
Corrélation croisée centrée normalisée	ZNCC	$NCC(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d)$	[-1; 1]	S	GB
Mesure de Moravec [Moravec 80]	MOR	$\frac{2(\mathbf{f}_g - \bar{\mathbf{f}}_g) \cdot (\mathbf{f}_d - \bar{\mathbf{f}}_d)}{\ \mathbf{f}_g - \bar{\mathbf{f}}_g\ ^2 + \ \mathbf{f}_d - \bar{\mathbf{f}}_d\ ^2}$	[-1; 1]	S	B

TAB. 3.3 – Famille CROISÉE – Les mesures de cette famille exploitent la corrélation croisée.

NOM	NOTATION	DÉFINITION	INT.	T	I
Distances	D_P	$\ \mathbf{f}_g - \mathbf{f}_d\ _P^P$	$[0; I_{\max}^P N_f]$	D	-
Distances centrées	ZD_P	$D_P(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d)$	$[0; I_{\max}^P N_f]$	D	B
Distances normalisées	ND_P	$\frac{D_P(\mathbf{f}_g, \mathbf{f}_d)}{\sqrt{\ \mathbf{f}_g\ _P^P \ \mathbf{f}_d\ _P^P}}$	$[0; I_{\max}^P N_f]$	D	-
Distances centrées normalisées	ZND_P	$ND_P(\mathbf{f}_g - \bar{\mathbf{f}}_g, \mathbf{f}_d - \bar{\mathbf{f}}_d)$	$[0; I_{\max}^P N_f]$	D	B
Distances localement centrées [Aschwanden 92]	LD_P	$D_P(\mathbf{f}_g, \frac{\bar{\mathbf{f}}_g}{\bar{\mathbf{f}}_d} \mathbf{f}_d)$	$[0; I_{\max}^P N_f]$	D	-
Variance des différences [Cox 95]	VD	$\text{var}(\mathbf{f}_g - \mathbf{f}_d)$	$[0; I_{\max}^2]$	D	-
Variance des valeurs absolues des différences [Cox 95]	VAD_P	$\text{var}(\mathbf{f}_g - \mathbf{f}_d ^P)$	$[0; I_{\max}^{2P}]$	D	-
Kurtosis [Rziza 01]	K_4	$\frac{\left \overline{(\mathbf{f}_g - \mathbf{f}_d)^4} \right }{-3 \left(\overline{(\mathbf{f}_g - \mathbf{f}_d)^2} \right)^2}$	$[0; I_{\max}^4]$	D	-

TAB. 3.4 – Famille CLASSIQUE – Les mesures de cette famille utilisent des outils des statistiques classiques de la distribution des différences des niveaux de gris.

NOM	NOTATION	DÉFINITION	INT.	T	I
Seitz [Seitz 89]	SEO _P	$\ \mathbf{R}_O(\mathbf{f}_g) - \mathbf{R}_O(\mathbf{f}_d)\ _P^P$	$[0; I_{\max}^P N_f]$	D	GB
Nishihara [Nishihara 84]	NIS	$\mathbf{R}_{LoG}(\mathbf{f}_g) \cdot \mathbf{R}_{LoG}(\mathbf{f}_d)$	$[0; N_f]$	S	-
Nack 1 [Nack 75]	NA ₁	$\frac{\mathbf{R}_{Roberts}(\mathbf{f}_g) \cdot \mathbf{R}_{Roberts}(\mathbf{f}_d)}{N_f \text{ moy}(\mathbf{R}_{Roberts}(\mathbf{f}_d))}$	$[0; 1]$	S	-
Nack 2 [Nack 75]	NA ₂	$\frac{\mathbf{NA}_1(\mathbf{f}_g, \mathbf{f}_d)}{N_f \text{ moy}(\mathbf{R}_{Roberts}(\mathbf{f}_g)) - \mathbf{R}_{Roberts}(\mathbf{f}_g) \cdot \mathbf{R}_{Roberts}(\mathbf{f}_d)}$	$[0; 1]$	S	-
Pratt [Pratt 78, p. 666–667]	PRATT	$\text{ZNCC}(\mathbf{R}_{Pratt}(\mathbf{f}_g), \mathbf{R}_{Pratt}(\mathbf{f}_d))$	$[-1; 1]$	S	-
<i>Orientation code matching</i> [Ullah 01]	OCM	$\frac{1}{N_f} \text{D}_{OCM}(\mathbf{R}_{OCM}(\mathbf{f}_g), \mathbf{R}_{OCM}(\mathbf{f}_d))$	$\left[0; \frac{N_{OCM}}{2}\right]$	S	-
Corrélation des champs de gradients [Crouzil 96]	GC	$\frac{\sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} \ \nabla I_g^{i+p, j+q} - \nabla I_d^{i+p, v+q}\ }{\sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} (\ \nabla I_g^{i+p, j+q}\ + \ \nabla I_d^{i+p, v+q}\)}$	$[0; \infty[$	D	B

TAB. 3.5 – Famille DÉRIVÉE – Les mesures de cette famille s'appuient sur les gradients des niveaux de gris des images.

NOM	NOTATION	DÉFINITION	INT.	T	I
Mesure du χ^2 [Puzicha 97]	χ^2	$\sum_{k=0}^{N_f-1} \frac{(f_g^k - f_d^k)^2}{2(f_g^k + f_d^k)}$	$[0; I_{\max} N_f]$	D	-
Mesure de Jeffrey [Puzicha 97]	JEFF	$\sum_{k=0}^{N_f-1} \left(f_g^k \log \frac{f_g^k}{(f_g^k + f_d^k)/2} + f_d^k \log \frac{f_d^k}{(f_g^k + f_d^k)/2} \right)$	$[0; I_{\max} N_f]$	D	-
<i>Increment sign correlation</i> [Kaneko 02]	ISC	$\frac{1}{N_f - 1} (\mathbf{b}_g \cdot \mathbf{b}_d + (\mathbf{1} - \mathbf{b}_g) \cdot (\mathbf{1} - \mathbf{b}_d))$	$[0; 1]$	S	GB
<i>Selective Coefficient Correlation</i> [Kaneko 03]	SCC	$\frac{\mathbf{E}(\mathbf{f}_g - \bar{\mathbf{f}}_g) \cdot (\mathbf{f}_d - \bar{\mathbf{f}}_d)}{\ \mathbf{E}(\mathbf{f}_g - \bar{\mathbf{f}}_g)\ \ \mathbf{E}(\mathbf{f}_d - \bar{\mathbf{f}}_d)\ }$	$[0; 1]$	S	GB
Rang [Zabih 94]	RANK _P	$\ \mathbf{R}_{rank}(\mathbf{f}_g) - \mathbf{R}_{rank}(\mathbf{f}_d)\ _P^P$	$[0; N_f^{P+1}]$	D	GB
Recensement [Zabih 94]	CENSUS	$\text{D}_{Ham}(\mathbf{R}_\tau(\mathbf{f}_g), \mathbf{R}_\tau(\mathbf{f}_d))$	$[0; N_f]$	S	GB
Mesures ordinales [Bhat 98]	α	$1 - \frac{2\text{D}_\alpha(\mathbf{R}_\pi(\mathbf{f}_g), \mathbf{R}_\pi(\mathbf{f}_d))}{\text{D}_{\max}}$	$[-1; 1]$	S	GB

TAB. 3.6 – Famille NON PARAMÉTRIQUE – Les mesures de cette famille utilisent des transformations non paramétriques des niveaux de gris des images.

NOM	NOTATION	DÉFINITION	INT.	T	I
Corrélation partielle [Zoghلامي 96, Lan 97, chap. 3]	RM_m	$\text{Mes}(R_m(\mathbf{f}_g), R_m(\mathbf{f}_d))$	$[-1; 1]$	S	GB
Quadrant correlation [Huber 81, p. 204–205]	QUAD	$\text{ZNCC}(R_{quad}(\mathbf{f}_g), R_{quad}(\mathbf{f}_d))$	$[0; 1]$	S	GB
ZNCC robuste [Trujillo 04]	ZNCC_R	$\frac{(\mathbf{f}_g - \text{med}(\mathbf{f}_g)) \cdot (\mathbf{f}_d - \text{med}(\mathbf{f}_d))}{\ \mathbf{f}_g - \text{med}(\mathbf{f}_g)\ _1 \ \mathbf{f}_d - \text{med}(\mathbf{f}_d)\ _1}$	$[-1; 1]$	S	GB
Pseudo-norme [Delon 01]	D_P	$\ \mathbf{f}_g - \mathbf{f}_d\ _P^P$ avec $0 < P < 1$	$[0; +\infty[$	D	-
Variance robuste [Chambon 03]	MAD	$\text{med}(\mathbf{f}_g - \mathbf{f}_d - \text{med}(\mathbf{f}_g - \mathbf{f}_d))$	$[0; +\infty[$	D	GB
Moindre médiane des puissances [Chambon 03]	LMP_P	$\text{med}(\mathbf{f}_g - \mathbf{f}_d ^P)$	$[0; I_{\max}^P]$	D	GB
Moindres puissances tronquées [Chambon 03]	LTP_P	$\sum_{k=0}^{h-1} (\mathbf{f}_g - \mathbf{f}_d ^P)_{k:N_f-1}$	$[0; I_{\max}^P h]$	D	GB
Smooth Median Powered Deviation [Chambon 03]	SMPD_P	$\sum_{k=0}^{h-1} (\mathbf{f}_g - \mathbf{f}_d - \text{med}(\mathbf{f}_g - \mathbf{f}_d) ^P)_{k:N_f-1}$	$[0; I_{\max}^P h]$	D	GB
M-estimateur [Chambon 03]	ME_m	$\sum_{k=0}^{N_f-1} \rho_m(f_g^k - f_d^k)$	$[\rho_m^{\min} N_f; \rho_m^{\max} N_f]$	D	B
R-estimateur [Chambon 03]	RE_m	$\sum_{k=0}^{N_f-1} J_m \left(\frac{R_\pi(f_g^k - f_d^k)}{N_f - 1} \right) (f_g^k - f_d^k)$	$[J_m^{\min} I_{\max}; J_m^{\max} I_{\max}]$	D	GB

TAB. 3.7 – Famille ROBUSTE – Les mesures de cette famille prennent en compte le problème des occultations. Les termes ρ_m^{\min} et ρ_m^{\max} correspondent aux valeurs inférieures et supérieures des fonctions ρ_m sur l'intervalle $[0; I_{\max}]$. Les termes J_m^{\min} et J_m^{\max} correspondent aux valeurs inférieures et supérieures des fonctions J_m sur l'intervalle $[0; I_{\max}]$.

3.5 Techniques d'accélération des calculs

La technique de mise en correspondance implique $N_z^{N_g}$ scores de corrélation à évaluer, ce qui donne $N_f N_z^{N_g}$ calculs à effectuer. Pour réduire cette complexité, nous pouvons réduire N_f la taille de la fenêtre de corrélation et réduire N_z , la taille de la zone de recherche. Cependant, ces deux possibilités ne sont pas judicieuses. En effet, réduire la fenêtre de corrélation entraîne plus de difficultés pour retrouver le pixel correspondant car plus la fenêtre de corrélation est petite et moins le voisinage pris en compte est représentatif du pixel considéré. Si nous réduisons la zone de recherche, nous risquons de ne plus prendre en compte le correspondant du pixel considéré. Pour accélérer les calculs, nous distinguons les cinq techniques suivantes :

- *La simplification des mesures* – Nous pouvons supprimer certains termes de la mesure de corrélation utilisée, comme les constantes où les expressions qui ne changeront pas l'ordre des scores de corrélation. Par exemple, les facteurs $1/N_f$ ou $1/(N_f - 1)$ peuvent être supprimés dans les expressions des mesures NA_1 , OCM et ISC , cf. équations (3.44), (3.49) et (3.56). Pour la mesure NCC , cf. équation (3.15), nous pouvons supprimer le terme $\|\mathbf{f}_g\|$. En effet, sur toute la zone de recherche, $\|\mathbf{f}_g\|$ a la même valeur. Nous pouvons faire la même remarque pour ZNCC et

toutes les mesures normalisées.

- *L'utilisation de la technique des « boîtes filtrantes » (box filtering)* – Il s'agit d'une des techniques les plus employées [Faugeras 93, Chen 99, Di Stefano 02, Sun 03, Mayer 03]. Elle est applicable uniquement aux mesures de corrélation qui sont de la forme $\text{Mes}(\mathbf{f}_g, \mathbf{f}_d) = \sum_{k=0}^{N_f-1} f_{\text{dissimilarité}}(f_g^k - f_d^k)$. Le principe est le suivant :

Lorsque le score $s^{j,v}$ de corrélation entre deux pixels $\mathbf{p}_g^{i,j}$ et $\mathbf{p}_d^{i,v}$ a été évalué, le score $s^{j+1,v+1}$ entre $\mathbf{p}_g^{i,j+1}$ et $\mathbf{p}_d^{i,v+1}$ est évalué grâce à $s^{j,v}$, c'est-à-dire $s^{j+1,v+1} = s^{j,v} - s^{j_1,v_1} + s^{j_{N_z},v_{N_z}}$ où s^{j_1,v_1} correspond au score de corrélation obtenu sur la première colonne de la fenêtre de corrélation associée à $\mathbf{p}_g^{i,j}$ et $s^{j_{N_z},v_{N_z}}$ correspond au score de corrélation obtenu sur la dernière colonne de la fenêtre de corrélation associée à $\mathbf{p}_g^{i,j+1}$ (cf. figure 3.6).

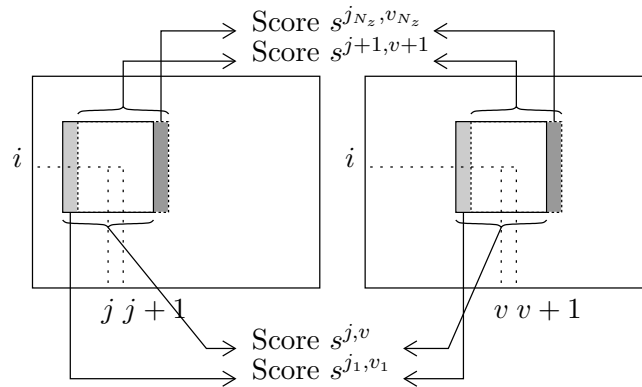


FIG. 3.6 – Boîtes filtrantes – Accélération des calculs grâce à la technique des boîtes filtrantes.

- *L'utilisation de seuils* – Cela peut être un seuil fixé (nous stoppons le calcul du score de corrélation lorsque celui-ci dépasse un seuil donné), un seuil local (nous stoppons le calcul du score de corrélation lorsque celui-ci dépasse le score minimal courant) ou un encadrement par une fonction qui nécessite moins de calculs. Cette méthode a surtout été utilisée pour des mesures classiques comme SAD et NCC [Di Stefano 03, Gong 04].
- *Un calcul préliminaire de certains termes de la mesure de corrélation* – Certaines méthodes utilisent la variance ou la moyenne des niveaux de gris sur la fenêtre de corrélation. Ces variances ou ces moyennes sont calculées, en prétraitement, pour tous les pixels de l'image [Falkenhagen 97, Menard 97, Arcara 00].
- *La mise en correspondance multirésolution* – nous avons déjà abordé cette approche au paragraphe 1.13 [Chen 03, Forstmann 04].
- *L'utilisation de matériel particulier* – Le calcul des scores de corrélation est optimisé en utilisant, par exemple, une carte graphique, cf. [Yang 93, Gong 05].

3.6 Résultats expérimentaux

3.6.1 Protocole d'évaluation

3.6.1.1 Images testées

Les images testées sont toutes celles présentées au paragraphe 2.3.4 du chapitre 2.

3.6.1.2 Critères d'évaluation

Les critères présentés au paragraphe 2.9.5 sont utilisés. Nous avons effectué une étape d'affinement des résultats au sous-pixel, aussi, ces critères sont évalués au sous-pixel près, ce qui signifie que le pourcentage d'appariements corrects correspond au pourcentage d'appariements précis. De plus, nous donnerons le rang de la mesure de corrélation parmi toutes les mesures de la même famille, noté RF.

Cependant, nous pouvons ajouter deux critères qui n'ont pas été abordés dans le chapitre 2 car ils sont très spécifiques au cas de la mise en correspondance par corrélation [De Joinville 00, De Joinville 01, Egnal 02, Trujillo 04, Mayoral 04]. Ils s'appuient sur le calcul d'une mesure d'imprécision et d'une mesure d'ambiguïté. La publication la plus significative concernant ces critères est celle de De Joinville *et al.* [De Joinville 01]. La méthode proposée dans cet article permet d'évaluer la qualité des résultats obtenus après le calcul des scores de corrélation et s'appuie sur deux notions : l'ambiguïté et l'imprécision de la mesure. Les deux critères évalués proviennent d'une analyse de la courbe constituée par les scores de corrélation sur la zone de recherche. Nous précisons que pour décrire ces critères, nos illustrations, cf. figures 3.7 et 3.8, concernent le cas de la mesure ZNCC (dont l'intervalle de variation est $[-1; 1]$). Dans un souci de clarté, contrairement au reste du mémoire, nous nous plaçons dans le cas où le score maximal est retenu. Les deux critères sont :

- *L'ambiguïté de la mesure* (cf. figure 3.7) – L'ambiguïté permet de mesurer la gravité d'une grosse erreur que l'on a pu commettre en choisissant un pixel alors qu'il en existe un autre pour lequel le score était proche. Plus précisément, s'il existe un score de corrélation $s^{j,v}$ proche de s^{j,v^*} , c'est-à-dire, tel que $|s^{j,v} - s^{j,v^*}| \leq T_a$ (T_a est un seuil à fixer) alors l'ambiguïté correspond à $|v^* - v|$.

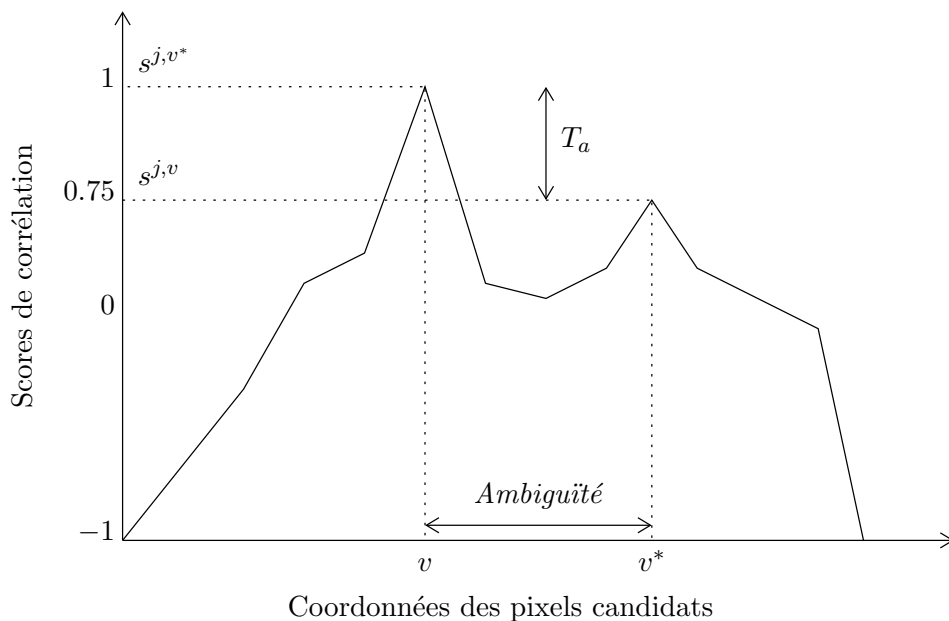


FIG. 3.7 – *Mesure d'ambiguïté* – Nous illustrons le calcul de l'ambiguïté, en utilisant la mesure ZNCC et en supposant que l'on cherche un maximum. Nous avons pris $T_a = 0.25$.

- *L'imprécision de la mesure* (cf. figure 3.8) – L'imprécision permet de quantifier l'erreur de localisation, c'est-à-dire la gravité d'une petite erreur que l'on a pu commettre. Plus précisément, il y a imprécision s'il existe v_1 et v_2 tels que $v^* \in [v_1; v_2]$ et quel que soit $v \in [v_1; v_2]$ le score $s^{j,v}$ est proche du score s^{j,v^*} , c'est-à-dire, tels que $|s^{j,v} - s^{j,v^*}| < T_i$ (T_i est un seuil à fixer) alors

l'imprécision correspond à $|v_1 - v_2|$.

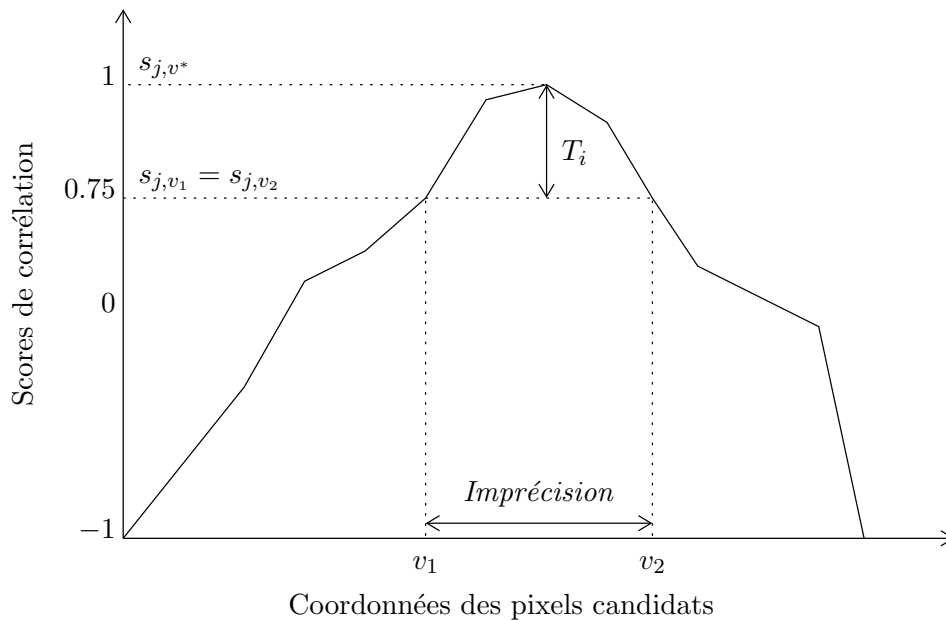


FIG. 3.8 – *Mesure d'imprécision* – Nous illustrons le calcul de l'imprécision, en utilisant la mesure ZNCC et en supposant que l'on cherche un maximum. Nous avons pris $T_i = 0.25$.

À la suite de ces calculs, nous présentons ainsi des cartes d'ambiguïté et des cartes d'imprécision : plus le pixel est clair et plus l'appariement est ambigu (respectivement imprécis).

Les tests que nous avons effectués reprennent et complètent les travaux réalisés dans le rapport [Chambon 02]. Ainsi, dans ce rapport, nous avons caractérisé le comportement de chaque mesure face à deux sortes de bruit : un bruit gaussien et un bruit impulsionnel. Nous ne reprendrons pas en détail ces évaluations, mais nous donnerons simplement un résumé des conclusions que nous avons pu en tirer.

3.6.1.3 Méthodes évaluées

Le but de cette évaluation est de pouvoir comparer les performances des mesures de corrélation, c'est pourquoi l'algorithme que nous avons choisi est minimal. Il ne fait ni appel à la multirésolution, ni à un prétraitement. Seule la contrainte de symétrie est appliquée pour tenter de détecter les pixels occultés (cf. algorithme 3.2). Les paramètres de cet algorithme sont :

- *La taille de la zone de recherche* – Nous avons choisi pour la zone de recherche $dl_{\min} = dl_{\max} = 0$, puisque nous ne testons que des images rectifiées et si l'intervalle de variation des disparités est $[\min; \max]$ (nous possédons la vérité terrain de toutes les images testées, nous connaissons donc les intervalles de variation des disparités), nous prenons $dc_{\min} = \min$ et $dc_{\max} = \max$.
- *La taille de la fenêtre de corrélation* – Nous avons choisi des fenêtres carrées centrées sur le pixel considéré, c'est-à-dire, de taille $(2N_v + 1)(2N_h + 1)$ avec $N_v = N_h \in [1; 7]$. Cette forme de fenêtre n'est pas la plus judicieuse car elle ne s'adapte pas à la forme réelle des objets et ainsi dans le voisinage du pixel considéré, nous pouvons prendre en compte des pixels qui sont des projections de points appartenant à un objet occultant. Nous souhaitons mettre en évidence le fait que certaines mesures de corrélation peuvent justement obtenir de bons résultats malgré la

présence de ces occultations ; il nous paraît donc justifié de n'utiliser que des fenêtres de cette forme pour ces tests.

Pour chaque passage faire

2.1 Si Mes \in famille DÉRIVÉE alors calcul des gradients des images

Si Mes \in {ISC, SCC, RANK_P, CENSUS, QUAD} alors Prétraiter les images

2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$

2.3 Pour chaque $\mathcal{S}^{i,j}$ faire

A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire calculer Mes($\mathbf{f}_g, \mathbf{f}_d$)

F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA

2.4 Si passage 1 alors appliquer la contrainte de symétrie puis calculer les disparités au sous-pixel

Les différents passages :

Passage 0 -- Mise en correspondance gauche -> droite

Passage 1 -- Mise en correspondance droite -> gauche

ALGO. 3.2 – *Algorithme minimal permettant de comparer les mesures de corrélation 1D – Cet algorithme est une instance de l'algorithme 1.2 à deux passages car nous appliquons la contrainte de symétrie. De plus, nous n'utilisons pas d'approche multirésolution et dans la suite de ce mémoire, pour tous les algorithmes de mise en correspondance à base de corrélation, nous allons omettre de marquer les étapes 1 et 2.5 ainsi que la boucle 2.*

3.6.1.4 Synthèse des critères et classement des méthodes

Dans ce chapitre, nous souhaitons comparer les performances des mesures de corrélation, en particulier dans les zones des occultations. Nous donnons les cartes d'ambiguïté et d'imprécision, ainsi que les cartes d'appariements corrects. Dans les tableaux de résultats, nous notons en gras, pour chaque colonne, la meilleure valeur obtenue pour le critère considéré dans la colonne.

3.6.2 Présentation et analyse des résultats

3.6.2.1 Influence de la taille des fenêtres de corrélation

La taille de la fenêtre influe sur le résultat. En effet, nous pouvons faire les deux remarques suivantes :

- Plus la taille de la fenêtre est petite, plus il est difficile de distinguer deux voisinages différents.
- Plus la taille de la fenêtre est grande et plus il y a de risque de prendre en compte des pixels qui possèdent des disparités différentes.

Ainsi, pour toutes les mesures qui ne sont pas robustes aux occultations, plus la taille de la fenêtre augmente, plus les résultats dans les zones des occultations sont mauvais. Les mesures efficaces avec de petites fenêtres (3×3 ou 5×5) sont : SAD, VOAD, K_4 , PRATT, GC, RANK_P, D_P , MAD, LMP_P, ME_m, RZSSD et RZNCC. Au contraire, les mesures peu performantes avec de petites fenêtres sont : SES_P, NIS, NA₁, NA₂, OCM, ISC, CENSUS, χ et QUAD. Grâce aux résultats que nous avons obtenus avec notre protocole et dans tous les tests que nous avons déjà effectués [Chambon 02], nous pouvons conclure que les résultats sont satisfaisants, avec toutes les images que nous avons testées, pour la majorité des critères, avec une fenêtre de taille 9×9 . C'est la taille qui semble être un bon compromis pour les images que nous avons testées ; nous la conserverons donc pour tous les tests effectués dans ce mémoire.

3.6.2.2 Temps de calcul

Suite à l'analyse des temps de calcul (cf. tableau 3.8), nous pouvons faire les remarques suivantes :

- Globalement, les mesures des familles CROISÉE, CLASSIQUE et DÉRIVÉE obtiennent des temps de calcul raisonnables.
- Les mesures des familles NON PARAMÉTRIQUE et ROBUSTE sont les plus coûteuses.
- La mesure de corrélation partielle de Lan [Lan 97] (RZNCC, famille ROBUSTE) est la plus coûteuse.

L'implémentation que nous avons choisie n'est pas la plus efficace et on peut donc penser qu'en l'améliorant (en utilisant notamment les techniques présentées au paragraphe 3.5), nous pourrions obtenir des temps de calcul raisonnables pour la plupart des mesures, excepté les mesures κ , χ et RZNCC.

MESURE	TEMPS	MESURE	TEMPS	MESURE	TEMPS	MESURE	TEMPS
NIS	2.38	ZNCC	5.4	SCC	13.02	SMPD ₂	39.77
VD	2.77	LD ₁	5.44	LMP ₂	16.65	R ₃	40.44
NCC	3.09	ME ₄	5.89	GC	17.6	R ₄	43.64
MOR	3.91	χ^2	5.93	MR ₃	24.1	ME ₅	49.48
SES ₁	4	ME ₆	5.94	LTP ₂	25.44	D _{0.1}	74.88
D ₁	4.13	K ₄	6.1	ME ₂	33.06	JEFF	83.84
NA ₁	4.57	ME ₁	6.25	MAD	33.23	QUAD	91.35
NA ₂	4.61	OCM	6.37	R ₁	35.5	κ	167.27
RANK ₁	4.73	ISC	6.67	ZNCC _R	35.55	χ	179.82
SEK ₁	4.83	ZNCC _{F₁}	8.12	R ₂	38.1	RZNCC	725.7
ZD ₁	5.23	CENSUS	8.41	R ₅	38.33		
PRATT	5.33	ME ₇	11.08	ME ₈	39.46		

TAB. 3.8 – Temps de calcul pour les différentes mesures de corrélation – Ce tableau donne les temps de calcul, en secondes, pour le couple d'images tsukuba avec une fenêtre de corrélation de taille 9×9 pour les principales mesures de corrélation. Les mesures sont classées par ordre croissant des temps de calcul.

3.6.2.3 Comportement face aux bruits

Bruit gaussien – Les différentes facettes du comportement des mesures de corrélation face à un bruit gaussien sont les suivantes :

- Plus le rapport signal sur bruit diminue, plus la taille de la fenêtre de corrélation doit être importante pour obtenir un pourcentage de correspondances correctes élevé.
- Les mesures des familles CROISÉE et CLASSIQUE sont assez résistantes face à un bruit gaussien, excepté la mesure K₄.
- Pour la famille DÉRIVÉE, seule la mesure GC a un comportement satisfaisant.
- Certaines mesures des familles NON PARAMÉTRIQUE et ROBUSTE n'obtiennent pas de bons résultats face à un bruit gaussien : RANK_P, CENSUS, les mesures de corrélation partielle, QUAD et MAD.
- Les mesures qui se comportent le mieux face à un bruit gaussien sont : SSD, ME₁ et ME₇.

Bruit impulsionnel – L'étude du comportement des mesures face à un bruit impulsionnel nous permet d'avoir une première appréciation de leur comportement face à une occultation. En effet, un

pixel « bruité » peut être considéré comme un pixel occulté. Voici les remarques que nous pouvons faire :

- Les mesures des familles CROISÉE et CLASSIQUE sont très peu robustes face à un bruit impulsionnel.
- Pratiquement toutes les mesures de la famille DÉRIVÉE se comportent correctement, excepté PRATT.
- Les mesures de la famille NON PARAMÉTRIQUE obtiennent de bons résultats, mais plus la proportion de pixels bruités augmente et plus il faut augmenter la taille de la fenêtre de corrélation pour obtenir de bons résultats.
- Les mesures de la famille ROBUSTE obtiennent les meilleurs résultats.

Synthèse – Ainsi, face à un bruit gaussien les mesures des familles CROISÉE et CLASSIQUE sont les plus performantes alors que face à un bruit impulsionnel, les mesures des familles NON PARAMÉTRIQUE et ROBUSTE sont les plus performantes. Il n'existe pas de mesure qui soit performante face à ces deux types de bruits. Ces résultats sont en adéquation, d'une part, avec les propriétés énoncées pour les familles CROISÉE et CLASSIQUE et, d'autre part, avec les prévisions faites sur les familles NON PARAMÉTRIQUE et ROBUSTE. Le lecteur intéressé peut trouver le détail des critères numériques de ce test dans [Chambon 02].

3.6.2.4 Analyse des performances

Le tableau 3.14 est une synthèse de tous les résultats que nous avons obtenus suivant chaque critère et pour la plupart des mesures. Nous exposons dans ce tableau le rang de chaque mesure pour chaque critère et le rang de chaque mesure en fonction de la moyenne des rangs sur chaque critère. Nous donnons aussi le rang de chaque mesure au sein de sa famille. Ce tableau nous permet de faire une première analyse globale de nos résultats et voici les remarques que nous pouvons faire :

- Parmi chaque famille :
 - CROISÉE – MOR est la mieux classée. Pour le pourcentage d'appariements corrects, NCC est la plus performante. Les trois mesures de cette famille obtiennent cependant des résultats assez similaires : elles font toutes les trois partie des dix premières mesures en ce qui concerne le pourcentage d'appariements corrects.
 - CLASSIQUE – D_1 (SAD) est la mieux classée, cependant, pour le pourcentage d'appariements corrects, ZD_1 (ZSAD) obtient les meilleurs résultats. Toutes les mesures de cette famille obtiennent un classement correct, notamment pour le pourcentage d'appariements corrects, excepté K_4 qui obtient de mauvais résultats.
 - DÉRIVÉE – GC est la première et nous pouvons aussi remarquer qu'elle obtient les meilleurs résultats parmi toutes les mesures pour le pourcentage d'appariements corrects et le pourcentage de faux négatifs. Hormis GC, les mesures de cette famille font partie des plus mal classées.
 - NON PARAMÉTRIQUE – ISC obtient les meilleurs résultats. Nous pouvons remarquer que $RANK_1$ est placée première pour le pourcentage d'appariements corrects dans la zone d'influence des occultations et dans la zone des discontinuités, ce qui ne l'empêche pas d'obtenir un classement moyen (rang 24 sur 44) car le pourcentage de faux positifs, notamment, est très mauvais.
 - ROBUSTE – $SMPD_2$ est la première et elle obtient aussi les meilleurs résultats parmi toutes les mesures pour le pourcentage d'appariements corrects dans la zone totale des occultations. Cette mesure est suivie de près par LTP_2 .

- La meilleure mesure, en suivant cette évaluation, est la mesure SMPD₂ et les dix premières mesures appartiennent à la famille ROBUSTE, excepté la mesure GC, classée quatrième, qui appartient à la famille DÉRIVÉE.
- En général, les mesures performantes dans les zones sans occultation, comme NCC, n'obtiennent pas les meilleurs résultats dans les différentes zones des occultations et, inversement, les mesures qui obtiennent de bons résultats dans les zones des occultations, comme SMPD₂, n'obtiennent pas les meilleurs résultats dans les zones sans occultation.

Nous ne donnons pas les classements obtenus pour chaque image séparément, cependant, en observant les différents tableaux que nous avons obtenus, nous pouvons encore faire les remarques suivantes :

- Bien qu'elle ne soit pas toujours classée première, SMPD₂ est toujours parmi les dix premières mesures, excepté pour le couple d'images *teddy* (elle est alors classée quatorzième). Ce couple d'images est en couleur. Comme dans ce chapitre, nous travaillons avec des images en niveaux de gris, nous avons transformé ces images en niveau de gris. Les images obtenues présentent alors de grandes zones non texturées. Ce qui explique les résultats de SMPD₂ qui sont moins bons qu'avec d'autres images. Nous avons un problème similaire avec les images « livres » SMPD₂ n'est classée que neuvième.
- Bien qu'elle ne soit pas toujours classée parmi les dix premières, LTP₂ est classée quatre fois première, pour les images « murs », *map*, *tsukuba* et *barn1*.

Pour les tableaux de résultats (cf. tableaux 3.9 à 3.13) et les cartes de résultats (cf. figures 3.9 à 3.12), nous donnons les résultats obtenus pour les mesures classées en première position dans leur famille respective dans le tableau 3.14. Nous donnons la variance suivant les différentes scènes des meilleures mesures par famille dans le tableau 3.9. Grâce à ce tableau, nous pouvons voir qu'il y a une variance des résultats assez importante suivant le type de scène. Par exemple, pour le pourcentage d'appariements corrects, il y a un écart-type qui varie entre 6.45 pour GC et 8.16 pour SAD. Cette variance semble cohérente avec la grande variabilité des images. En effet, nous testons un couple d'images très texturées et faciles à mettre en correspondance, le stéréogramme, des images réelles peu complexes mais peu texturées, comme *map*, des images réelles complexes, comme *cones*. À présent, nous allons affiner notre évaluation en montrant le détail des résultats les plus significatifs pour le stéréogramme aléatoire et pour les images « murs », *cones* et « livres ».

MESURE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD
MOR	83.78 (6.98)	1.14 (1.13)	0.58 (0.53)	1.94 (1.03)	1.33 (1.1)	11.24 (4.9)	68.77 (6.27)	75.99 (16.78)	56.01 (11.78)	76.23 (6)
SAD	80.7 (8.16)	1.34 (1.3)	0.69 (0.51)	2.47 (1.46)	1.18 (1)	12.85 (5.68)	69.95 (6.24)	78.54 (15.68)	56.32 (11.3)	73.94 (6.18)
GC	85.45 (6.45)	1.28 (1.19)	0.54 (0.56)	1.92 (1.04)	1.36 (1.12)	9.45 (4.06)	71.09 (5.84)	75.45 (17.21)	60.96 (11.65)	78.87 (6.93)
ISC	82.32 (7.61)	1.68 (2.37)	0.51 (0.41)	2.73 (1.02)	2 (1.34)	10.77 (5.95)	66.69 (9.45)	62.76 (20.9)	63.82 (12.64)	80.47 (5.77)
SMPD ₂	78.99 (7.22)	1.77 (1.41)	0.72 (0.45)	2.2 (1.36)	0.42 (0.86)	15.5 (4.96)	75.14 (5.73)	86.11 (10.48)	60.52 (12.82)	72.49 (6.29)

TAB. 3.9 – *Variances des mesures sur les différentes scènes – Dans chaque case de ce tableau, nous donnons la moyenne et, entre parenthèse, l'écart-type des résultats obtenus sur les quatorze couples d'images.*

Pour les tableaux de résultats (cf. tableaux 3.10 à 3.13), dans chaque case, nous donnons la valeur du critère ainsi que, entre parenthèse, le classement général de la mesure parmi toutes les mesures.

Enfin, nous précisons le rang global de la mesure par rapport à toutes les autres mesures (colonne RT) et également au sein de sa propre famille (colonne RF).

Stéréogramme aléatoire – Les résultats obtenus pour le stéréogramme aléatoire sont exposés dans le tableau 3.10 et sur la figure 3.9. Les résultats obtenus avec les différentes mesures sont assez proches (les pourcentages de pixels corrects varient entre 97.3% et 98.26%); c’est pourquoi, il est difficile de les distinguer. Toutefois, nous pouvons remarquer que $SMPD_2$ n’est pas classée première. En réalité, les mesures MAD et LMP_2 obtiennent les meilleurs résultats. Nous pouvons aussi souligner le bon comportement de SEK_1 , CENSUS et κ qui sont classées troisième, huitième et septième.

Murs – Les résultats obtenus pour le couples d’images « murs » sont exposés dans le tableau 3.11 et sur la figure 3.10. Neuf des mesures les mieux classées appartiennent à la famille ROBUSTE et SAD est classée neuvième. Pour ce couple d’images, c’est LTP_2 qui est la mieux classée. Elle est classée première pour le pourcentage d’appariements corrects dans la zone totale des occultations et dans la zone d’influence des occultations. Elle obtient aussi de meilleurs résultats que $SMPD_2$ pour le pourcentage d’appariements corrects dans toute l’image. La différence entre ces deux mesures est que $SMPD_2$ est centrée (on utilise la médiane) alors que LTP_2 ne l’est pas. Le fait de centrer la mesure peut, entre autres, rendre la mesure plus robuste à un changement de luminosité. Or, dans ces images de synthèse, il n’y a pas de changement de luminosité. Nous pouvons donc supposer que l’utilisation d’une mesure centrée détériore les résultats au lieu de les améliorer. Nous observons d’ailleurs la même différence entre les résultats obtenus par SAD et ZSAD.

Cones – Les résultats obtenus pour le couples d’images *cones* sont exposés dans le tableau 3.12 et sur la figure 3.11. Malgré un grand nombre d’occultations, ces images présentent moins de difficultés d’appariement que les autres couples que nous avons testés. En effet, il n’y a pas de zones uniformes et les zones des occultations sont nombreuses mais peu étendues. Ainsi, les mesures des familles CROISÉE et CLASSIQUE obtiennent de meilleurs résultats qu’avec les autres images.

Livres – Les résultats obtenus pour le couples d’images « livres » sont exposés dans le tableau 3.13 et sur la figure 3.12. Ces images sont difficiles à apparier car il y a de très nombreuses zones non texturées. Ce sont globalement les mesures de la famille ROBUSTE qui sont les mieux classées, excepté GC et SAD qui sont respectivement classées dixième et cinquième.

3.7 Conclusion

Ce chapitre nous a permis, d’une part, de présenter les différentes mesures de corrélation existantes, en les classant et en les caractérisant de manière précise et, d’autre part, de proposer de nouvelles mesures s’appuyant sur des outils de statistiques robustes, pour tenter de prendre en compte les occultations.

Les résultats que nous avons obtenus montrent que les mesures de la famille ROBUSTE obtiennent généralement les meilleurs résultats, notamment dans la zone totale des occultations. En revanche, elles se comportent moins bien dans le reste de l’image contrairement à la mesure GC de la famille DÉRIVÉE, aux mesures de la famille CROISÉE ainsi qu’aux mesures SAD et ZSAD, de la famille CLASSIQUE, qui se comportent bien dans le reste de l’image.

Les résultats montrent aussi que les méthodes de mise en correspondance par corrélation rencontrent des difficultés dans les zones non texturées.

Pour tenter d'améliorer les performances des méthodes à base de corrélation, nous pouvons envisager, d'une part, d'utiliser une information plus riche que les niveaux de gris, la couleur, et, d'autre part, d'utiliser la combinaison de différentes mesures de corrélation. Nous allons aborder ces deux aspects dans les chapitres suivants.

MESURE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT	RF
MOR	97.31 (26)	0.02 (12)	0.01 (13)	0.87 (29)	0.55 (16)	1.27 (26)	64.66 (15)	66 (17)	63.24 (13)	97.7 (33)	13	1
SAD	97.49 (13)	0 (38)	0.01 (13)	0.79 (11)	0.61 (33)	1.1 (11)	62.81 (29)	62.6 (33)	63.03 (17)	98.72 (13)	18	1
GC	97.3 (29)	0.06 (2)	0.02 (9)	1.06 (42)	0.53 (11)	1.11 (14)	66.56 (8)	67.6 (11)	65.47 (2)	98.3 (21)	4	2
ISC	97.37 (24)	0.02 (12)	0.03 (7)	0.85 (27)	0.5 (9)	1.28 (31)	66.77 (7)	69 (10)	64.41 (5)	98.02 (26)	7	1
SMPD ₂	98.26 (1)	0.02 (12)	0.01 (13)	0.1 (2)	0.63 (38)	1.02 (7)	62.14 (36)	61.4 (38)	62.92 (21)	99.9 (1)	5	3

TAB. 3.10 – Résultats obtenus avec le stéréogramme aléatoire – Contrairement aux résultats que nous verrons par la suite, SMPD₂ obtient les meilleurs résultats pour le pourcentage d'appariements corrects, notamment dans la zone des discontinuités.

MESURE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT	RF
MOR	85.56 (14)	0.99 (24)	0.17 (24)	0.99 (27)	0.33 (39)	11.96 (14)	80.37 (33)	94.14 (37)	37.73 (19)	73.71 (6)	27	1
SAD	88.43 (6)	0.63 (37)	0.08 (40)	0.6 (6)	0.17 (14)	10.1 (7)	82.72 (14)	96.98 (13)	38.56 (17)	70.74 (12)	9	1
GC	89.39 (1)	1 (23)	0.1 (34)	0.81 (20)	0.34 (40)	8.35 (3)	80.86 (28)	94.03 (38)	40.06 (11)	77.78 (1)	15	1
ISC	60.02 (34)	2.24 (9)	0.74 (7)	6.18 (40)	0.37 (41)	30.45 (34)	80.5 (32)	93.58 (39)	40.02 (12)	57.1 (33)	35	6
SMPD ₂	78.76 (24)	2.06 (11)	0.44 (13)	1.04 (29)	0.07 (3)	17.64 (26)	86.06 (4)	98.81 (1)	46.56 (4)	71.15 (10)	5	5

TAB. 3.11 – Résultats obtenus avec le couple d'images « murs » – Ces résultats illustrent le fait que les mesures ordinales n'obtiennent pas de très bons résultats. GC obtient le pourcentage d'appariements corrects le plus élevé aussi bien dans toute l'image que dans la zone des discontinuités.

MESURE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT	RF
MOR	80.56 (13)	1.23 (17)	0.59 (17)	2.85 (25)	3.99 (43)	10.78 (7)	69.21 (22)	71.48 (41)	67.63 (6)	75.83 (5)	9	1
SAD	74.61 (25)	1.75 (12)	0.81 (13)	3.34 (31)	3.53 (35)	5.96 (3)	69.75 (16)	74.8 (33)	66.23 (14)	70.04 (19)	14	3
GC	82.66 (4)	1.23 (17)	0.5 (24)	2.54 (19)	4.09 (44)	8.98 (5)	72.87 (4)	70.74 (42)	74.36 (3)	80.24 (2)	3	1
ISC	82.87 (3)	0.37 (40)	0.31 (36)	2.34 (16)	3.18 (21)	10.92 (10)	73.79 (2)	77.26 (19)	71.37 (4)	76.81 (4)	2	1
SMPD ₂	85.86 (1)	0.46 (39)	0.2 (42)	1.22 (5)	2.91 (19)	9.07 (6)	77.4 (1)	79.2 (17)	76.14 (1)	78.87 (3)	1	1

TAB. 3.12 – Résultats obtenus avec le couple d'images cones – Pour ce couple, MOR obtient un meilleur classement qu'avec les autres images et c'est aussi le cas pour toutes les mesures de sa famille (famille CROISÉE). SMPD₂ obtient les meilleurs résultats aussi bien dans la zone d'influence des occultations que dans le reste de l'image.

MESURE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT	RF
MOR	66.45 (4)	3.81 (25)	1.3 (24)	2.14 (14)	0.33 (32)	25.97 (6)	69.73 (17)	94.26 (31)	29.47 (18)	64.68 (8)	15	13
SAD	64.33 (16)	4.73 (11)	1.61 (12)	2.3 (19)	0.23 (11)	26.8 (12)	70.8 (9)	95.92 (11)	29.58 (16)	63.85 (14)	5	7
GC	69.22 (1)	4.18 (16)	1.37 (18)	2.93 (25)	0.49 (41)	21.8 (3)	70.64 (11)	91.42 (39)	36.54 (1)	81.41 (1)	10	4
ISC	59.69 (22)	2.77 (38)	0.94 (35)	4.46 (34)	0.63 (44)	31.5 (25)	67.36 (35)	88.91 (42)	31.99 (8)	64.47 (9)	32	9
SMPD ₂	64.91 (12)	4.14 (17)	1.31 (21)	2.2 (16)	0.17 (3)	27.27 (15)	72.36 (3)	97.04 (2)	31.86 (9)	75.16 (2)	1	1

TAB. 3.13 – Résultats obtenus avec le couple d'images « livres » – Les résultats dans la zone d'influence des occultations sont globalement mauvais (il y a au plus 36.54% d'appariements corrects dans cette zone). Ces mauvais résultats sont sûrement dus au manque de texture dans les zones d'influence. SMPD₂ est toujours première, mais c'est tout de même GC qui se comporte le mieux dans la zone d'influence des occultations et qui obtient le meilleur pourcentage d'appariements corrects.

FAMILLE	MESURE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT	RF
CROISÉE	NCC	7	35	34	12	30	9	32	29	34	23	32	3
	ZNCC	10	37	34	14	37	12	28	37	8	5	23	2
	MOR	8	30	28	17	40	7	25	41	9	3	19	1
CLASSIQUE	D ₁	16	22	21	22	24	15	15	23	10	12	11	1
	ZD ₁	4	34	32	10	28	5	24	28	23	9	14	2
	LD ₁	5	39	40	6	20	8	26	20	29	21	20	3
	VD	5	30	29	13	34	6	31	33	31	20	30	4
	K ₄	39	4	3	42	39	35	41	40	40	38	40	5
DÉRIVÉE	SES ₁	42	11	14	37	5	42	39	14	43	43	37	2
	SEK ₁	43	6	6	43	32	41	42	31	41	41	42	5
	NIS	44	21	12	41	4	44	44	4	44	44	38	3
	OCM	37	10	11	40	41	37	38	39	37	37	43	6
	PRATT	41	33	1	44	36	39	36	35	26	34	41	4
	GC	1	19	31	15	42	1	4	42	2	2	4	1
NON PARAMÉTRIQUE	JEFF	35	3	2	38	38	31	34	38	24	28	35	7
	χ_2	17	17	22	27	33	17	33	34	33	26	34	6
	ISC	23	36	13	35	21	21	8	19	6	19	15	1
	SCC	21	14	16	28	35	22	27	36	11	22	30	5
	RANK ₁	15	24	36	30	43	3	29	43	1	1	24	3
	CENSUS	40	15	18	29	9	43	37	7	39	42	36	8
	κ	27	16	33	9	14	29	18	13	30	31	22	2
	χ	31	13	18	19	10	32	23	8	36	36	26	4
ROBUSTE	ZNCC _{F1}	25	40	30	26	44	28	43	44	32	28	44	22
	RM _{LMS}	30	32	42	2	1	34	9	1	35	39	24	16
	QUAD	22	27	27	11	12	27	10	11	25	30	17	13
	ZNCC _R	33	42	43	3	17	33	40	17	42	40	39	21
	D_{0,1}	29	5	7	33	15	23	7	15	12	18	6	5
	MAD	34	44	44	1	3	38	12	3	19	33	28	18
	LMP ₂	36	29	37	7	2	40	13	2	21	32	21	15
	LTP ₂	24	9	8	31	7	24	2	6	4	14	2	2
	SMPD₂	20	12	14	21	6	20	1	5	3	13	1	1
	ME₁	13	28	24	16	23	11	16	25	13	10	10	9
	ME₂	18	20	17	24	27	16	11	22	7	6	8	7
	ME₃	19	18	25	20	18	19	3	18	5	6	3	3
	ME₄	28	7	9	34	11	26	5	9	16	25	9	8
	ME ₅	32	2	5	36	13	30	19	12	28	27	18	14
	ME ₆	38	1	4	39	16	36	35	15	38	35	33	20
	ME ₇	14	26	23	18	22	14	14	24	15	11	12	10
	ME₈	26	8	10	32	8	25	5	10	13	24	5	4
	R ₁	2	43	37	5	26	2	21	26	22	17	15	12
	R ₂	12	25	20	25	31	13	30	32	27	16	28	18
	R₃	3	37	41	4	19	4	17	20	18	4	7	6
R ₄	11	41	39	8	25	18	22	27	20	15	26	17	
R ₅	9	23	26	22	29	10	20	29	17	8	13	11	

TAB. 3.14 – Classement des mesures de corrélation – Nous avons synthétisé les résultats obtenus sur les quatorze images de tests avec le protocole que nous avons décrit dans le paragraphe 2.9. Les mesures en gras sont les dix premières. Les mesures encadrées correspondent aux mesures classées premières dans leur famille et la mesure sur fond grisé est celle qui obtient les meilleurs résultats.

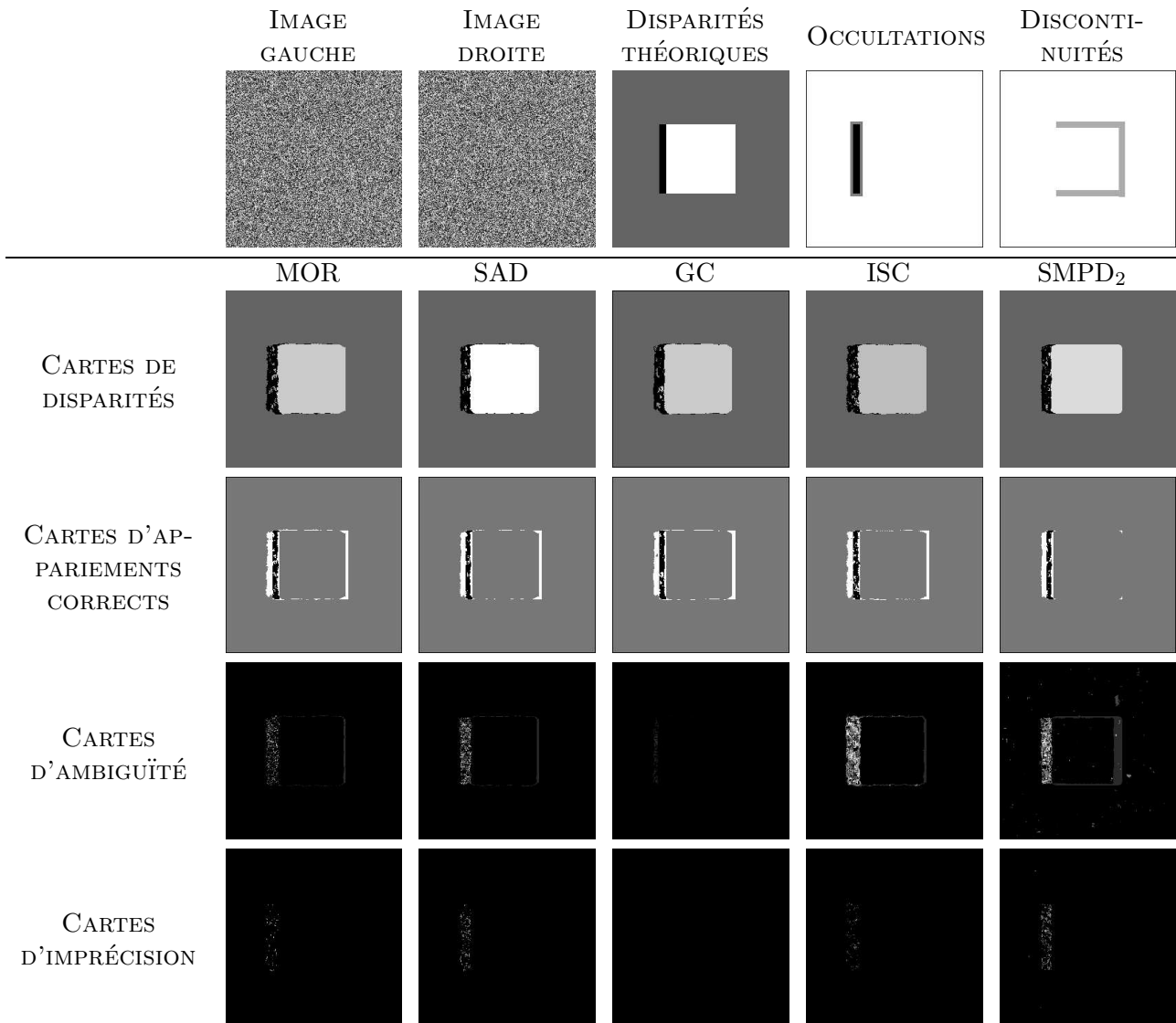


FIG. 3.9 – Cartes de résultats obtenus avec le stéréogramme aléatoire – La carte de disparités la plus « nette » est celle obtenue par la mesure SMPD₂. La carte d'appariements corrects permet d'illustrer le bon comportement de cette mesure dans la zone des discontinuités où elle est ambiguë.

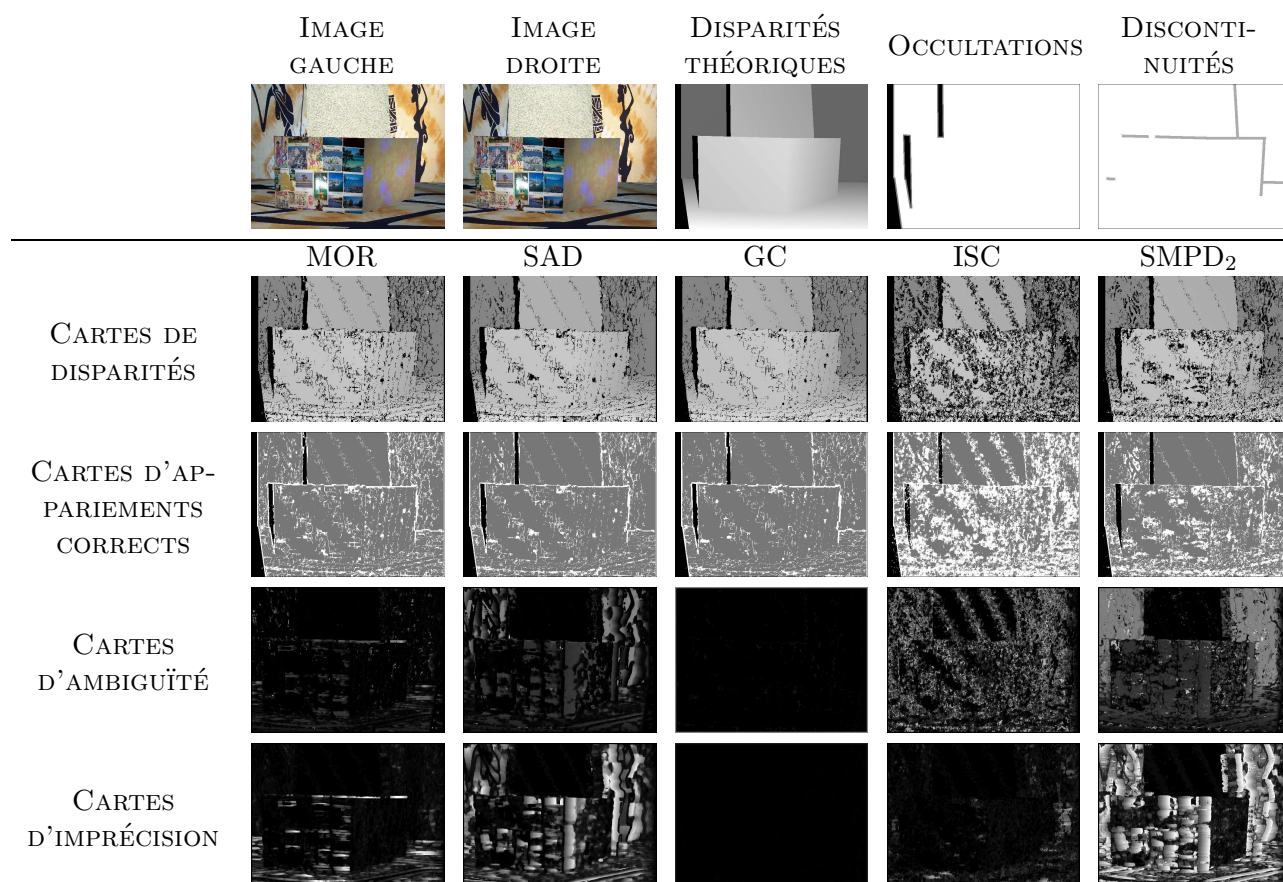


FIG. 3.10 – Cartes de résultats obtenus avec le couple d'images « murs » – Ces résultats montrent que la carte de disparités obtenue avec la mesure SMPD₂ n'est pas la plus dense (il y a trop de pixels détectés comme occultés) mais en revanche, les zones des occultations sont beaucoup plus nettes. Nous pouvons aussi remarquer le peu d'ambiguïté de la mesure MOR et le taux de faux négatifs beaucoup trop élevé de la mesure ISC. La mesure GC est très peu ambiguë et imprécise.

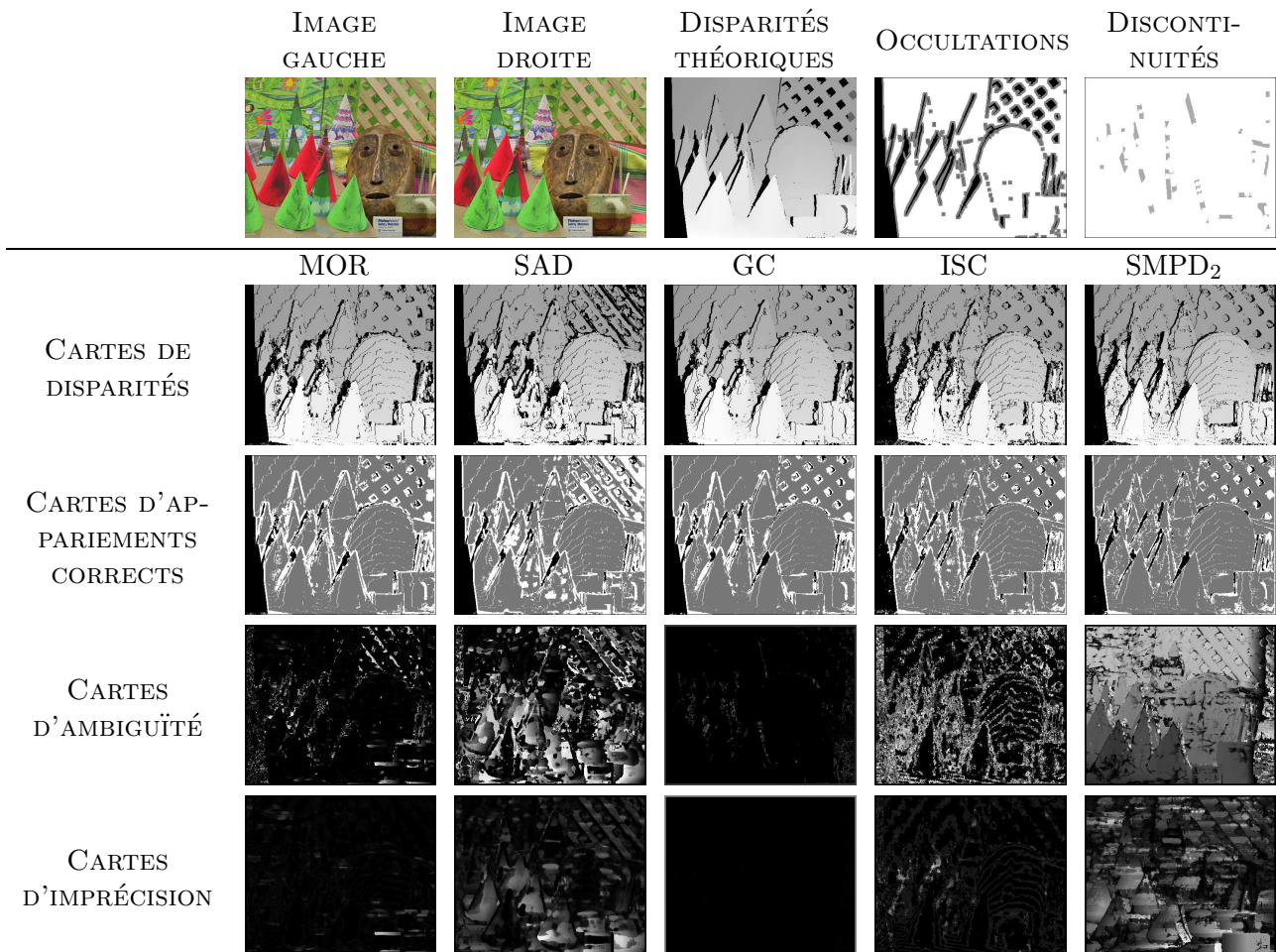


FIG. 3.11 – Cartes de résultats obtenus avec le couple d'images cones – Pour ce couple d'images, les différences entre les différentes cartes de disparités sont visiblement moins importantes. Nous pouvons toutefois toujours remarquer le bon comportement de la mesure SMPD₂ dans les zones proches des occultations où elle est la plus ambiguë.

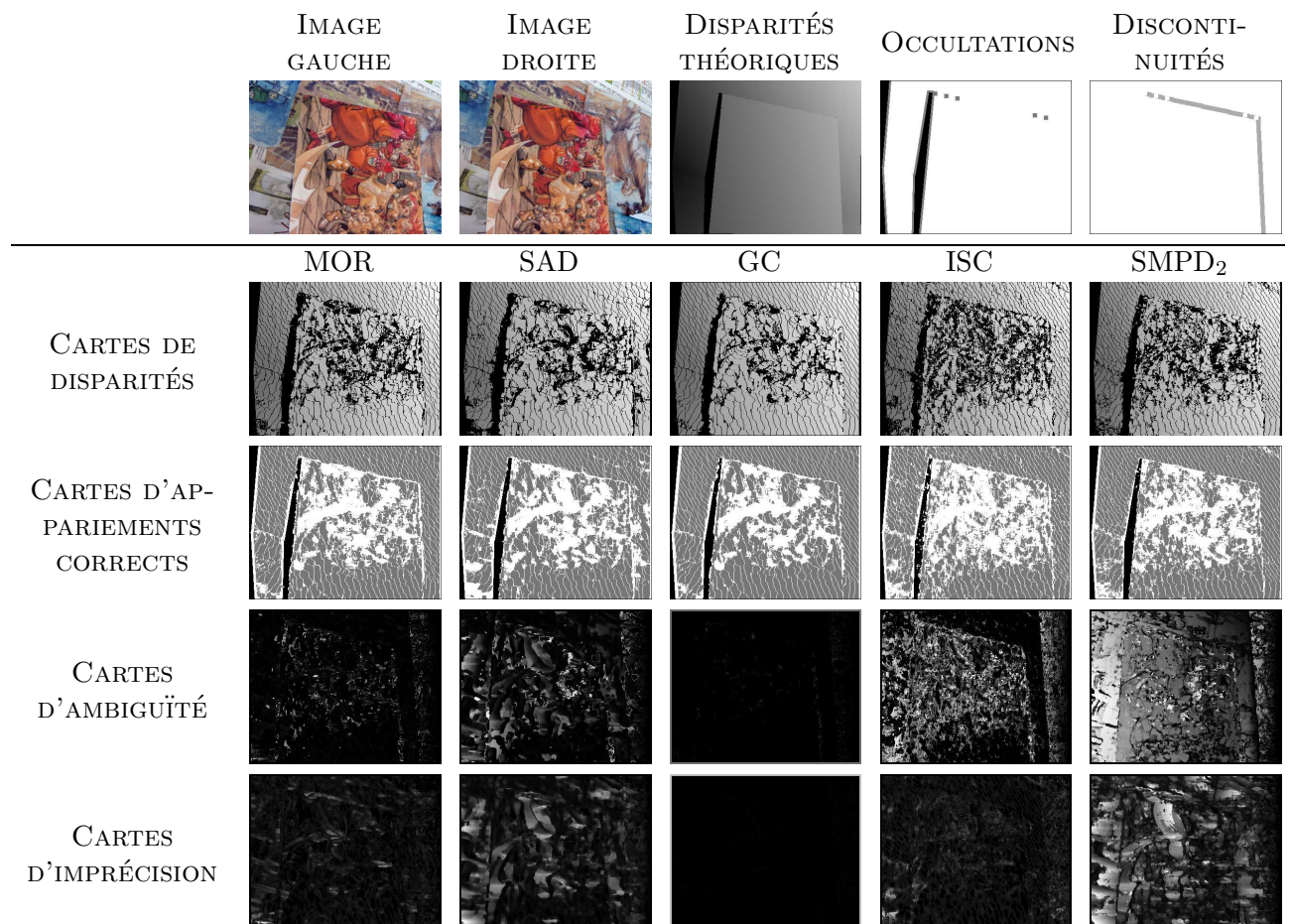


FIG. 3.12 – Cartes de résultats obtenus avec le couple d'images « livres » – Pour ce couple d'images, le premier plan est apparemment assez difficile à apparier (il manque de textures). Même si GC n'est pas classée première, nous pouvons voir qu'elle obtient des bons résultats.

Chapitre 4

Mise en correspondance par corrélation couleur – Corrélation 3D

4.1 Introduction

4.1.1 Objectifs

Jusqu'à présent, les travaux présentés dans ce mémoire ne prennent en compte que des images en niveaux de gris. Or, actuellement, la majorité des capteurs numériques permettent l'acquisition d'images couleur. De plus, les images couleur sont de plus en plus employées en traitement d'images et en vision par ordinateur, cf. [Duffy 98, Wesolkowski 99, Schlüns 00, Gevers 03, Jaffré 03, Smolka 03, Buessler 05, Busin 05]. Certains travaux qui concernent la mise en correspondance, comme dans [Okutomi 92b, Rubner 01, Perwass 01, Georgescu 04], ont montré le gain de performance apporté par la couleur. En effet, alors que les résultats obtenus avec les niveaux de gris sont souvent ambigus [Chambon 03], l'utilisation de plusieurs composantes permet de diminuer cette ambiguïté et ainsi, entre autres, d'obtenir plus de robustesse près des zones des occultations. Il nous est ainsi apparu naturel d'étendre à la couleur nos recherches sur la mise en correspondance par corrélation, cf. [Chambon 05], et d'observer l'amélioration des résultats dans toute l'image mais aussi près des zones des occultations, cf. [Chambon 04a]. Pour cela, il est nécessaire d'adapter à la couleur la mise en correspondance par corrélation.

4.1.2 Stratégie de généralisation

À notre connaissance, il existe peu de travaux sur la généralisation même des mesures de corrélation à la couleur, et ils ne prennent souvent en compte que les mesures de corrélation les plus utilisées : corrélation croisée, cf. [Belli 00], somme des carrés des différences, cf. [Koschan 96, Coquin 02, Hansen 03], somme des carrés des différences centrée normalisée, cf. [Devernay 97], somme des valeurs absolues des différences, cf. [Mühlmann 01], et corrélation des champs de gradients, cf. [Crouzil 97, p. 61–65]. En réalité, la généralisation des mesures de corrélation aux images couleur est délicate et pose plusieurs problèmes :

- *Le choix d'un système de représentation* – De nombreux travaux sont consacrés à la comparaison des espaces de couleur pour des opérations spécifiques en traitement d'images numériques, notamment dans [Vandenbroucke 00, p. 51–89], [Hanbury 03] et [Trémeau 04, chap. 5 et 7], mais ils ne concernent pas la mise en correspondance par corrélation. Ainsi, dans ce cadre précis, il est important d'étudier les différents systèmes pour déterminer celui qui permet d'obtenir les meilleurs résultats.

- *Le choix d'une méthode de généralisation* – Pour cela, trois possibilités se présentent :
 - évaluer la mesure de corrélation sur chaque composante, puis effectuer une fusion des résultats obtenus sur chaque composante : cette approche peut être appelée « approche marginale », cf. [Lambert 02, p. 13] ;
 - effectuer une analyse en composantes principales (ACP), et appliquer la mesure de corrélation sur la première composante principale ;
 - généraliser directement les mesures de corrélation aux données à trois dimensions : cette méthode peut être vue comme une « approche vectorielle », cf. [Lambert 02, p. 13].

La dernière méthode qui consiste à adapter directement les mesures de corrélation est plus délicate que les autres. En effet, elle implique le problème de l'adaptation de certains opérateurs à des données tridimensionnelles comme, par exemple, le calcul de la distance entre deux couleurs, le calcul des rangs ou de la médiane d'un ensemble de couleurs.

4.1.3 Plan

Dans un premier temps, les deux aspects de la généralisation des mesures de corrélation à la couleur vont être présentés : les systèmes de représentation de la couleur et les trois méthodes de généralisation. Puis, les résultats obtenus avec des images de synthèse et des images réelles couleur, suivant les différentes méthodes et les différents systèmes seront présentés.

4.2 Représentation de la couleur

De nombreux travaux portent sur le processus de formation des couleurs et le système de perception des couleurs de la vision humaine, cf. [Levine 85, chap. 3, 4 et 7], [Carron 95, chap. 1], [Brun 96, chap. 1], [Hardeberg 99, chap. 1] et [Chambah 01, chap. 2], mais, ici, seuls les systèmes de représentation de la couleur seront abordés. Les termes utilisés sont :

- *La luminance* – Il s'agit de la notion d'émettre plus ou moins de lumière.
- *La teinte* – Elle correspond à une couleur, une longueur d'onde dominante.
- *La saturation* – Elle indique le niveau de coloration d'une teinte, la pureté de la couleur.

La teinte et la saturation définissent la chrominance d'une couleur. Un système de représentation de la couleur est un système de coordonnées qui permet de représenter la couleur. Ce système, d'après la CIE (Commission Internationale de l'Éclairage), doit posséder trois composantes par analogie au système visuel humain qui comporte trois cônes différents, réceptifs à trois longueurs d'ondes différentes. Une couleur est représentée par les valeurs de ces trois composantes.

Comme dans [Vandenbroucke 00, chap. 2], cinq catégories de systèmes peuvent être distinguées :

- les systèmes de primaires ;
- les systèmes luminance-chrominance ;
- les systèmes perceptuels ;
- les systèmes d'axes indépendants ;
- les systèmes hybrides.

Les quatre premières catégories sont classées dans les systèmes classiques, par opposition aux systèmes hybrides. Nous allons présenter les systèmes les plus couramment utilisés pour chaque catégorie.

4.2.1 Systèmes de primaires

Ces systèmes s'appuient sur l'utilisation de trois couleurs fondamentales (les primaires). Ils se différencient par le choix des primaires et du blanc de référence (codage du blanc). Le système le plus

courant est *RGB* (*Red Green Blue*), introduit par la CIE. Les composantes couleur de ce système sont définies par $R \in [R_{min}; R_{max}]$, $G \in [G_{min}; G_{max}]$ et $B \in [B_{min}; B_{max}]$. Pour modéliser certaines couleurs du domaine visible, la composante R doit parfois être négative, cf. [Brun 96, p. 15], or, généralement, les intervalles de variation sont $R \in [0; 255]$, $G \in [0; 255]$ et $B \in [0; 255]$. Ce qui signifie que certaines couleurs ne peuvent pas être modélisées. Il y a autant de systèmes *RGB* que de blancs de référence, cf. [Vandenbroucke 00, p. 195] et [Trémeau 04, p. 82].

Un autre système a été proposé par la CIE, cf. [Cie 15.2 86], il s'agit de *XYZ*. Il est utilisé dans de nombreux articles, notamment dans [Gevers 98, Koschan 93, Sharma 97]. Ce système a été proposé pour prendre en compte le problème des valeurs négatives de la composante R du système *RGB*. Chaque composante de ce système est une combinaison linéaire des composantes *RGB*. Pour chaque système *RGB*, une matrice de passage a donc été définie pour passer des coordonnées R , G et B aux coordonnées X , Y et Z , cf. [Vandenbroucke 00, p. 195–197]. Ce système est rarement utilisé directement car il sert plutôt de système de transition pour passer d'un système *RGB* vers un autre système. Dans le cadre de ce mémoire, nous avons utilisé la transformation la plus couramment employée, cf. [Vandenbroucke 00, p. 196] et [Trémeau 04, p. 90] :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \quad (4.1)$$

Lorsque les intervalles de variation du système *RGB* sont $R \in [0; 255]$, $G \in [0; 255]$ et $B \in [0; 255]$, les intervalles de variation des composantes *XYZ* sont $X \in [0; 250.16]$, $Y \in [0; 255]$ et $Z \in [0; 301.41]$. Dans la suite de ce rapport, nous considérons toujours ces intervalles de variation pour les composantes *RGB*.

Les trois composantes de ces systèmes sont liées à la luminance. Or, deux couleurs peuvent avoir le même caractère chromatique avec des luminances différentes. Pour ne tenir compte que de la chrominance, les coordonnées peuvent être normalisées de la façon suivante, cf. [Garbay 79, p. 24] et [Vandenbroucke 00, p. 52–56] :

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B} \quad \text{et} \quad b = \frac{B}{R+G+B}. \quad (4.2)$$

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z} \quad \text{et} \quad z = \frac{Z}{X+Y+Z}. \quad (4.3)$$

Cette normalisation permet d'obtenir des coordonnées comprises entre 0 et 1 et comme $x + y + z = 1$, la couleur peut être représentée dans un plan. Ces nouvelles coordonnées sont indépendantes des variations scalaires de la luminance de la couleur étudiée. En effet, si chaque composante R , G et B est multipliée par le même scalaire, alors les valeurs de r , g et b ne changent pas. L'inconvénient de cette normalisation est qu'elle n'est pas définie pour $R = G = B = 0$. Quand on se trouve dans ce cas particulier, une solution est de prendre $r = g = b = \frac{1}{3}$, cf. [Garbay 79].

Enfin, le dernier système de primaires, *CMY* (*Cyan Magenta Yellow*), utilisé dans [Gevers 98, Trémeau 03], s'appuie sur une synthèse soustractive et est utilisé pour la télévision et les imprimantes. Une quatrième composante peut être ajoutée (composante « noire »). La transformation suivante peut être utilisée, cf. [Trémeau 04, p. 89] :

$$\begin{pmatrix} C \\ M \\ Y \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \quad (4.4)$$

Les intervalles de variation sont $C \in [0; 510]$, $M \in [0; 510]$ et $Y \in [0; 510]$.

Un des inconvénients d'un système de synthèse additive est qu'il ne permet pas de modéliser certaines couleurs (certaines nuances de vert s'obtiennent en ajoutant des lumières bleues et vertes et en retranchant du rouge, cf. [Brun 96, p. 15] or les couleurs sont représentées avec des valeurs positives). Dans de nombreuses applications en imagerie, comme l'affichage, l'impression, la restauration d'images, etc., il est souhaitable que les propriétés d'un système de représentation de la couleur se rapprochent le plus possible de celles de la vision humaine. Si nous notons D , la distance qui sépare deux couleurs dans un système donné et si nous appelons différence, la différence entre deux couleurs perçue par l'œil humain, un système qui se rapproche d'un système visuel humain doit posséder la propriété (P):

$\forall \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$, trois couleurs telles que la différence de couleur perçue entre \mathbf{c}_1 et \mathbf{c}_3 est la même que celle perçue entre \mathbf{c}_2 et \mathbf{c}_3 alors $D(\mathbf{c}_1, \mathbf{c}_3) = D(\mathbf{c}_2, \mathbf{c}_3)$.

Tous ces systèmes de primaires ne possèdent pas de distance D qui vérifie la propriété (P) et constituent donc une approche limitée pour ces applications. C'est pourquoi d'autres systèmes possédant cette propriété (P) ont été proposés, comme les systèmes luminance-chrominance.

4.2.2 Systèmes luminance-chrominance

Parmi les systèmes luminance-chrominance, quatre types sont différenciés :

- les systèmes perceptuellement uniformes (une distance D vérifiant (P) peut leur être associée) ;
- les systèmes de télévision que nous ne détaillerons pas ici puisqu'ils sont particulièrement dédiés à la télévision et sont largement abordés dans [Vandenbroucke 00, p. 64–67] ;
- les systèmes antagonistes ;
- les autres systèmes.

Tous ces systèmes possèdent une composante de luminance et deux composantes de chrominance.

4.2.2.1 Systèmes perceptuellement uniformes

Parmi les systèmes perceptuellement uniformes, deux systèmes sont proposés par la CIE : CIELUV noté aussi $L^*u^*v^*$ et CIELAB noté aussi $L^*a^*b^*$, cf. [Trémeau 04, p. 95–102]. Ces deux systèmes se déduisent de XYZ par les transformations suivantes :

$$L^* = \begin{cases} 116 \left(\frac{Y}{Y_b} \right)^{\frac{1}{3}} - 16 & \text{si } \frac{Y}{Y_b} > 0.01 \\ 903.3 \frac{Y}{Y_b} & \text{sinon,} \end{cases} \quad (4.5)$$

$$u^* = 13L^*(u' - u'_b) \text{ avec } u' = \frac{4X}{X + 15Y + 3Z}, \quad (4.6)$$

$$v^* = 13L^*(v' - v'_b) \text{ avec } v' = \frac{9Y}{X + 15Y + 3Z}, \quad (4.7)$$

$$a^* = 500 \left(f \left(\frac{X}{X_b} \right) - f \left(\frac{Y}{Y_b} \right) \right), \quad (4.8)$$

$$b^* = 200 \left(f \left(\frac{Y}{Y_b} \right) - f \left(\frac{Z}{Z_b} \right) \right), \quad (4.9)$$

$$\text{avec } f(x) = \begin{cases} x^{\frac{1}{3}} & \text{si } x > 0.008856 \\ 7.787x + \frac{16}{116} & \text{sinon.} \end{cases}$$

Les termes X_b, Y_b, Z_b, u'_b et v'_b sont à associer au blanc de référence. Dans les images que nous utilisons dans ce rapport, le blanc de référence est codé $(255 \ 255 \ 255)^T$, en RGB . Ainsi en utilisant

l'équation (4.1), nous obtenons : $X_b = 250.155$, $Y_w = 255$ et $Z_w = 301.41$. Les intervalles de variation sont donc : $L^* \in [0; 100]$, $u^* \in [-131.95; 220.8]$, $v^* \in [-139.05; 121.47]$, $a^* \in [-137.72; 96.14]$ et $b^* \in [-99.23; 115.65]$.

4.2.2.2 Systèmes antagonistes

Les systèmes antagonistes de Faugeras [Faugeras 79] ont pour but de représenter le mieux possible la perception visuelle humaine. En s'appuyant sur des études physiologiques montrant que l'activité ou la réponse des cônes récepteurs de la rétine humaine, notés LMS (*Long Medium Short*) est proportionnelle non pas à l'intensité du stimulus mais à son logarithme, Faugeras utilise le logarithme des trois signaux L, M et S. Un système semblable s'appuyant sur le système *RGB* a aussi été proposé par Garbay *et al.* [Garbay 81] :

$$A = \frac{1}{3}(\log(R) + \log(G) + \log(B)), \quad (4.10)$$

$$C_1 = \frac{\sqrt{3}}{2}(\log(R) - \log(G)) \text{ et} \quad (4.11)$$

$$C_2 = \log(B) - \frac{\log(R) + \log(G)}{2}. \quad (4.12)$$

Les intervalles de variation sont donc $A \in [0; \log(255)]$, $C_1 \in [-\frac{\sqrt{3}}{2} \log(255); \frac{\sqrt{3}}{2} \log(255)]$ et $C_2 \in [-\log(255); \log(255)]$. Même si la modélisation de la non linéarité de la réponse des cônes de la rétine humaine n'est pas conservée, généralement, le système de Swain et Ballard [Swain 91] est préféré car il n'utilise pas le logarithme :

$$A = \frac{R + G + B}{3}, \quad (4.13)$$

$$C_1 = \frac{\sqrt{3}}{2}(R - G) \text{ et} \quad (4.14)$$

$$C_2 = B - \frac{R + G}{2}. \quad (4.15)$$

Les intervalles de variations sont donc : $A \in [0; 255]$, $C_1 \in [-\frac{\sqrt{3}}{2}255; \frac{\sqrt{3}}{2}255]$ et $C_2 \in [-255; 255]$.

4.2.2.3 Autres systèmes

Enfin, d'autres systèmes semblables ont été proposés par Simonetto et Soadane [Simonetto 05], ou Carron [Carron 95, p. 17]. Nous présentons le système le plus couramment utilisé, celui de Carron :

$$Y = \frac{R + G + B}{3}, \quad (4.16)$$

$$Ch_1 = R - \frac{G + B}{2} \text{ et} \quad (4.17)$$

$$Ch_2 = \frac{\sqrt{3}}{2}(B - G). \quad (4.18)$$

Ainsi, les intervalles de variation sont $Y \in [0; 255]$, $Ch_1 \in [-255; 255]$ et $Ch_2 \in [-\frac{\sqrt{3}}{2}255; \frac{\sqrt{3}}{2}255]$.

4.2.3 Systèmes perceptuels

Les systèmes perceptuels distinguent trois grandeurs : la luminance, la teinte et la saturation. Ils sont notés *HSI* (*Hue Saturation Intensity*). De nombreuses propositions ayant été faites pour le calcul de ces trois composantes, cf. [Hanbury 03, Vandenbroucke 00, p. 74–82], nous ne présentons que les équations les plus utilisées :

$$H = \begin{cases} \pi & \text{si } R = G = B \\ \arccos \frac{\frac{1}{2}((R - G) + (R - B))}{\sqrt{(R - G)^2 + (R - B)(G - B)}} & \text{si } B \leq G \\ 2\pi - \arccos \frac{\frac{1}{2}((R - G) + (R - B))}{\sqrt{(R - G)^2 + (R - B)(G - B)}} & \text{sinon,} \end{cases} \quad (4.19)$$

$$S = \begin{cases} 0 & \text{si } R = G = B = 0 \\ 1 - \frac{3 \min(R, G, B)}{R + G + B} & \text{sinon} \end{cases} \quad \text{et} \quad (4.20)$$

$$I = \frac{R + G + B}{3}. \quad (4.21)$$

Avec ces équations, les intervalles de variation sont $H \in [0; 2\pi]$, $S \in [0; 1]$ et $I \in [0; 255]$.

4.2.4 Systèmes d'axes indépendants

Dans les systèmes de primaires, les trois composantes sont fortement corrélées car elles possèdent toutes une information commune qui est la luminance. Des systèmes d'axes indépendants ont été proposés. Ils s'appuient sur l'utilisation de l'analyse en composantes principales qui permet d'obtenir des composantes décorrélées. Les deux systèmes les plus utilisés sont :

- Le système d'Ohta *et al.* [Ohta 80] dont les composantes sont définies par :

$$I_1 = \frac{R + G + B}{3}, \quad (4.22)$$

$$I_2 = \frac{R - B}{2} \quad \text{et} \quad (4.23)$$

$$I_3 = \frac{2G - R - B}{4}, \quad (4.24)$$

avec $I_1 \in [0; 255]$, $I_2 \in [-\frac{255}{2}; \frac{255}{2}]$ et $I_3 \in [-\frac{255}{2}; \frac{255}{2}]$. Ce système est fondé sur une approximation de la transformation de Karhunen-Loeve, cf. [Horowitz 96, Uenohara 97, Levy 00], où I_1 est la composante la plus discriminante.

- Le système $H_1H_2H_3$ de Braquelaire et Brun [Braquelaire 97] dont les composantes sont définies par :

$$H_1 = R + G, \quad (4.25)$$

$$H_2 = R - G \quad \text{et} \quad (4.26)$$

$$H_3 = B - \frac{R + G}{2}, \quad (4.27)$$

où $H_1 \in [0; 510]$, $H_2 \in [-255; 255]$ et $H_3 \in [-255; 255]$. Levine [Levine 85, chap. 7] montre que les trois directions privilégiées par la vision humaine sont les directions : rouge-vert, bleu-jaune et blanc-noir. Ce système met en évidence ces trois directions privilégiées.

4.2.5 Systèmes hybrides

L'approche hybride de Vandembroucke [Vandembroucke 00, chap. 4] consiste à choisir, pour un ensemble d'images données, parmi toutes les composantes couleur de tous les systèmes de représentation de la couleur, les trois composantes les plus pertinentes pour une application donnée, en utilisant une technique d'apprentissage supervisé. Ces trois composantes forment le système hybride.

4.2.6 Synthèse des systèmes de représentation de la couleur

Parmi les systèmes qui viennent d'être brièvement présentés, nous avons testé les dix suivants (cf. tableau 4.1): *RGB*, *XYZ*, *CMY*, $L^*u^*v^*$, $L^*a^*b^*$, AC_1C_2 , YCh_1Ch_2 , *HSI*, $I_1I_2I_3$ et $H_1H_2H_3$. Ce sont les plus représentatifs et les plus utilisés en traitement d'images. La remarque qui peut être faite dès à présent est la suivante: les modèles luminance-chrominance $L^*a^*b^*$ et $L^*u^*v^*$, et le modèle perceptuel *HSI* se rapprochent le plus de la vision humaine mais il reste à déterminer si cela a une influence sur les résultats obtenus au niveau de la mise en correspondance. De plus, le passage vers ces modèles est coûteux en temps de calcul.

À présent, nous allons présenter la généralisation de la mise en correspondance par corrélation à la couleur. Tout d'abord, nous abordons la généralisation des mesures de corrélation à la couleur que nous désignons par « mesures 3D » et ensuite nous exposons les trois méthodes de mise en correspondance proposées.

4.3 Généralisation des mesures de corrélation à la couleur

4.3.1 Notations

Dans la suite, nous notons x , y et z , les trois composantes de n'importe quel espace de couleur. Nous utilisons aussi les notations suivantes :

- Les vecteurs $\mathbf{c}_l^{i,j} = (x_l^{i,j} \ y_l^{i,j} \ z_l^{i,j})^T$, sont les couleurs des pixels de coordonnées $(i \ j)^T$ dans I_l .
- Les matrices \mathbf{F}_l de taille $N \times 3$, contiennent les composantes des couleurs des pixels dans les fenêtres de corrélation gauche et droite :

$$\mathbf{F}_l = (\dots \ \mathbf{c}_l^{i+p,j+q} \ \dots)^T = (\dots \ \mathbf{c}_l^k \ \dots)^T,$$

où \mathbf{c}_l^k est la couleur d'indice $k \in [0; N_f - 1]$, $p \in [-N_v; N_v]$ et $q \in [-N_h; N_h]$.

- La matrice ∇I_l^k , contient les vecteurs gradients de chaque composante, au pixel k de la fenêtre de corrélation dans I_l :

$$\nabla I_l^k = \begin{pmatrix} \frac{\partial x_l^k}{\partial i} & \frac{\partial y_l^k}{\partial i} & \frac{\partial z_l^k}{\partial i} \\ \frac{\partial x_l^k}{\partial j} & \frac{\partial y_l^k}{\partial j} & \frac{\partial z_l^k}{\partial j} \end{pmatrix}^T. \quad (4.28)$$

Les termes $\frac{\partial x_l^k}{\partial i}$, $\frac{\partial y_l^k}{\partial i}$, $\frac{\partial z_l^k}{\partial i}$, $\frac{\partial x_l^k}{\partial j}$, $\frac{\partial y_l^k}{\partial j}$ et $\frac{\partial z_l^k}{\partial j}$ représentent les dérivées partielles suivant les lignes et les colonnes de chaque composante. La norme et l'orientation du gradient sont notées $\|\nabla I_l^k\|$ et θ_l^k et la manière de les calculer est abordée au paragraphe 4.3.2.3.

4.3.2 Généralisation d'opérateurs

L'adaptation de mesures s'appuyant sur des données 1D, les niveaux de gris, à des données 3D, les couleurs, demande la généralisation de plusieurs opérateurs : calcul de la différence entre deux couleurs, convolution d'une image par un filtre et calcul des rangs d'un ensemble de données tridimensionnelles.

RÉFÉRENCE	TYPE	DÉFINITION DES COMPOSANTES	INTERVALLES
XYZ [Cie 15.2 86]	Système de primaires	$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	$X \in [0; 250.16]$ $Y \in [0; 255]$ $Z \in [0; 301.41]$
CMY [Trémeau 04, p. 89]		$\begin{pmatrix} C \\ M \\ Y \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	$C \in [0; 510]$ $M \in [0; 510]$ $Y \in [0; 510]$
$L^*u^*v^*$ [Trémeau 04, p. 95–102]	Système perceptuelle-ment uniforme	$L^* = \begin{cases} 116 \left(\frac{Y}{Y_b}\right)^{\frac{1}{3}} - 16 & \text{si } \frac{Y}{Y_b} > 0.01 \\ 903.3 \frac{Y}{Y_b} & \text{sinon} \end{cases}$	$L^* \in [0; 100]$ $u^* \in [-131.95; 220.8]$ $v^* \in [-139.05; 121.47]$
$L^*a^*b^*$ [Trémeau 04, p. 95–102]		$u^* = 13L^*(u' - u'_b)$ avec $u' = \frac{4X}{X+15Y+3Z}$ $v^* = 13L^*(v' - v'_b)$ avec $v' = \frac{9Y}{X+15Y+3Z}$ $a^* = 500 \left(f\left(\frac{X}{X_b}\right) - f\left(\frac{Y}{Y_b}\right) \right)$ et $b^* = 200 \left(f\left(\frac{Y}{Y_b}\right) - f\left(\frac{Z}{Z_b}\right) \right)$ avec $f(x) = \begin{cases} x^{\frac{1}{3}} & \text{si } x > 0.008856 \\ 7.787x + \frac{16}{116} & \text{sinon} \end{cases}$	$L^* \in [0; 100]$ $a^* \in [-137.72; 96.14]$ $b^* \in [-99.23; 115.65]$
AC_1C_2 [Swain 91]	Système antagoniste	$\begin{pmatrix} A \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	$A \in [0; 255]$ $C_1 \in \left[-\frac{\sqrt{3}}{2}255; \frac{\sqrt{3}}{2}255\right]$ $C_2 \in [-255; 255]$
YCh_1Ch_2 [Carron 95, p. 17]	Système luminance-chrominance	$\begin{pmatrix} Y \\ Ch_1 \\ Ch_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	$Y \in [0; 255]$ $Ch_1 \in [-255; 255]$ $Ch_2 \in \left[-\frac{255\sqrt{3}}{2}; \frac{255\sqrt{3}}{2}\right]$
HSI [Hanbury 03]	Système perceptuel	$H_1 = \left(\frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right)$ $H = \begin{cases} \pi & \text{si } R = G = B \\ \arccos H_1 & \text{si } B \leq G \\ 2\pi - \arccos H_1 & \text{sinon} \end{cases}$ $S = \begin{cases} 0 & \text{si } R = G = B = 0 \\ 1 - \frac{\min(R,G,B)}{R+G+B} & \text{sinon} \end{cases}$ $I = \frac{R+G+B}{3}$	$I \in [0; 255]$ $S \in [0; 1]$ $H \in [0; 2\pi]$
$I_1I_2I_3$ [Ohta 80]	Système d'axes	$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	$I_1 \in [0; 255]$ $I_2 \in [-127.5; 127.5]$ $I_3 \in [-127.5; 127.5]$
$H_1H_2H_3$ [Braquelaire 97]	indépendants	$\begin{pmatrix} H_1 \\ H_2 \\ H_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -\frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	$H_1 \in [0; 510]$ $H_2 \in [-255; 255]$ $H_3 \in [-255; 255]$

TAB. 4.1 – Dix systèmes testés (en comptant RGB).

4.3.2.1 Opérateurs de base

- Les normes L_P avec $P \in \mathbb{N}^*$ données par l'équation (3.1) deviennent :

$$\|\mathbf{F}_l\|_P = \left(\sum_{k=0}^{N_f-1} \|\mathbf{c}_l^k\|_P^P \right)^{\frac{1}{P}} \quad (4.29)$$

avec $\|\mathbf{c}_l^k\|_P^P = (x_l^k)^P + (y_l^k)^P + (z_l^k)^P$.

La norme de Frobenius est notée $\|\mathbf{F}_l\| = \|\mathbf{F}_l\|_2$.

- Le produit scalaire défini par l'équation (3.2) devient :

$$\mathbf{F}_g \cdot \mathbf{F}_d = \sum_{k=0}^{N_f-1} x_g^k x_d^k + y_g^k y_d^k + z_g^k z_d^k. \quad (4.30)$$

- Les vecteurs des moyennes définis par l'équation (3.4) deviennent les matrices des moyennes de chaque composante :

$$\overline{\mathbf{F}}_l = \frac{1}{N_f} \left(\underbrace{1 \cdots 1}_{N_f \text{ colonnes}} \right)^T \begin{pmatrix} \sum_{k=0}^{N_f-1} x_l^k & \sum_{k=0}^{N_f-1} y_l^k & \sum_{k=0}^{N_f-1} z_l^k \end{pmatrix}. \quad (4.31)$$

- Dans certains cas, une autre définition de la moyenne d'une matrice, $\mathbf{A}_{P \times Q}$, est utilisée :

$$\text{moy}(\mathbf{A}) = \frac{1}{PQ} \sum_{i=0}^{P-1} \sum_{j=0}^{Q-1} A^{i,j}. \quad (4.32)$$

- En généralisant les mesures, les divisions entre scalaires sont parfois remplacées par des divisions entre matrices. Dans ce cas, la division d'Hadamard est utilisée :

$$\mathbf{F}_g \oslash \mathbf{F}_d = \begin{pmatrix} \frac{x_g^1}{x_d^1} & \frac{y_g^1}{y_d^1} & \frac{z_g^1}{z_d^1} \\ \vdots & \vdots & \vdots \\ \frac{x_g^{N_f}}{x_d^{N_f}} & \frac{y_g^{N_f}}{y_d^{N_f}} & \frac{z_g^{N_f}}{z_d^{N_f}} \end{pmatrix}. \quad (4.33)$$

4.3.2.2 Distance entre deux couleurs

Pour tous les systèmes de primaires, il n'existe pas de distance D qui vérifie la propriété (P) (cf. paragraphe 4.2.1). Il a souvent été proposé, notamment dans [Crouzil 97, p. 57], d'utiliser une distance euclidienne. D'autres distances ont été proposées par Koschan [Koschan 93] et nous les généralisons par une norme L_P :

$$D_P(\mathbf{c}_g, \mathbf{c}_d) = ((x_g - x_d)^P + (y_g - y_d)^P + (z_g - z_d)^P)^{\frac{1}{P}}. \quad (4.34)$$

Pour les systèmes $L^*a^*b^*$ et $L^*u^*v^*$, la distance euclidienne est la plus utilisée, cf. [Brun 96, p. 17], [Vandenbroucke 00, p. 63–64] et [Rital 04]. Elle est adaptée à ces deux systèmes et elle respecte la propriété (P).

Concernant les systèmes perceptuels, la distance entre deux couleurs $\mathbf{c}_g = (H_g \ S_g \ I_g)^T$ et $\mathbf{c}_d = (H_d \ S_d \ I_d)^T$ est définie par Koschan [Koschan 93] par :

$$D_{HSI}(\mathbf{c}_g, \mathbf{c}_d) = \sqrt{(I_g - I_d)^2 + S_g^2 + S_d^2 - 2S_g S_d \cos\theta} \quad (4.35)$$

$$\theta = \begin{cases} |H_g - H_d| & \text{si } |H_g - H_d| \leq \pi \\ 2\pi - |H_g - H_d| & \text{sinon.} \end{cases}$$

Pour tous les autres systèmes, une distance utilisant une norme L_P , cf. équation (4.34), peut être utilisée. Ainsi, lorsque nous parlerons de distance D_* entre deux couleurs, nous utiliserons l'équation (4.34) pour tous les systèmes, excepté *HSI*, pour lequel nous utiliserons l'équation (4.35).

Le vecteur $\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)$ contient toutes les distances entre les couleurs contenues dans \mathbf{F}_g et \mathbf{F}_d :

$$\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d) = (\dots \ D_*(\mathbf{c}_g^i, \mathbf{c}_d^i) \ \dots)^T. \quad (4.36)$$

4.3.2.3 Calcul des vecteurs gradients d'une image couleur

Pour toutes les mesures utilisant les dérivées, il faut appliquer un filtre à l'image avant de calculer la mesure de corrélation classique. Il nous faut donc adapter ces filtres à des données tridimensionnelles. Des travaux ont déjà été réalisés, cf. [Carron 95, chap. 1.2], [Chanussot 98, chap. A.IV], [Lambert 02, chap. 3] et [Trémeau 04, chap. 5] et les techniques employées vont être présentées.

Le résultat $I_l^{i,j}$ de la convolution d'une image I_l par un masque \mathbf{H} , matrice de taille $(2N_v + 1)(2N_h + 1)$, s'évalue en chaque pixel de l'image par :

$$I_l^{i,j} = \sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} h^{N_v-p, N_h-q} I_l^{i+p, j+q}, \text{ avec } h^{i,j} \text{ les éléments de la matrice } \mathbf{H}. \quad (4.37)$$

Une première approche (approche marginale) consiste à effectuer la convolution sur chaque image de chaque composante, cf. [Lezoray 03, Levesque 04]. Dans ce cas, la corrélation qui existe entre chaque composante n'est pas prise en compte pour calculer le gradient. Une seconde approche (approche vectorielle) consiste à prendre en compte les trois composantes couleur pour filtrer l'image. Pour cela il faut adapter le filtre \mathbf{H} et l'équation (4.37) devient :

$$\mathbf{c}_l^{i,j} = \sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} \mathbf{h}^{N_v-p, N_h-q} \mathbf{c}_l^{i+p, j+q}. \quad (4.38)$$

Il faut donc trouver les valeurs des matrices $\mathbf{h}^{i,j}$ en réalisant une extension vectorielle du filtre \mathbf{H} tout en conservant ses propriétés. La mise en œuvre de la seconde approche est bien plus délicate que la première approche et n'est pratiquement pas utilisée, cf. [Chanussot 98, chap. A.IV] et [Lambert 02, chap. 3]. Dans ce mémoire, l'approche marginale sera donc préférée.

Il reste cependant le problème du calcul de la norme et de l'orientation du vecteur gradient de l'image. Il faut établir la fusion des trois vecteurs gradients obtenus. Deux possibilités ont déjà été envisagées, cf. [Di Zenzo 86, Lee 91] et [Crouzil 97, 61–65] :

- effectuer la somme des vecteurs gradients ;
- calculer la norme et l'orientation du gradient multi-composantes selon la méthode de Lee et

Cok [Lee 91] grâce aux équations suivantes :

$$\|\nabla I_l^k\|^2 = \frac{1}{2} \left(p + q + \sqrt{(p - q)^2 + 4t^2} \right), \quad (4.39)$$

$$\theta_l^k = \begin{cases} \arctan \left(\frac{t}{\|\nabla I_l^k\|^2 - q} \right) & \text{si } t = 0 \text{ ou } \|\nabla I_l^k\|^2 - p = 0 \\ \arctan \left(\frac{\|\nabla I_l^k\|^2 - p}{t} \right) & \text{sinon,} \end{cases} \quad (4.40)$$

$$p = \left(\frac{\partial x_l^k}{\partial i} \right)^2 + \left(\frac{\partial y_l^k}{\partial i} \right)^2 + \left(\frac{\partial z_l^k}{\partial i} \right)^2,$$

$$q = \left(\frac{\partial x_l^k}{\partial j} \right)^2 + \left(\frac{\partial y_l^k}{\partial j} \right)^2 + \left(\frac{\partial z_l^k}{\partial j} \right)^2 \text{ et}$$

$$t = \frac{\partial x_l^k}{\partial i} \frac{\partial x_l^k}{\partial j} + \frac{\partial y_l^k}{\partial i} \frac{\partial y_l^k}{\partial j} + \frac{\partial z_l^k}{\partial i} \frac{\partial z_l^k}{\partial j}.$$

La première méthode peut entraîner une perte d'information mais elle est beaucoup plus simple et moins coûteuse. Dans cette étude, les deux méthodes seront testées.

4.3.2.4 Rangs de données tridimensionnelles

Certaines mesures que nous avons présentées au chapitre 3 ne s'appuient pas sur les niveaux de gris mais sur l'ordre des niveaux de gris. Pour adapter ces mesures, il nous faut donc pouvoir comparer les couleurs et les ordonner. Pour calculer les rangs de données tridimensionnelles, plusieurs possibilités sont envisageables :

- *Ordre vectoriel par entrelacement de bits* – Pour Chanussot [Chanussot 98, p. 32–37], il s'agit de construire une chaîne de bits à partir des N_b bits codant chaque composante couleur. Soient $\{X_k\}$, $\{Y_k\}$ et $\{Z_k\}$, $k = 0..N_b - 1$, les codes des trois composantes, alors la chaîne est obtenue en entrelaçant les bits de la manière suivante :

$$X_0 Y_0 Z_0 X_1 Y_1 Z_1 \dots X_{N_b-1} Y_{N_b-1} Z_{N_b-1}.$$

À partir de cette chaîne, les vecteurs sont ordonnés.

- *Ordre selon une composante* – Lambert [Lambert 02, p. 31–32] propose d'utiliser seulement une des composantes couleur. Celle-ci est donc favorisée.
- *Ordre lexicographique* – Lambert [Lambert 02, p. 31–32] propose une autre possibilité qui consiste à utiliser une des trois composantes pour effectuer le tri. Lorsqu'il y a des valeurs égales, une autre composante est prise en compte. Cette méthode favorise une composante en particulier.
- *Ordre par mesure d'une distance cumulée* – Lambert [Lambert 02, p. 34] s'appuie sur une distance calculée entre tous les couples de couleurs de l'ensemble. Pour chaque couleur, la somme des distances avec les autres couleurs est effectuée. Cette somme sert à classer les couleurs. Cette méthode est très coûteuse.
- *Ordre sur la première composante d'une ACP* – Cheng et Hsia [Cheng 03] proposent d'effectuer le tri des données en ne considérant que la première composante issue d'une ACP, cf. paragraphe 4.4.2.

4.3.3 Mesures 3D

4.3.3.1 Corrélation croisée couleur

La généralisation des mesures NCC et ZNCC a déjà été abordée dans [Qian 02, Tsai 03, Crouzil 97, p. 59–60] et le même modèle a été suivi pour la mesure de Moravec. Cette généralisation s'appuie sur les équations (4.29), (4.30) et (4.31) (cf. tableau 4.2).

NOM	NOTATION	DÉFINITION	INT.
Corrélation croisée normalisée	NCC	$\frac{\mathbf{F}_g \cdot \mathbf{F}_d}{\ \mathbf{F}_g\ \ \mathbf{F}_d\ }$	[0; 1]
Corrélation croisée centrée normalisée	ZNCC	$\frac{(\mathbf{F}_g - \overline{\mathbf{F}}_g) \cdot (\mathbf{F}_d - \overline{\mathbf{F}}_d)}{\ \mathbf{F}_g - \overline{\mathbf{F}}_g\ \ \mathbf{F}_d - \overline{\mathbf{F}}_d\ }$	[-1; 1]
Mesure de Moravec	MOR	$\frac{2(\mathbf{F}_g - \overline{\mathbf{F}}_g) \cdot (\mathbf{F}_d - \overline{\mathbf{F}}_d)}{\ \mathbf{F}_g - \overline{\mathbf{F}}_g\ ^2 + \ \mathbf{F}_d - \overline{\mathbf{F}}_d\ ^2}$	[-1; 1]

TAB. 4.2 – Famille CROISÉE couleur – Les mesures de cette famille exploitent la corrélation croisée couleur.

4.3.3.2 Statistiques de la distribution des différences de couleur

Le modèle suivi pour les mesures SAD et SSD, comme dans [Crouzil 97, p. 58], et par les normes L_P , cf. [Stricker 95, Coquin 02], a été utilisé. Les équations (4.29), (4.31), (4.33) et (4.36) sont nécessaires pour cette généralisation (cf. tableau 4.3).

NOM	NOTATION	DÉFINITION	INT.
Distances	D_P	$\ \mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)\ _{P^P}$	$[0; D_{\max}^P N_f]$
Distances normalisées	ND_P	$\frac{D_P(\mathbf{F}_g, \mathbf{F}_d)}{\sqrt{\ \mathbf{F}_g\ _P \ \mathbf{F}_d\ _P}}$	$[0; D_{\max}^P N_f]$
Distances centrées	ZD_P	$D_P(\mathbf{F}_g - \overline{\mathbf{F}}_g, \mathbf{F}_d - \overline{\mathbf{F}}_d)$	$[0; D_{\max}^P N_f]$
Distances centrées et normalisées	ZND_P	$ND_P(\mathbf{F}_g - \overline{\mathbf{F}}_g, \mathbf{F}_d - \overline{\mathbf{F}}_d)$	$[0; D_{\max}^P N_f]$
Distances localement centrées	LD_P	$\ \mathbf{D}_*(\mathbf{F}_g, (\overline{\mathbf{F}}_g \circledast \overline{\mathbf{F}}_d))\mathbf{F}_d\ _{P^P}$	$[0; +\infty[$
Variance des différences	VD_P	$\text{var}(\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d))$	$[0; D_{\max}^{2P}]$
Variance des valeurs absolues des différences	VAD_P	$\text{var}(\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d) ^P)$	$[0; D_{\max}^{2P}]$
Kurtosis	K_4	$ \overline{\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)^4} - 3\overline{\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)^2} $	$[0; D_{\max}^4]$

TAB. 4.3 – Famille CLASSIQUE couleur – Les mesures de cette famille utilisent des outils des statistiques classiques de la distribution des différences des couleurs. Le terme D_{\max}^P correspond à la distance maximale entre deux couleurs (cette distance dépend de la norme L_P utilisée).

4.3.3.3 Mesures utilisant les dérivées des images couleur

Les mesures présentées dans cette famille (cf. tableau 4.4) utilisent les opérateurs de Sobel ($R_{Sobel}(\mathbf{F}_l)$), de Kirsch ($R_{Kirsch}(\mathbf{F}_l)$), du laplacien de gaussien ($R_{LoG}(\mathbf{F}_l)$), de Roberts ($R_{Roberts}(\mathbf{F}_l)$), de Pratt ($R_{Pratt}(\mathbf{F}_l)$), ou encore le filtrage d'Ullah ($R_{OCM}(\mathbf{F}_l)$). Ces filtres sont appliqués respectivement sur chaque canal. En ce qui concerne la mesure GC, pour le calcul de la norme et de l'orientation du gradient, les deux méthodes décrites dans le paragraphe 4.3.2.3 sont utilisées. La généralisation de ces mesures exploite les équations (4.29), (4.30), (4.31) et (4.36).

NOM	NOTATION	DÉFINITION	INT.
Seitz	SEO	$\ \mathbf{D}_*(\mathbf{R}_O(\mathbf{F}_g), \mathbf{R}_O(\mathbf{F}_d))\ _{L^P}^P$	$[0; \mathbf{D}_{\max}^P N_f]$
Nishihara	NIS	$\mathbf{R}_{LoG}(\mathbf{F}_g) \cdot \mathbf{R}_{LoG}(\mathbf{F}_d)$	$[0; N_f]$
Nack 1	NA ₁	$\frac{\mathbf{R}_{Roberts}(\mathbf{F}_g) \cdot \mathbf{R}_{Roberts}(\mathbf{F}_d)}{N_f \text{ moy}(\mathbf{R}_{Roberts}(\mathbf{F}_d))}$	$[0; 1]$
Nack 2	NA ₂	$\frac{\mathbf{NA}_1(\mathbf{F}_g, \mathbf{F}_d)}{N_f \text{ moy}(\mathbf{R}_{Roberts}(\mathbf{F}_g)) - \mathbf{R}_{Roberts}(\mathbf{F}_g) \cdot \mathbf{R}_{Roberts}(\mathbf{F}_d)}$	$[0; 1]$
Pratt	PRATT	$\text{ZNCC}(\mathbf{R}_{Pratt}(\mathbf{F}_g), \mathbf{R}_{Pratt}(\mathbf{F}_d))$	$[-1; 1]$
Orientation code matching	OCM	$\frac{1}{N_f} \text{DOCM}(\mathbf{R}_{OCM}(\mathbf{F}_g), \mathbf{R}_{OCM}(\mathbf{F}_d))$	$[0; \frac{N_{OCM}}{2}]$
Corrélation des champs de gradients	GC	$\frac{\sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} \ \nabla I_g^{i+p, j+q} - \nabla I_d^{i+p, v+q}\ }{\sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} (\ \nabla I_g^{i+p, j+q}\ + \ \nabla I_d^{i+p, v+q}\)}$	$[0; \infty[$

TAB. 4.4 – Famille DÉRIVÉE couleur – Les mesures de cette famille s'appuient sur les dérivées des images couleur. Le terme \mathbf{D}_{\max}^P correspond à la distance maximale entre deux couleurs (cette distance dépend de la norme L_P utilisée).

NOM	NOTATION	DÉFINITION	INT.
Mesure du χ^2	χ^2	$\sum_{k=0}^{N_f-1} \frac{\mathbf{D}_*(\mathbf{c}_g^k, \mathbf{c}_d^k)^2}{4\mathbf{c}_g^k + \mathbf{c}_d^k}$	$[0; \mathbf{D}_{\max}^P N_f]$
Mesure de Jeffrey	JEFF	$\sum_{k=0}^{N_f-1} \left(\overline{\mathbf{c}_g^k} \log \frac{\overline{\mathbf{c}_g^k}}{\mathbf{c}_g^k + \mathbf{c}_d^k} + \overline{\mathbf{c}_d^k} \log \frac{\overline{\mathbf{c}_d^k}}{\mathbf{c}_g^k + \mathbf{c}_d^k} \right)$	$[0; \mathbf{D}_{\max}^P N_f]$
Increment sign correlation	ISC	$\frac{1}{N_f - 1} (\mathbf{b}_g \cdot \mathbf{b}_d + (\mathbf{1} - \mathbf{b}_g) \cdot (\mathbf{1} - \mathbf{b}_d))$	$[0; 1]$
Selective Coefficient Correlation	SCC	$\frac{\mathbf{E}(\mathbf{F}_g - \overline{\mathbf{F}_g}) \cdot (\mathbf{F}_d - \overline{\mathbf{F}_d})}{\ \mathbf{E}(\mathbf{F}_g - \overline{\mathbf{F}_g})\ \ \mathbf{E}(\mathbf{F}_d - \overline{\mathbf{F}_d})\ }$	$[0; 1]$
Rang	RANK _P	$\ \mathbf{R}_{rank}(\mathbf{F}_g) - \mathbf{R}_{rank}(\mathbf{F}_d)\ _{L^P}^P$	$[0; N_f^{P+1}]$
Recensement	CENSUS	$\mathbf{D}_{Ham}(\mathbf{R}_\tau(\mathbf{F}_g), \mathbf{R}_\tau(\mathbf{F}_d))$	$[0; N_f]$
Mesures ordinales	α	$1 - \frac{2\mathbf{D}_\alpha(\mathbf{R}_\pi(\mathbf{F}_g), \mathbf{R}_\pi(\mathbf{F}_g))}{\mathbf{D}_{\max}}$	$[-1; 1]$

TAB. 4.5 – Famille NON PARAMÉTRIQUE couleur – Les mesures de cette famille utilisent des transformations non paramétriques des images couleur. Le terme \mathbf{D}_{\max}^P correspond à la distance maximale entre deux couleurs (cette distance dépend de la norme L_P utilisée).

4.3.3.4 Mesures non paramétriques couleur

Pour les mesures ordinales de cette famille (cf. tableau 4.5), la solution envisagée est de calculer les rangs sur les données tridimensionnelles puis d'évaluer la mesure. Les opérateurs décrits dans le paragraphe 4.3.2.4 sont utilisés : ACP, ordre suivant une composante, ordre lexicographique, ordre

vectorel par entrelacement de bits. Toutes les transformations non paramétriques sont exposées dans le paragraphe 3.3.4. Les équations (4.29) et (4.30) sont aussi utilisées pour adapter ces mesures.

4.3.3.5 Mesures robustes couleur

Pour les mesures de Lan [Lan 97, chap. 3], il faudrait adapter l'estimation de la moindre médiane des carrés et de l'ellipse de volume minimal à des données de dimension 6. Cette généralisation semble délicate et surtout très coûteuse en temps de calcul. En effet, la corrélation 1D seule est déjà très coûteuse en temps de calcul, cf. résultats du paragraphe 3.6.2.2. C'est pourquoi, pour l'instant, nous ne proposons pas de méthode pour généraliser ces mesures. Pour le calcul des rangs et de la médiane, les opérateurs proposés dans le paragraphe 4.3.2.4 sont utilisés. Toutes les mesures robustes que nous avons proposées dans [Chambon 03] sont fondées sur le calcul des différences entre chaque pixel des fenêtres de corrélation et l'utilisation des rangs et de la médiane.

NOM	NOTATION	DÉFINITION	INT.
Corrélation partielle	RM_m	$RM_m(\mathbf{F}_g, \mathbf{F}_d)$	$[-1; 1]$
<i>Quadrant correlation</i>	QUAD	$ZNCC(R_{quad}(\mathbf{F}_g), R_{quad}(\mathbf{F}_d))$	$[-1; 1]$
Pseudo-norme	D_P	$\ \mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)\ _{P^P}$ avec $0 < P < 1$	$[0; D_{\max}^P N_f]$
Variance robuste	MAD	$\text{med} \mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d) - \text{med}(\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)) $	$[0; D_{\max}^P]$
Moindre médiane des puissances	LMP_P	$\text{med} \left((\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d))^P \right)$	$[0; D_{\max}^P]$
Moindres puissances tronquées	LTP_P	$\sum_{k=0}^{h-1} (\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d))_{k:N_f-1}$	$[0; D_{\max}^P h]$
<i>Smooth Median Powered Deviation</i>	$SMPD_P$	$\sum_{k=0}^{h-1} (\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d) - \text{med}(\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)) ^P)_{k:N_f-1}$	$[0; D_{\max}^P h]$
M-estimateur	MR_m	$\sum_{k=0}^{N_f-1} \rho_m(\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d))$	$[\rho_m^{\min} N_f; \rho_m^{\max} N_f]$
R-estimateur	R_l	$\sum_{k=0}^{N_f-1} J_m \left(\frac{R_\pi(\mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d))}{N_f - 1} \right) \mathbf{D}_*(\mathbf{F}_g, \mathbf{F}_d)$	$[J_m^{\min} D_{\max}^P N_f; J_m^{\max} D_{\max}^P N_f]$

TAB. 4.6 – Famille ROBUSTE couleur – Les mesures de cette famille prennent en compte le problème des occultations. Le terme D_{\max}^P correspond à la distance maximale entre deux couleurs (cette distance dépend de la norme L_P utilisée). Les termes ρ_m^{\min} , ρ_m^{\max} et J_m^{\min} , J_m^{\max} correspondent respectivement aux valeurs inférieures et supérieures des fonctions ρ_m et J_m sur les intervalles de variations des trois composantes couleur.

Ainsi, pour certaines mesures, deux solutions sont envisageables :

- Pour la mesure $SMPD_P$ –
 - La première solution consiste à évaluer le vecteur des distances entre les couleurs puis la médiane de ce vecteur et enfin la somme des h premières différences à la puissance P .
 - La seconde solution se fonde sur l'évaluation de la médiane et des différences à la puissance P directement sur les données tridimensionnelles.
- Pour les mesures utilisant les R-estimateurs –
 - La première solution est de calculer le vecteur des distances entre les couleurs puis les rangs des données de ce vecteur.

- La seconde solution consiste à appliquer directement le calcul des rangs aux différences de couleur.

Pour l'adaptation de toutes ces mesures, l'équation (4.36) est nécessaire (cf. tableau 4.6).

4.4 Généralisation des méthodes de mise en correspondance à la couleur

Dans ce paragraphe, nous allons décrire les trois possibilités envisagées pour généraliser la mise en correspondance par corrélation à la couleur :

- *Mise en correspondance sur chaque composante suivi d'une fusion* – La fusion peut avoir lieu au niveau du calcul du score de corrélation ou au niveau des trois disparités obtenues.
- *Mise en correspondance à partir de la première composante issue d'une analyse en composante principale* – L'ACP peut être effectuée une seule fois sur toute l'image ou à chaque calcul du score de corrélation, sur la fenêtre de corrélation.
- *Mise en correspondance 3D* – Les mesures de corrélation 3D que nous avons présentées au paragraphe 4.3.3 sont utilisées.

4.4.1 Mise en correspondance sur chaque composante et fusion

4.4.1.1 Fusion des scores

Les méthodes les plus souvent employées sont le calcul du score de corrélation sur une des composantes, cf. [Garcia 01c], le calcul de la moyenne des scores de corrélation sur chaque composante, cf. [Mayer 03] et la fusion des scores de corrélation sur chaque composante, cf. [Belli 00]. Dans la méthode de Belli *et al.* [Belli 00], la fusion est effectuée en utilisant un « barycentre pondéré » déjà employé dans le cadre de la fusion de données en analyse d'images par Lefèvre *et al.* [Lefèvre 96]. Nous généralisons toutes ces approches en proposant la méthode FUSION-SCORE, cf. algorithme 4.1 et figure 4.1. Son principe consiste à appliquer la mesure de corrélation sur chaque composante puis à fusionner les résultats obtenus :

$$\text{Mes}_{\text{couleur}}(\mathbf{F}_g, \mathbf{F}_d) = \text{fusion}(\text{Mes}_{\text{gris}}(\mathbf{x}_g, \mathbf{x}_d), \text{Mes}_{\text{gris}}(\mathbf{y}_g, \mathbf{y}_d), \text{Mes}_{\text{gris}}(\mathbf{z}_g, \mathbf{z}_d)) \quad (4.41)$$

avec $\text{fusion} \in \{\text{moyenne, maximum, minimum, médiane, barycentre pondéré}\}$.

Les termes $\text{Mes}_{\text{couleur}}$ et Mes_{gris} correspondent respectivement à la mesure de corrélation 3D (couleur) et la mesure de corrélation 1D (niveaux de gris). Les vecteurs \mathbf{x}_l , \mathbf{y}_l et \mathbf{z}_l contiennent respectivement toutes les composantes x_l^k , y_l^k et z_l^k de la fenêtre de corrélation. La fusion peut être réalisée en calculant la moyenne, le maximum, le minimum ou la médiane des trois scores de corrélation obtenus. La formule du barycentre pondéré, qui a été proposée dans [Belli 00] dans le cas de l'utilisation de la corrélation croisée, est la suivante :

$$\text{Mes}_{\text{couleur}}(\mathbf{F}_g, \mathbf{F}_d) = \frac{\text{Mes}_{\text{gris}}(\mathbf{x}_g, \mathbf{x}_d)^2 + \text{Mes}_{\text{gris}}(\mathbf{y}_g, \mathbf{y}_d)^2 + \text{Mes}_{\text{gris}}(\mathbf{z}_g, \mathbf{z}_d)^2}{\text{Mes}_{\text{gris}}(\mathbf{x}_g, \mathbf{x}_d) + \text{Mes}_{\text{gris}}(\mathbf{y}_g, \mathbf{y}_d) + \text{Mes}_{\text{gris}}(\mathbf{z}_g, \mathbf{z}_d)}. \quad (4.42)$$

Pour chaque passage faire

2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$

2.3 Pour chaque $\mathcal{S}^{i,j}$ faire

A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire

A.1. Calculer $\text{Mes}_{\text{gris}}(\mathbf{x}_g, \mathbf{x}_d)$, $\text{Mes}_{\text{gris}}(\mathbf{y}_g, \mathbf{y}_d)$ et $\text{Mes}_{\text{gris}}(\mathbf{z}_g, \mathbf{z}_d)$

A.2. Fusionner les trois scores de corrélation

F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA

2.4 Si passage 1 alors appliquer la contrainte de symétrie puis calculer les disparités au sous-pixel

Les différents passages :

Passage 0 -- Mise en correspondance gauche -> droite

Passage 1 -- Mise en correspondance droite -> gauche

ALGO. 4.1 – Méthode FUSION-SCORE – Cet algorithme est une instance de l’algorithme 1.2. L’utilisation de la couleur en fusionnant les scores de corrélation 1D est effectuée à l’étape A. Il n’y a pas de prétraitement (disparition de l’étape 2.1).

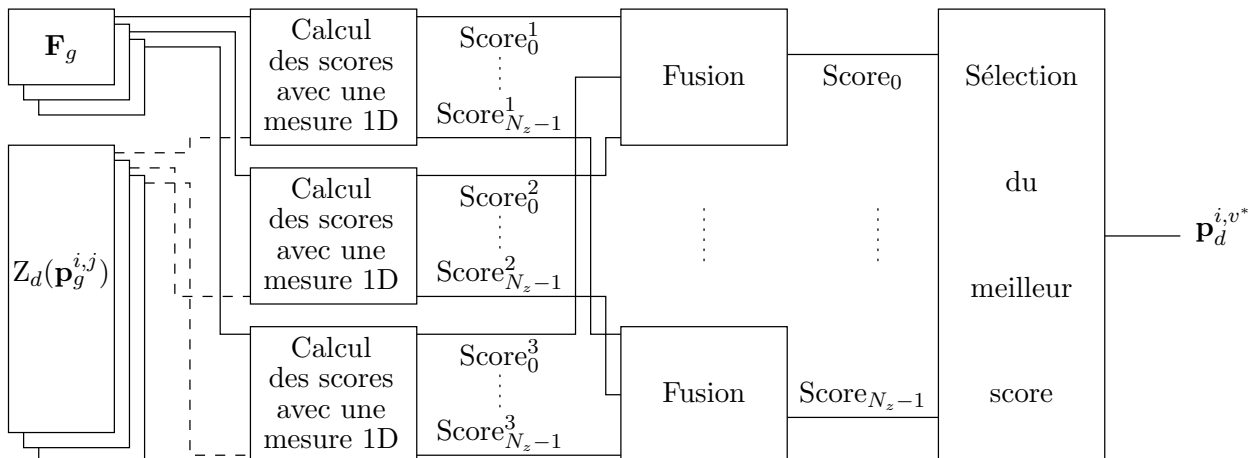
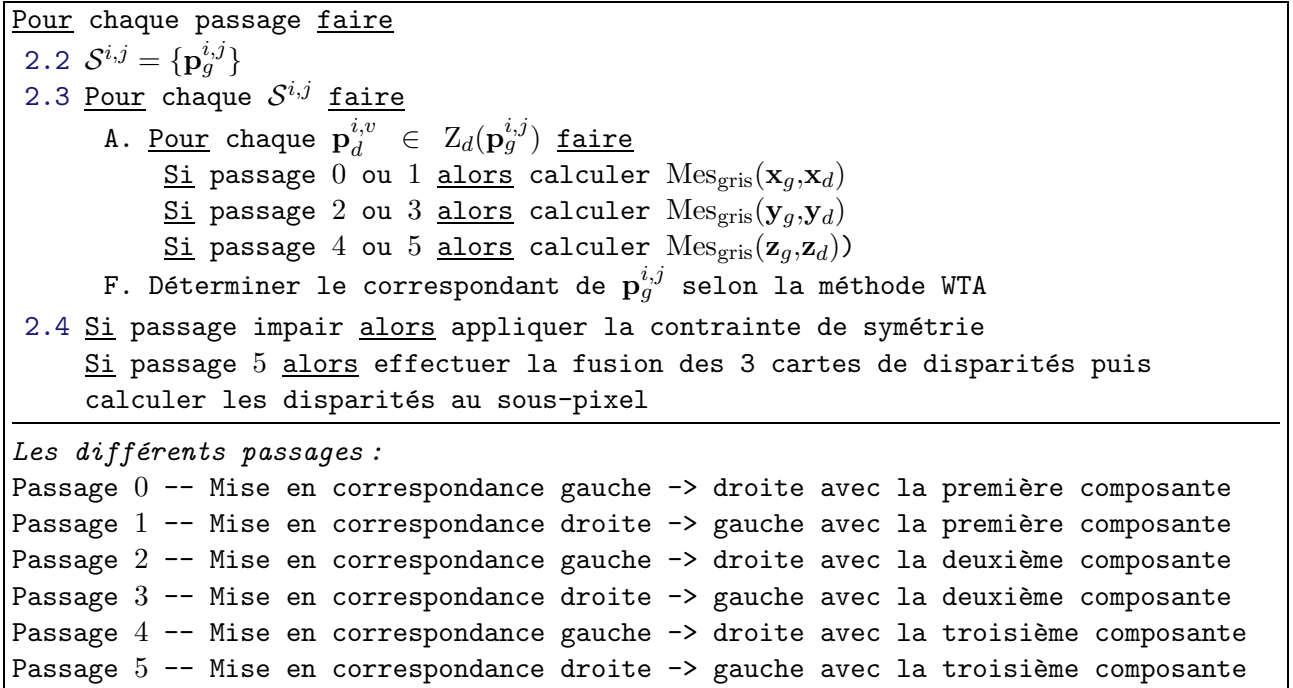


FIG. 4.1 – Méthode FUSION-SCORE – Recherche d’un correspondant en utilisant trois composantes et la fusion des scores (voir figure 3.2(b) pour le calcul des scores avec une mesure 1D).

4.4.1.2 Fusion des cartes de disparités

Une autre solution, la méthode FUSION-CARTE, consiste à calculer les trois cartes de disparités suivant les trois composantes, puis de les fusionner (cf. algorithme 4.2 et figure 4.2). En chaque pixel, nous préconisons l’application des règles suivantes :

- Si au moins deux des trois cartes de disparités donnent le même résultat alors ce résultat est conservé.
- Si les trois cartes donnent trois résultats différents alors la disparité correspondant au meilleur score de corrélation est conservée.



ALGO. 4.2 – Méthode FUSION-CARTE – Cet algorithme est une instance de l'algorithme 1.2. C'est un algorithme à six passages car nous appliquons la contrainte de symétrie sur chaque carte de disparités obtenue sur chaque composante. Il n'y a pas d'étape de prétraitement (disparition de l'étape 2.1).

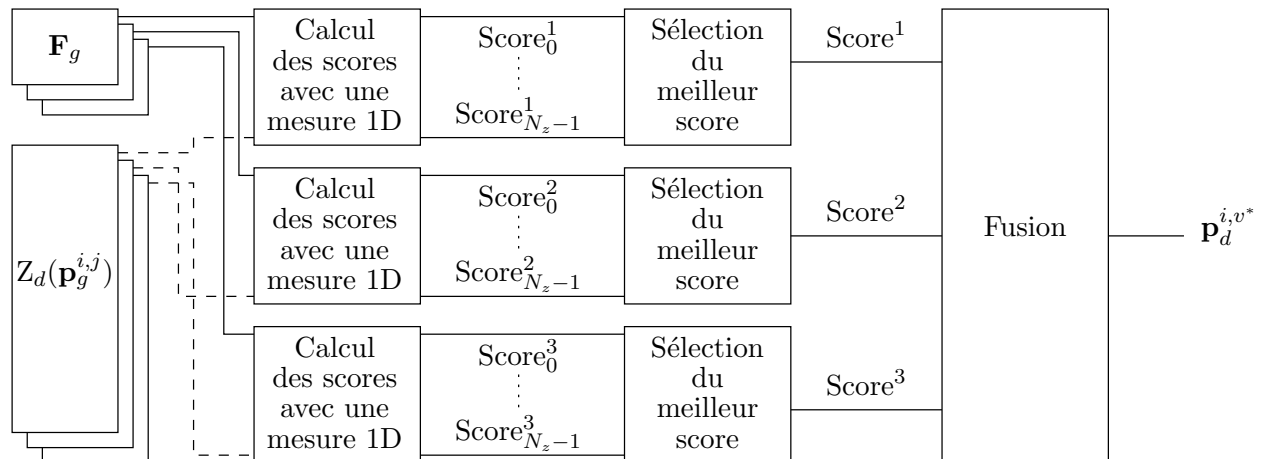


FIG. 4.2 – Méthode FUSION-CARTE – Recherche d'un correspondant en utilisant trois composantes et la fusion des meilleurs scores (voir figure 3.2(b) pour le calcul des scores avec une mesure 1D).

4.4.2 Analyse en composantes principales

Le but de l'analyse en composantes principales, ACP, est de réduire la dimension des données en ne conservant que les dimensions les plus pertinentes. Pour le problème que nous nous posons, il s'agit de réduire des données tridimensionnelles à des données unidimensionnelles. Deux méthodes différentes sont envisagées :

- Méthode ACP-IMAGE (cf. figure 4.3) – L'ACP est globale.
- Méthode ACP-CORRÉLATION (cf. figure 4.4) – L'ACP est locale. Nous pouvons penser que la composante retenue sera plus représentative. En revanche, cette méthode est plus coûteuse que la méthode ACP-IMAGE.

Ces méthodes de généralisation peuvent être très coûteuses en temps de calcul ; c'est pourquoi, seule l'ACP adaptée à l'utilisation en vision par ordinateur proposée par Cheng et Hsia [Cheng 03] sera utilisée (cf. algorithme 4.3).

1. Calculer le vecteur des moyennes \mathbf{m}_l

$$\mathbf{m}_l = \frac{1}{N_f} \left(\sum_{k=0}^{N_f-1} x_l^k \quad \sum_{k=0}^{N_f-1} y_l^k \quad \sum_{k=0}^{N_f-1} z_l^k \right)^T \quad (4.43)$$

2. Calculer le vecteur \mathbf{D}_l des distances entre chaque couleur et le vecteur des moyennes

$$\mathbf{D}_l = (\dots D_*(\mathbf{c}_l^k, \mathbf{m}_l) \dots)^T \quad (4.44)$$

3. Calculer la plus grande valeur D_1^{\max} du vecteur \mathbf{D}_l obtenue pour la couleur \mathbf{c}_l^{\max}

4. Pour chaque couleur \mathbf{c}_l^k faire calculer le score de projection proj_l^k

$$\text{proj}_l^k = D_1^{\max} - \frac{(\mathbf{c}_l^k - \mathbf{m}_l) \cdot (\mathbf{c}_l^{\max} - \mathbf{m}_l)}{D_1^{\max}} \quad (4.45)$$

5. Utiliser les valeurs des projections proj_l^k pour effectuer la mise en correspondance

ALGO. 4.3 – Analyse en composantes principales – Pour réduire des données 3D à des données 1D, nous utilisons la méthode de Cheng et Hsia [Cheng 03].

4.4.2.1 ACP sur l'image

La méthode ACP-IMAGE (cf. algorithme 4.4 et figure 4.3) consiste à effectuer une ACP sur toute l'image puis à réaliser la mise en correspondance à base de corrélation 1D à partir de la première composante.

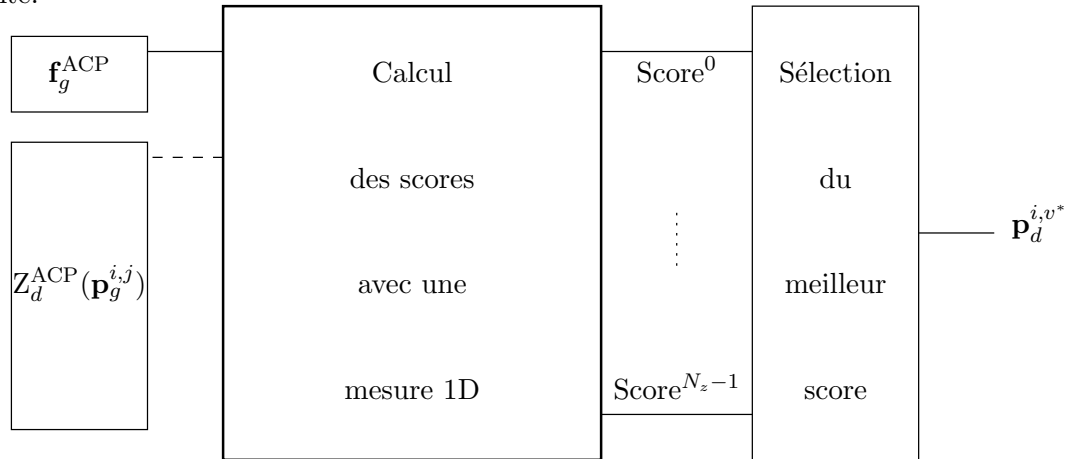


FIG. 4.3 – Méthode ACP-IMAGE – Le vecteur $\mathbf{f}_g^{\text{ACP}}$ contient la première composante principale des couleurs des pixels des fenêtres de corrélation de I_g et Z_d^{ACP} est la zone de recherche associée au pixel considéré (voir figure 3.2(b) pour le calcul des scores avec une mesure 1D).

Pour chaque passage faire

2.1 Analyse en composantes principales

2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$

2.3 Pour chaque $\mathcal{S}^{i,j}$ faire

A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire calculer $\text{Mes}_{\text{gris}}(\mathbf{f}_g^{\text{ACP}}, \mathbf{f}_d^{\text{ACP}})$

F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA

2.4 Si passage l alors appliquer la contrainte de symétrie puis calculer les disparités au sous-pixel

Les différents passages :

Passage 0 -- Mise en correspondance gauche -> droite

Passage 1 -- Mise en correspondance droite -> gauche

ALGO. 4.4 – Méthode ACP-IMAGE – Le vecteur $\mathbf{f}_g^{\text{ACP}}$ contient la première composante principale des couleurs des pixels des fenêtres de corrélation de I_g et Z_d^{ACP} est la zone de recherche associée au pixel étudié. Cet algorithme est une instance de l'algorithme 1.2. L'analyse en composantes principales est effectuée au cours de l'étape 2.1 (prétraitement).

4.4.2.2 ACP sur la fenêtre de corrélation

La méthode ACP-CORRÉLATION (cf. algorithme 4.5 et figure 4.4) consiste à effectuer une ACP sur chaque fenêtre de corrélation. Le score est alors calculé en utilisant :

$$\text{Mes}_{\text{couleur}}(\mathbf{F}_g, \mathbf{F}_d) = \text{Mes}_{\text{gris}}(\text{ACP}(\mathbf{F}_g), \text{ACP}(\mathbf{F}_d)), \quad (4.46)$$

où $\text{ACP}(\mathbf{F}_l)$ est le vecteur contenant la première composante principale des trois composantes couleur contenues dans le vecteur \mathbf{F}_l .

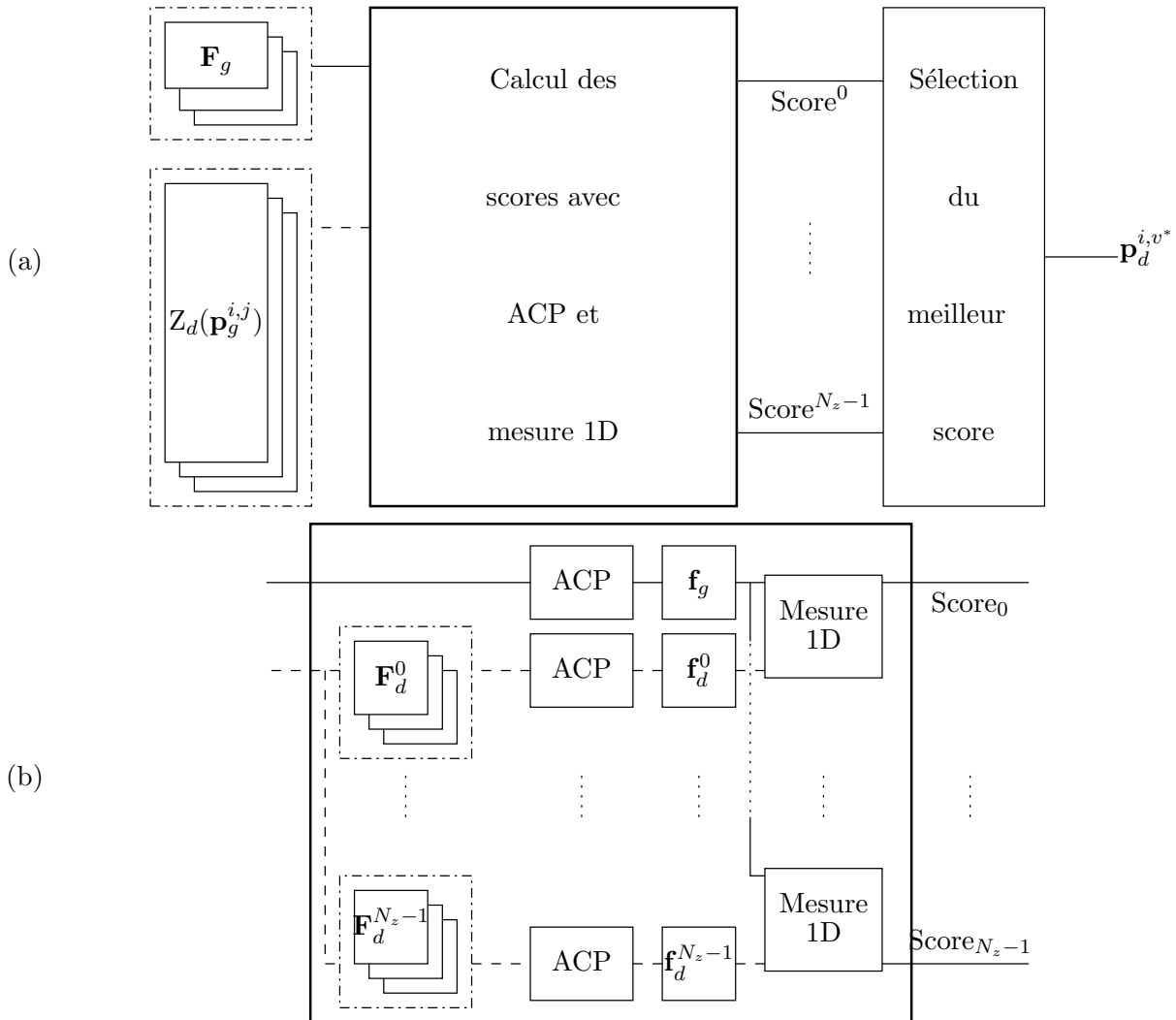


FIG. 4.4 – Méthode ACP-CORRÉLATION – (a) Recherche d'un correspondant en utilisant trois composantes et une mesure 1D. (b) Détails du calcul des scores avec une ACP et une mesure 1D.

<p><u>Pour</u> chaque passage <u>faire</u></p> <p>2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$</p> <p>2.3 <u>Pour</u> chaque $\mathcal{S}^{i,j}$ <u>faire</u></p> <p style="padding-left: 2em;">A. <u>Pour</u> chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ <u>faire</u> calculer $\text{Mes}(\text{ACP}(\mathbf{F}_g), \text{ACP}(\mathbf{F}_d))$</p> <p style="padding-left: 2em;">F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA</p> <p>2.4 <u>Si</u> passage 1 <u>alors</u> appliquer la contrainte de symétrie puis calculer les disparités au sous-pixel</p> <hr/> <p><i>Les différents passages :</i></p> <p>Passage 0 -- Mise en correspondance gauche -> droite</p> <p>Passage 1 -- Mise en correspondance droite -> gauche</p>

ALGO. 4.5 – Méthode ACP-CORRÉLATION – *Cet algorithme est une instance de l'algorithme 1.2. C'est à l'étape A que l'ACP est effectuée. Il n'y a pas de prétraitement (disparition de l'étape 2.1).*

4.4.3 Mise en correspondance avec des mesures 3D

Cette dernière méthode, appelée méthode DIRECTE, consiste à remplacer la mesure de corrélation 1D par la mesure de corrélation 3D correspondante. Les mesures 3D sont décrites dans le paragraphe 4.3 (cf. algorithme 3.2 et figure 4.5).

4.5 Résultats expérimentaux

4.5.1 Protocole d'évaluation

4.5.1.1 Images testées

Les images testées sont toutes celles présentées au paragraphe 2.3.4 excepté le couple *map* car Scharstein et Szeliski [Scharstein 02] ne fournissent que des images en niveaux de gris pour ce couple.

4.5.1.2 Critères d'évaluation

Le protocole présenté dans le chapitre 2 est utilisé. Cependant, nous pouvons ajouter deux zones d'évaluation différentes qui n'ont pas été abordées dans le chapitre 2 car elles sont très spécifiques à l'utilisation de la couleur. Nous avons distingué les zones chromatiques des zones achromatiques en utilisant la technique exposée dans [Tseng 92, Koschan 93]. Les régions achromatiques peuvent être distinguées des régions chromatiques en utilisant des seuils spécifiques sur les composantes *I* et *S* du système de couleur *HSI* (cf. paragraphe 4.2.3) :

$$C(\mathbf{p}_g^{i,j}) = \begin{cases} 0 & \text{Si } I_g^{i,j} > 95 \text{ ou } (I_g^{i,j} > 80 \text{ et } S_g^{i,j} < 18) \text{ ou} \\ & (I_g^{i,j} > 60 \text{ et } S_g^{i,j} < 20) \text{ ou } (I_g^{i,j} > 50 \text{ et } S_g^{i,j} < 30) \text{ ou} \\ & (I_g^{i,j} > 40 \text{ et } S_g^{i,j} < 40) \text{ ou } (I_g^{i,j} > 25 \text{ et } S_g^{i,j} < 60) \\ 1 & \text{sinon.} \end{cases} \quad (4.47)$$

Si $C(\mathbf{p}_g^{i,j}) = 0$ alors $\mathbf{p}_g^{i,j}$ est un pixel qui se trouve dans une région achromatique, sinon dans une région chromatique. Nous avons détecté ces zones dans les différentes images testées (cf. figure 4.6) et nous avons évalué les différents critères dans chacune des zones. Tous les résultats fournis dans les tableaux de ce chapitre ont été obtenus sur la zone chromatique.

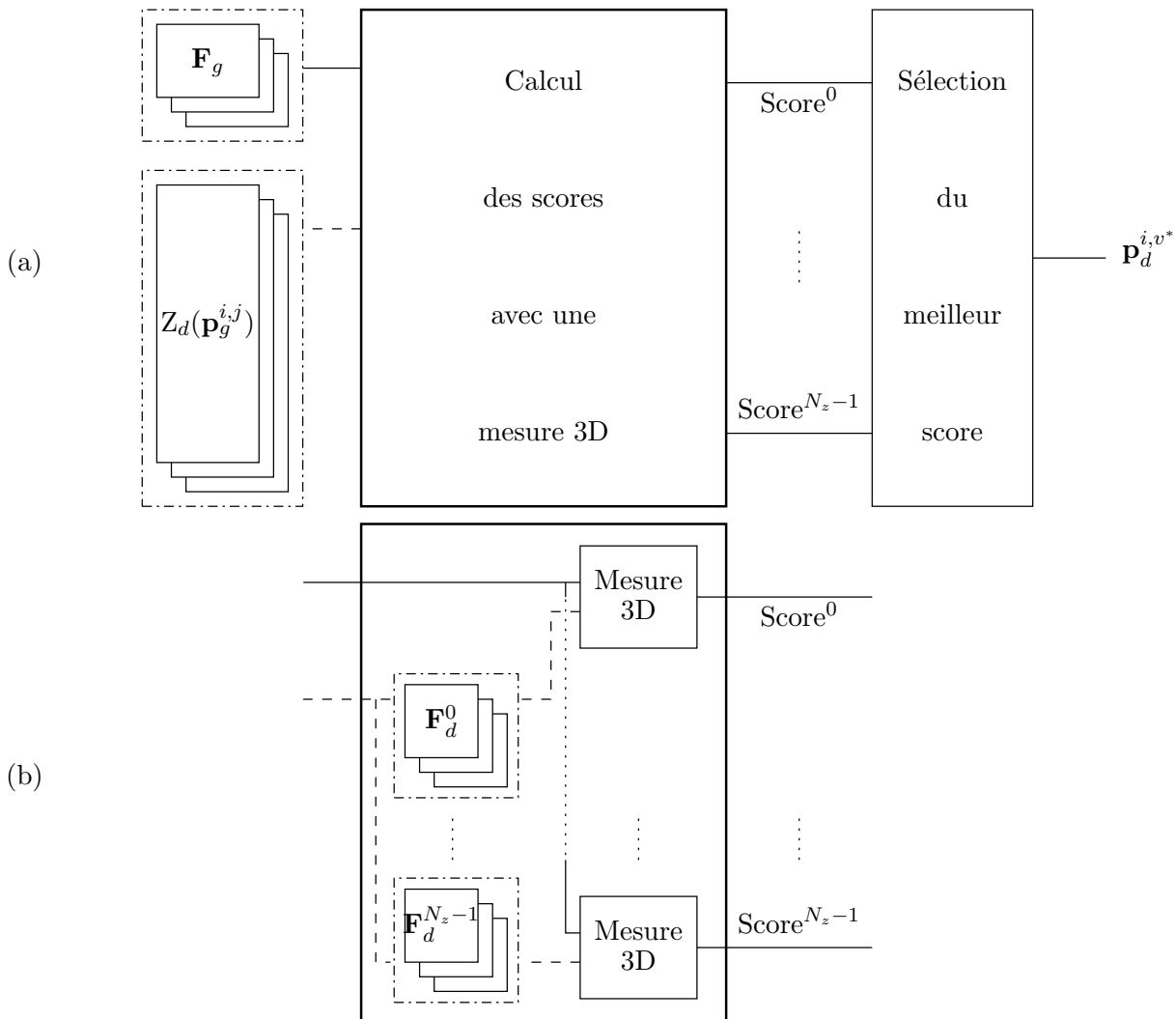


FIG. 4.5 – Méthode DIRECTE – (a) Recherche d'un correspondant en utilisant trois composantes et une mesure 3D. (b) Détails du calcul des scores avec une mesure 3D.

4.5.1.3 Méthodes évaluées

Nous avons évalué les trois méthodes proposées et les variantes associées, avec les dix systèmes de représentation de la couleur et pour chaque mesure de corrélation. Nous avons repris les évaluations que nous avons effectuées dans [Chambon 04a, Chambon 05], en utilisant une zone d'évaluation plus fine (la zone chromatique) et nous n'avons testé que les systèmes et les méthodes qui donnent les résultats les plus performants selon notre évaluation dans [Chambon 05], c'est-à-dire :

- les systèmes RGB , XYZ , $I_1I_2I_3$ et $H_1H_2H_3$;
- les méthodes FUSION-SCORE et DIRECTE.

Enfin, nous ne donnons les résultats que pour cinq des mesures pour lesquelles les améliorations sont les plus importantes (une par famille) : NCC, SAD, PRATT, ISC et SMPD₂.


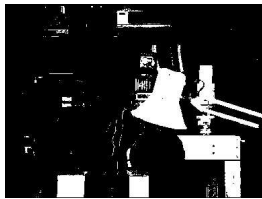



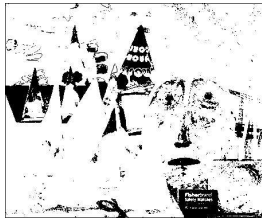
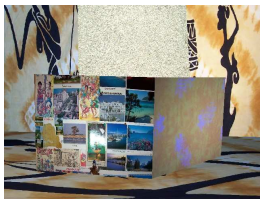



NOM	IMAGE GAUCHE	ZONES ACHROMATIQUES	COMMENTAIRE
<i>Tsukuba</i>			Cette image contient une zone chromatique qui couvre 14.47% de l'image.
<i>Sawtooth</i>			Cette image contient une zone chromatique qui couvre 32.74% de l'image. Tout le premier plan (le plan en dents de scie) est achromatique.
<i>Cones</i>			Cette image possède une des zones chromatiques les plus importantes, 86.34% de l'image.
<i>Murs</i>			Cette image de synthèse possèdent une zone chromatique qui couvre 33.75% de l'image.
<i>Livres</i>			Cette image réelle possède une grande zone chromatique qui représente 58.45% de l'image.

FIG. 4.6 – Zones achromatiques et zones chromatiques dans les images testées – Les pixels blancs correspondent aux pixels qui se trouvent dans une zone chromatique.

4.5.1.4 Synthèse des critères et classement des méthodes

Dans ce chapitre, nous prenons en compte uniquement les critères utilisés au chapitre 3. De plus, nous exposons les cartes de disparités et les cartes d'appariements corrects pour la méthode couleur qui a donné le meilleur résultat et la méthode à base de corrélation 1D. Dans les tableaux de résultats et les figures avec les différentes cartes, nous notons « gris » la méthode à base de corrélation 1D et « couleur » la méthode à base de couleur. De plus, dans chaque tableau, nous précisons les caractéristiques de la méthode « couleur », c'est-à-dire, le type de la méthode dans la colonne MET (méthode FUSION, notée FUS ou méthode DIRECTE, notée DIR), le système de couleur utilisé et la variante choisie dans la colonne DET (minimum, noté Min, moyenne, noté Moy ou médiane, noté Med). De plus, nous précisons la mesure dans la colonne MES et nous notons en gras la meilleure valeur pour chaque critère (c'est-à-dire chaque colonne).

4.5.2 Présentation et analyse des résultats

4.5.2.1 Temps de calcul

Les résultats obtenus montrent que, comme cela était prévisible, les temps de calcul pour la méthode FUSION sont trois fois plus élevés que pour la méthode à base de corrélation 1D. La méthode DIRECTE est toujours moins coûteuse que la méthode FUSION mais elle est toujours plus coûteuse que la méthode à base de corrélation 1D.

4.5.2.2 Résultats préliminaires

Les tests que nous avons réalisés représentent une masse de résultats très importante dont une première analyse dans [Chambon 04a, Chambon 05], nous permet de faire les remarques suivantes :

- Les systèmes CMY , $L^*a^*b^*$, $L^*u^*v^*$, AC_1C_2 , YCh_1YCh_2 et HSI obtiennent rarement de meilleurs résultats que les résultats obtenus avec les autres systèmes.
- La méthode FUSION-SCORE est plus performante que la méthode FUSION-CARTE.
- Les méthodes ACP ne sont pas satisfaisantes.
- Pour la méthode DIRECTE, quelle que soit la distance utilisée, les résultats sont assez proches les uns des autres. Pour la mesure GC, les résultats obtenus sont les plus performants en utilisant la fusion des vecteurs gradients suivant la méthode de Lee et Cok [Lee 91]. Pour les mesures utilisant les rangs, les meilleurs résultats sont obtenus en utilisant l'ordre lexicographique. Pour les mesures $SMPD_2$ et R_m , il est préférable de calculer le vecteur des distances entre les couleurs puis la médiane et le rang des données de ces vecteurs.

4.5.2.3 Analyse des performances

D'une manière générale, quelle que soit la famille et la mesure, les meilleurs résultats sont obtenus avec la méthode FUSION-SCORE, en utilisant l'opérateur minimum pour la fusion (la fusion du barycentre pondéré est préférable pour les mesures ordinales) et le système XYZ , aussi bien dans les zones chromatiques que dans les zones achromatiques. Dans nos précédents travaux, cf. [Chambon 05], nous avons aussi remarqué que la méthode DIRECTE, en utilisant le système $H_1H_2H_3$, donne aussi de bons résultats. En comparant les résultats obtenus dans la zone chromatique et dans la zone achromatique, nous avons pu remarquer que cette méthode donne effectivement de bons résultats dans la zone chromatique mais, en revanche, elle a tendance à dégrader les résultats dans la zone achromatique. Nous pouvons penser que dans la zone achromatique, certaines des composantes couleur sont uniformes et peuvent entacher le calcul du score de corrélation d'erreur. La méthode DIRECTE va prendre en compte ces composantes alors que la méthode FUSION-SCORE, en utilisant le minimum, ne va pas prendre en compte ces composantes. Cette différence de comportement explique en partie la raison pour laquelle la méthode FUSION-SCORE obtient souvent de meilleurs résultats que la méthode DIRECTE.

Quelle que soit la mesure, l'utilisation de la couleur améliore globalement, c'est-à-dire pour la plupart des critères, les résultats. Les améliorations les plus importantes sont obtenues avec les mesures 1D les moins performantes, cf. les résultats du chapitre 3, paragraphe 3.6.

Afin d'affiner notre analyse des résultats, nous donnons le détail des résultats obtenus sur quatre des couples d'images les plus représentatifs : *journaux*, *barn1*, *teddy* et « livres ».

MES	MET	DET	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
NCC	GRIS		2	8	8	4	8	4	8	8	9	8	8
	FUS- SCORE	XYZ Min	1	10	10	3	5	3	7	5	8	9	6
SAD	GRIS		5	6	4	6	6	5	6	6	7	5	5
	FUS- SCORE	XYZ Min	3	3	6	2	7	1	5	7	6	2	3
PRATT	GRIS		10	7	1	10	10	10	9	9	10	1	10
	FUS- SCORE	XYZ Min	9	4	2	9	9	8	10	10	4	7	9
ISC	GRIS		8	9	3	8	3	9	4	3	5	10	7
	FUS- SCORE	XYZ Belli	6	1	7	7	4	6	3	4	2	4	4
SMPD ₂	GRIS		7	2	5	5	2	7	2	2	3	6	2
	FUS- SCORE	XYZ Min	4	5	9	1	1	2	1	1	1	3	1

TAB. 4.7 – *Classement des méthodes de mise en correspondance d’images couleur* – Nous avons synthétisé les résultats obtenus sur les treize images de tests avec le protocole que nous avons décrit dans le paragraphe 2.9. Nous remarquons que la couleur améliore les résultats sur la plupart des critères. La meilleure méthode est FUSION-SCORE, avec le système XYZ et en utilisant le minimum pour la fusion.

Journaux – Pour ces images, l’amélioration est faible, cf. figure 4.7 et tableau 4.8, surtout pour les deux premières mesures SAD et NCC. Cette faible amélioration est due au fait que la zone chromatique est assez petite (16.76% de l’image est chromatique). Les résultats obtenus avec ces images permettent d’illustrer le fait que la méthode utilisant la couleur ne dégrade pas les résultats même si les images sont quasiment achromatiques.

Barn1 – Nous pouvons voir dans cet exemple que la méthode FUSION-SCORE est toujours la meilleure. En revanche, le système de couleur utilisé varie d’une mesure à l’autre. Pour PRATT, les résultats avec la méthode FUSION-SCORE sont nettement meilleurs que ceux obtenus avec la corrélation 1D, mais cela n’empêche pas cette mesure d’être la moins performante des cinq présentées dans le tableau 4.9. Sur la figure 4.8, l’amélioration des résultats est visible si nous analysons la proportion de faux négatifs dans les cartes de disparités et la densité des cartes d’appariements corrects.

Teddy – Pour ces images, l’amélioration des résultats est visible dans la figure 4.9 où les cartes de disparités sont beaucoup plus nettes et les cartes d’appariements corrects beaucoup plus denses, surtout dans la zone de l’image qui comporte le toit de la maison. Sur le tableau 4.10, nous remarquons que l’utilisation de la couleur améliore les résultats pour la plupart des critères et SMPD₂ en 3D est classée première pour la plupart des critères.

Livres – Les meilleurs résultats ne sont pas obtenus avec les mêmes méthodes que pour les autres images, cf. tableau 4.11 et figure 4.10. En effet, pour les mesures NCC et PRATT, c’est la méthode DIRECTE avec $I_1I_2I_3$ ou $H_1H_2H_3$ qui obtient les meilleurs résultats. De plus, notons que pour ces images, l’amélioration globale des critères est moins importante qu’avec les autres images. Cela est dû au fait que ces images possèdent plus de zones uniformes que les autres images.

4.6 Conclusion

Nous avons proposé trois manières d’adapter la mise en correspondance stéréoscopique de pixels à la couleur. L’évaluation que nous avons faite nous permet de conclure qu’il vaut mieux utiliser le système XYZ et la méthode la plus performante consiste à fusionner, en utilisant l’opérateur minimum, les scores de corrélation évalués sur chaque composante. Cette méthode améliore les résultats dans les zones chromatiques et ne détériore pas les résultats dans les zones achromatiques, contrairement à la méthode qui consiste à appliquer une mesure de corrélation 3D.

Le chapitre 3 nous a permis de montrer l’efficacité de la mesure $SMPD_2$ dans les différentes zones des occultations. Ce chapitre a mis en évidence le fait que la couleur améliore ces résultats. Cependant, nous avons toujours un pourcentage d’appariements corrects meilleurs avec des mesures comme ZNCC et SAD. Nous allons donc aborder à présent des méthodes hybrides utilisant deux mesures de corrélation 3D pour tenter de combiner les avantages des mesures $SMPD_2$ et ZNCC ou SAD.

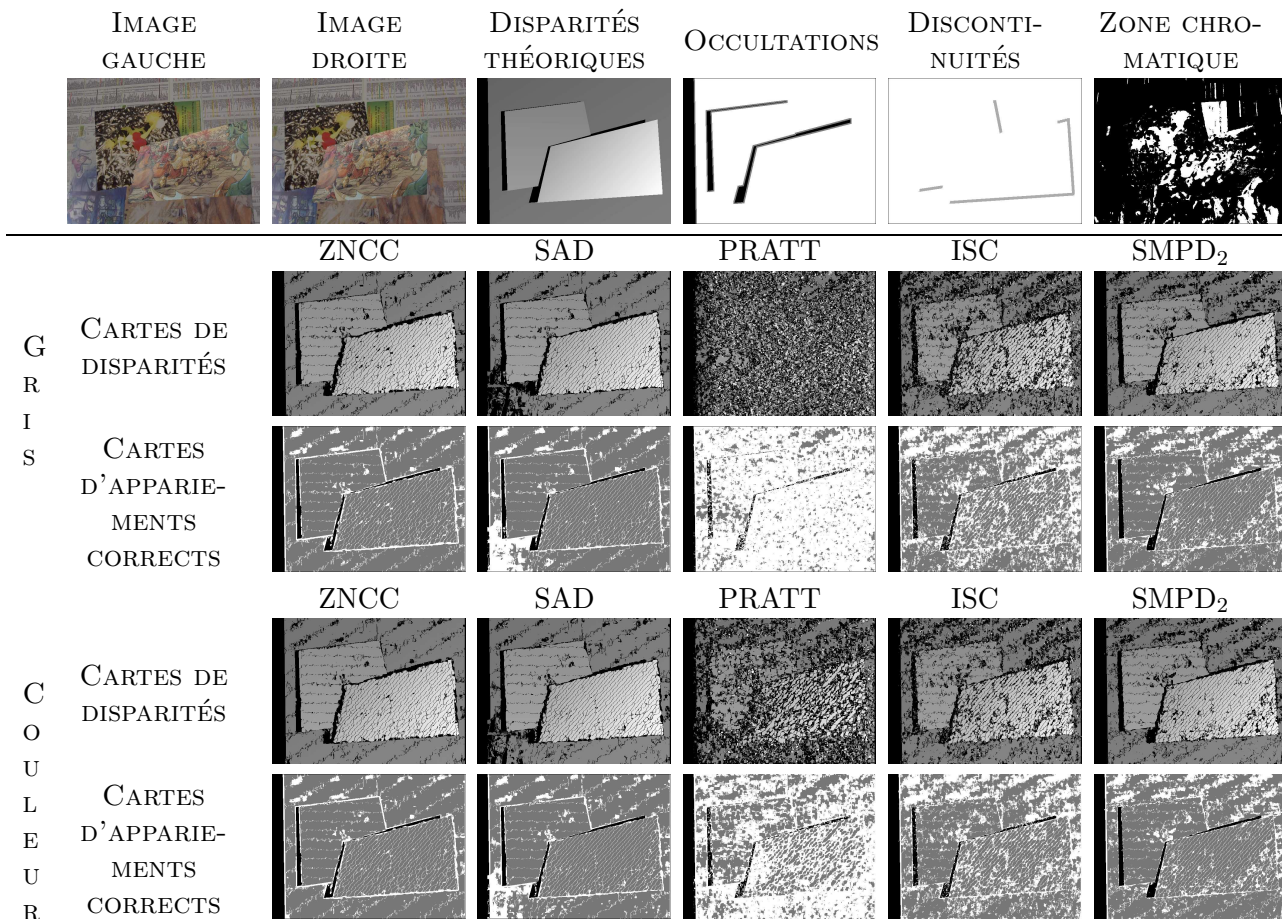


FIG. 4.7 – Cartes de résultats obtenus avec le couple d’images « journaux » – Pour ces images, les améliorations sont peu visibles (cela est dû au fait que la zone chromatique couvre une faible partie de l’image, 16.76%).

MES	MET	DET	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
NCC	GRIS		87.42	0.19	0.05	0.05	0.95	0.38	11.01	65.84	85.86	48.74	7
			(3)	(10)	(9)	(4)	(6)	(6)	(8)	(7)	(6)	(7)	
	DIR	RGB	87.77	0.45	0.03	0.03	0.86	0.36	10.53	65.91	86.79	48.07	6
			(2)	(7)	(10)	(2)	(5)	(5)	(10)	(6)	(5)	(8)	
SAD	GRIS		87.32	0.28	0.52	0.03	0.72	0.28	11.37	72.12	89.73	57.08	4
			(4)	(9)	(6)	(2)	(4)	(4)	(7)	(4)	(4)	(4)	
	DIR	$H_1H_2H_3$ L_1	87.79	0.37	0.51	0.01	0.66	0.23	10.94	72.39	91.46	56.1	3
			(1)	(8)	(7)	(1)	(3)	(3)	(9)	(3)	(3)	(5)	
PRATT	GRIS		10.81	1.01	1	1.02	38.43	1.14	47.37	53.84	57.72	50.52	10
			(10)	(4)	(1)	(10)	(10)	(10)	(1)	(10)	(10)	(6)	
	FUS- SCORE	XYZ Min	34.32	1.02	0.85	0.68	18.23	0.75	45	56.35	72.36	42.67	8
			(9)	(3)	(2)	(9)	(9)	(9)	(2)	(9)	(9)	(10)	
ISC	GRIS		64.66	0.75	0.71	0.25	4.32	0.45	29.56	62.53	83.2	44.88	8
			(8)	(6)	(5)	(8)	(8)	(8)	(3)	(8)	(8)	(9)	
	FUS- SCORE	RGB Belli	73.21	1.42	0.16	0.16	2.34	0.39	22.49	67.49	85.71	61.72	5
			(7)	(2)	(8)	(6)	(7)	(7)	(4)	(5)	(7)	(1)	
SMPD ₂	GRIS		82.64	0.86	0.81	0.19	0.59	0.19	15.52	75.99	92.89	61.56	2
			(6)	(5)	(3)	(7)	(2)	(2)	(5)	(1)	(2)	(2)	
	FUS- SCORE	XYZ Min	84.82	1.5	0.77	0.12	0.36	0.16	13.05	75.89	94.04	60.39	1
			(5)	(1)	(4)	(5)	(1)	(1)	(6)	(2)	(1)	(3)	

TAB. 4.8 – Résultats obtenus avec le couple d’images « journaux » – Pour les mesures PRATT, ISC et SMPD₂, l’utilisation de la couleur permet d’obtenir de meilleurs résultats pour la plupart des critères. Pour les autres mesures, les améliorations sont moins importantes. Cette image possède une petite zone chromatique, ainsi, les critères ont été évalués sur une petite partie de l’image. Le codage des différentes cartes est donné aux paragraphes 2.2.1 et 2.9.5 du chapitre 2.

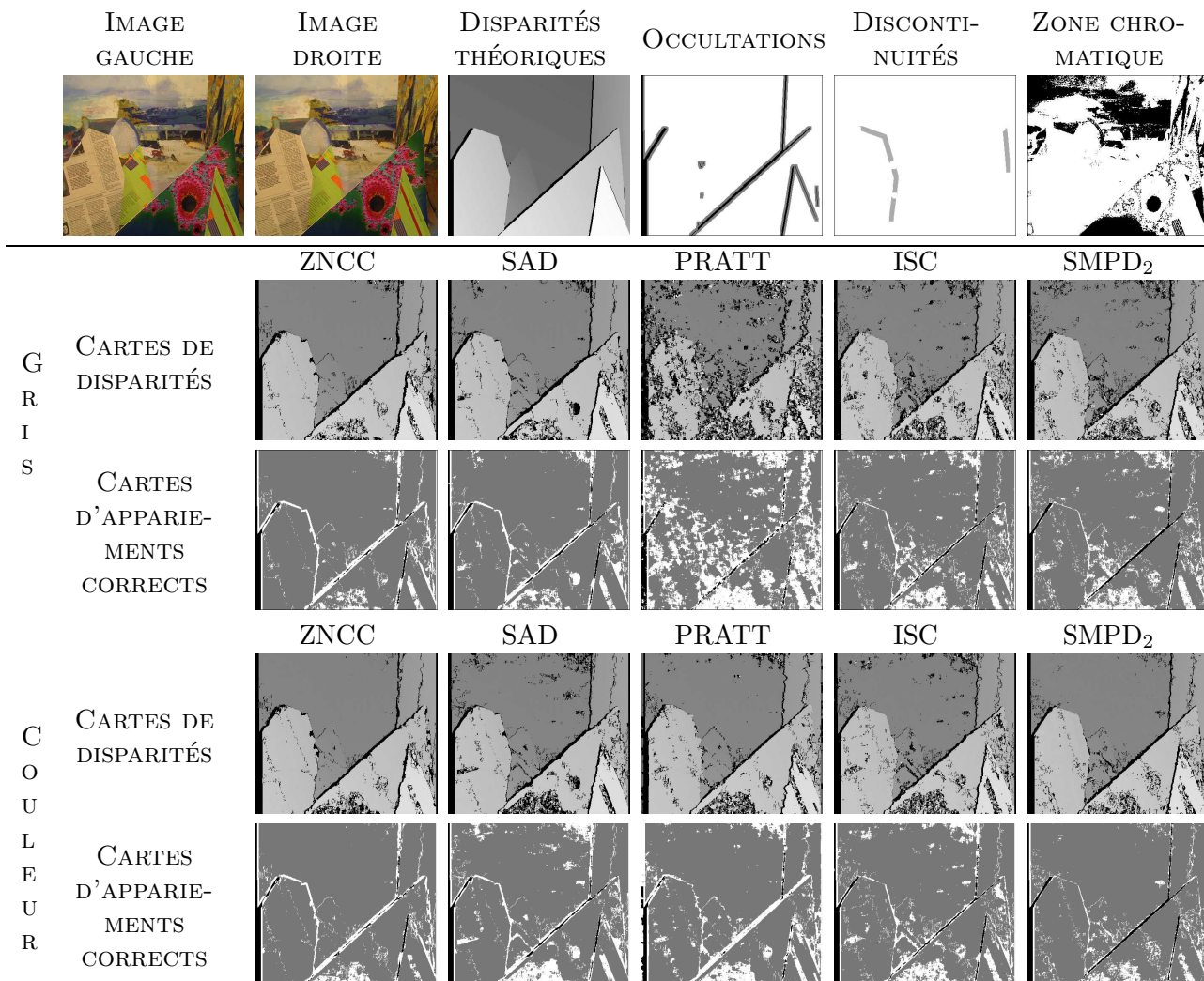


FIG. 4.8 – Cartes de résultats obtenus avec le couple d'images barn1 – L'amélioration des résultats est surtout visible pour la mesure PRATT. Globalement, la proportion de faux négatifs est moins importante dans les cartes de disparités obtenues avec la méthode couleur.

MES	MET	DET	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
NCC	GRIS		89.81	0.21	0.27	0.93	1.54	7.25	60.4	65.84	54.52	57.28	8
			(3)	(7)	(7)	(4)	(9)	(3)	(9)	(9)	(9)	(10)	
	FUS-SCORE	$H_1H_2H_3$ Min	90.84	0.16	0.17	0.69	1.5	6.63	61.21	66.7	55.77	58.05	7
			(2)	(8)	(9)	(2)	(8)	(1)	(8)	(8)	(8)	(9)	
SAD	GRIS		89.64	0.43	0.3	0.93	1.29	7.42	65.33	71.4	56.95	61.84	6
			(6)	(3)	(5)	(4)	(6)	(4)	(7)	(7)	(6)	(7)	
	FUS-SCORE	RGB Med	89.75	0.44	0.3	0.86	1.23	7.42	66.57	72.68	56.18	63.06	3
			(4)	(2)	(5)	(3)	(5)	(4)	(5)	(5)	(7)	(6)	
PRATT	GRIS		69.42	0.33	1.3	6.65	1.3	19.33	65.49	72.51	50.9	61.45	10
			(10)	(4)	(2)	(10)	(7)	(10)	(6)	(6)	(10)	(8)	
	FUS-SCORE	XYZ Min	85.9	0.15	1.31	2.4	2.51	7.73	57.05	44.84	67.32	64.07	9
			(9)	(10)	(1)	(9)	(10)	(7)	(10)	(10)	(1)	(5)	
ISC	GRIS		86.41	0.25	0.34	1.62	1.04	10.34	73.06	76.97	59.87	70.81	5
			(8)	(6)	(4)	(8)	(3)	(9)	(3)	(3)	(5)	(3)	
	FUS-SCORE	RGB Belli	89.74	0.16	0.23	1 (6)	1.15	7.72	72.19	74.54	62.45	70.85	4
			(5)	(8)	(8)	(4)	(6)	(4)	(4)	(4)	(2)	(2)	
SMPD ₂	GRIS		87.04	0.71	0.38	1.3	0.55	10.04	76.36	87.87	61.05	69.74	2
			(7)	(1)	(3)	(7)	(2)	(8)	(2)	(2)	(4)	(4)	
	FUS-SCORE	XYZ Min	91.61	0.32	0.11	0.43	0.48	7.06	78.05	89.34	61.16	71.56	1
			(1)	(5)	(10)	(1)	(1)	(2)	(1)	(1)	(3)	(1)	

TAB. 4.9 – Résultats obtenus avec le couple d'images barn1 – Après PRATT, SMPD₂ obtient la meilleure amélioration pour le pourcentage d'appariements corrects.

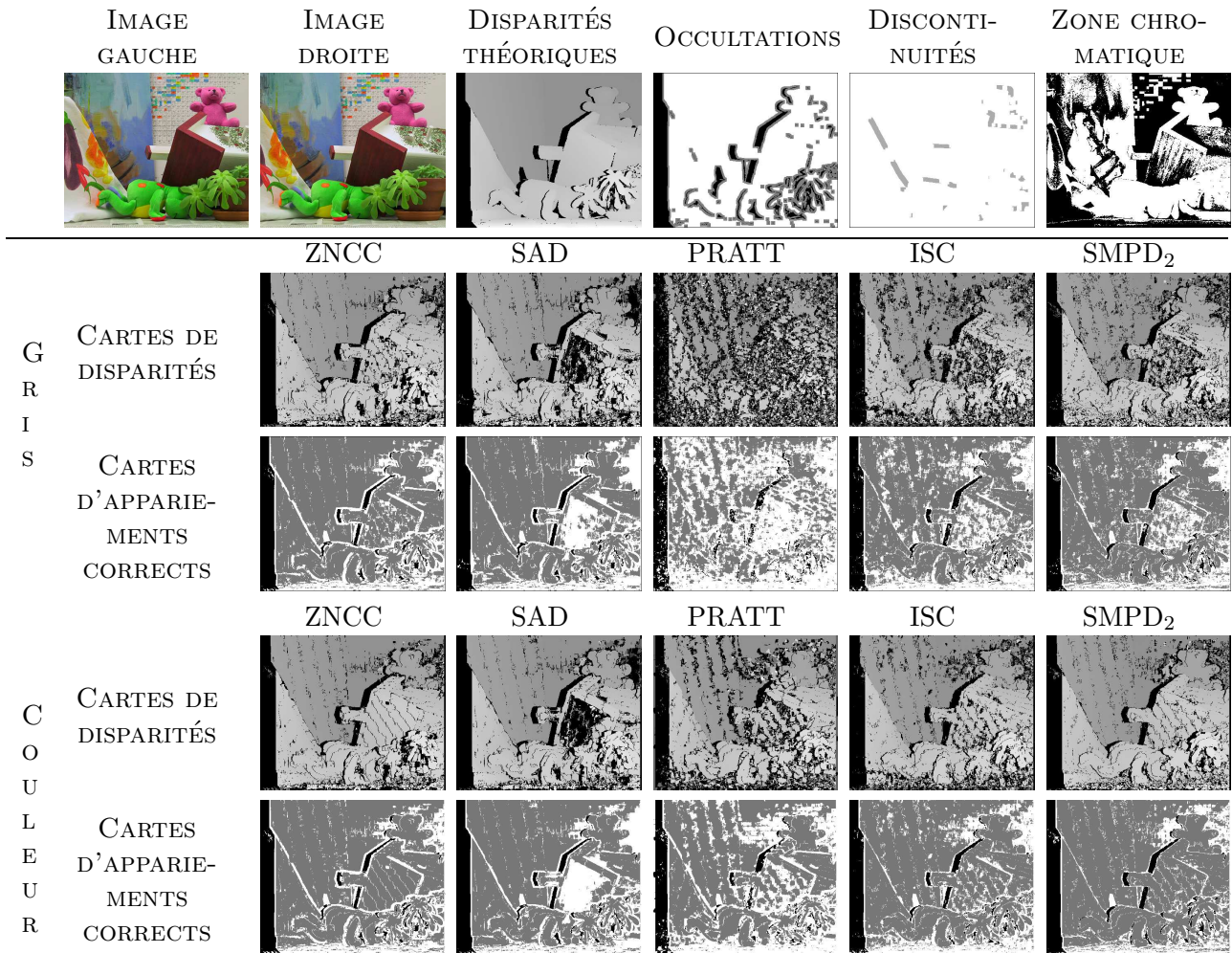


FIG. 4.9 – Cartes de résultats obtenus avec le couple d'images teddy – L'amélioration des résultats est visible pour ZNCC, PRATT, ISC et SMPD₂. Pour la mesure SAD, les résultats ne semblent pas meilleurs en utilisant la couleur, nous pouvons d'ailleurs remarquer que la carte d'appariements corrects obtenue avec la méthode couleur est la moins dense au niveau du toit de la maison. Si nous analysons de manière quantitative les résultats, cf. tableau 4.10, nous pouvons confirmer que l'utilisation de la couleur n'améliore pas beaucoup les résultats pour SAD.

MES	MET	DET	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
NCC	GRIS		75.87	2.37	1.42	2.62	2.4	10.83	70.95	65.11	72.98	78.73	4
			(4)	(3)	(3)	(3)	(5)	(3)	(8)	(6)	(7)	(7)	
	FUS-SCORE	$H_1H_2H_3$ Min	76.98	2.5	1.24	2.31	2.28	10.09	71.41	66.9	72.97	78.75	2
			(3)	(1)	(5)	(2)	(3)	(2)	(6)	(4)	(8)	(6)	
SAD	GRIS		72.09	2.29	1.31	3.48	2.58	13.94	72.19	62.6	75.52	80.39	7
			(6)	(4)	(4)	(6)	(9)	(7)	(5)	(9)	(5)	(4)	
	FUS-SCORE	XYZ Min	72.07	2.39	1.21	3.62	2.43	13.81	72.96	64.68	75.84	81.1	6
			(7)	(2)	(6)	(7)	(7)	(6)	(4)	(7)	(4)	(3)	
PRATT	GRIS		41.42	0.43	1.48	16.58	2.4	32.82	60.52	70.61	57.01	59.91	9
			(10)	(10)	(2)	(10)	(5)	(10)	(10)	(1)	(10)	(10)	
	FUS-SCORE	XYZ Min	66.04	2.02	1.85	6.5	2.66	17.07	69.35	61.38	72.11	77.88	9
			(9)	(5)	(1)	(9)	(10)	(8)	(9)	(10)	(9)	(8)	
ISC	GRIS		68.1	1.17	1.01	4.87	2.34	17.96	71.37	66.01	73.23	76.26	8
			(8)	(9)	(7)	(8)	(4)	(9)	(7)	(5)	(6)	(9)	
	FUS-SCORE	XYZ Belli	77.03	1.27	0.89	2.69	2.48	11.24	74.79	64	78.54	81.62	5
			(2)	(8)	(9)	(4)	(8)	(4)	(3)	(8)	(2)	(2)	
SMPD ₂	GRIS		74.17	1.68	0.9	2.7	2.2	13.66	75.14	68.07	77.6	78.99	3
			(5)	(6)	(8)	(5)	(2)	(5)	(2)	(3)	(3)	(5)	
	FUS-SCORE	XYZ Min	81.56	1.31	0.55	1.32	2.09	8.37	78.37	69.62	81.4	83	1
			(1)	(7)	(10)	(1)	(1)	(1)	(1)	(2)	(1)	(1)	

TAB. 4.10 – Résultats obtenus avec le couple d’images teddy – Quelle que soit la mesure, la méthode utilisant la couleur, qui est toujours la méthode FUSION-SCORE, est toujours mieux classée que celle utilisant la corrélation 1D. L’utilisation de la couleur accentue l’écart entre la mesure SMPD₂ et les autres, car cette mesure est classée première pour la plupart des critères. Notons que la mesure NCC en 3D est mieux classée que la mesure SMPD₂ en 1D.

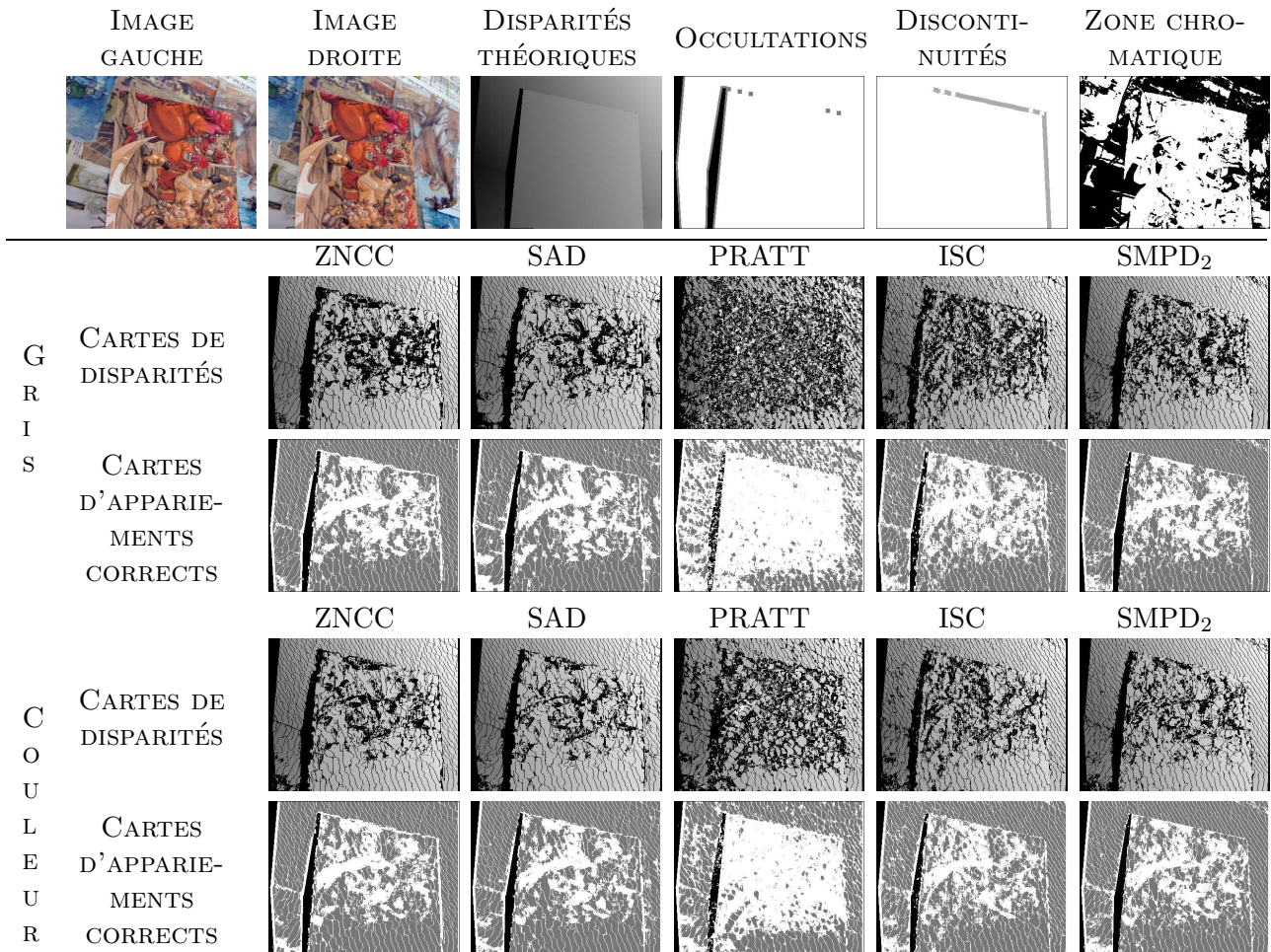


FIG. 4.10 – Cartes de résultats obtenus avec le couple d'images « livres » – Une fois de plus, les améliorations sont surtout visibles pour la mesure PRATT.

MES	MET	DET	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
NCC	GRIS		60.25	5.17	1.81	2.29	0.47	0.06	30.41	69.82	96.91	36.74	5
			(6)	(8)	(7)	(3)	(4)	(5)	(4)	(7)	(6)	(7)	
	DIR	$I_1I_2I_3$	64	6.21	2.05	1.9	0.51	0.05	25.78	72.3	97.42	41.62	3
			(2)	(4)	(6)	(1)	(5)	(4)	(9)	(5)	(4)	(5)	
SAD	GRIS		60.76	6.12	2.12	2.86	0.52	0.03	28.12	74.25	98.65	44.45	3
			(5)	(5)	(5)	(6)	(6)	(2)	(7)	(3)	(2)	(4)	
	FUS-SCORE	$H_1H_2H_3$ Moy	65.04	7.34	2.33	2.16	0.46	0.03	23.1	76.12	98.65	48.62	1
			(1)	(2)	(3)	(2)	(3)	(2)	(10)	(1)	(2)	(1)	
PRATT	GRIS		20.07	4.07	2.66	27.21	2.41	0.36	45.64	59.83	82.99	31.55	10
			(10)	(9)	(2)	(10)	(10)	(10)	(1)	(10)	(10)	(9)	
	DIR	$H_1H_2H_3$	33.09	11.1	4.64	12.32	1.35	0.19	38.66	63.41	90.91	29.82	8
			(9)	(1)	(1)	(9)	(8)	(9)	(2)	(9)	(9)	(10)	
ISC	GRIS		54.4	3.68	1.34	5.82	1.39	0.1	34.66	70.92	95.17	41.31	9
			(8)	(10)	(10)	(8)	(9)	(8)	(3)	(6)	(8)	(6)	
	FUS-SCORE	$H_1H_2H_3$ Moy	61.39	5.54	1.72	2.46	1.18	0.07	28.81	73.75	96.71	45.71	5
			(4)	(6)	(9)	(5)	(7)	(7)	(6)	(4)	(7)	(2)	
SMPD ₂	GRIS		60.02	5.51	1.79	2.93	0.33	0.06	29.69	69.64	97.36	35.8	7
			(7)	(7)	(8)	(7)	(1)	(5)	(5)	(8)	(5)	(8)	
	FUS-SCORE	$I_1I_2I_3$ Moy	62.89	6.63	2.21	2.43	0.33	0.02	25.82	74.96	98.9	45.71	2
			(3)	(3)	(4)	(4)	(1)	(1)	(8)	(2)	(1)	(2)	

TAB. 4.11 – Résultats obtenus avec le couple d'images « livres » – Les systèmes de couleur qui permettent d'obtenir les meilleurs résultats sont $I_1I_2I_3$ et $H_1H_2H_3$, ce qui est différent des résultats analysés jusqu'à présent. La mesure SAD est mieux classée que la mesure SMPD₂. Quelle que soit la mesure, la méthode utilisant la couleur est toujours mieux classée que celle utilisant la corrélation 1D.

Chapitre 5

Prise en compte des occultations

5.1 Introduction

Un des problèmes les plus délicats en vision par ordinateur, et notamment dans le cadre de la mise en correspondance stéréoscopique, est le problème des occultations. Dans la littérature, de nombreux articles, comme dans [Meer 00], ont été proposés pour prendre en compte ce problème. Les occultations représentent une difficulté majeure pour les techniques d'appariement et de très nombreux articles ont tenté de prendre en compte cette difficulté. Les différentes techniques proposées reposent sur des méthodes avec prétraitements, des méthodes locales à passages multiples, des méthodes globales intégrant une contrainte d'occultation ou des méthodes avec affinement.

Dans le chapitre 3, nous avons mis en évidence le fait que la principale difficulté des méthodes de mise en correspondance par corrélation concerne les occultations (cf. paragraphe 3.2.3). Dans le paragraphe 3.3.5.1, pour prendre en compte cette difficulté, nous avons proposé de nouvelles mesures de corrélation robustes aux occultations. Cependant, les résultats que nous avons obtenus avec ces mesures sont les meilleurs dans les zones des occultations que nous considérons alors que dans les zones sans occultation, ces mesures n'obtiennent pas toujours les meilleurs résultats. Il serait donc intéressant de pouvoir utiliser ces nouvelles mesures uniquement dans les zones des occultations.

5.1.1 Objectifs

Ainsi, nous souhaitons proposer de nouvelles méthodes s'appuyant sur deux mesures de corrélation :

- une mesure de corrélation des familles CROISÉE ou CLASSIQUE à utiliser dans les zones sans occultation ;
- une mesure de corrélation de la famille ROBUSTE à utiliser dans la zone des occultations et plus particulièrement dans la zone d'influence des occultations (définitions au paragraphe 2.9.4).

Il faut donc déterminer, d'une part, les mesures à utiliser et, d'autre part, la technique de détection des zones des occultations, le dernier problème étant le plus délicat. Il existe de très nombreuses techniques qui prennent en compte les occultations, pour la segmentation, comme dans [Triantafyllidis 00], pour la classification, comme dans [Withagen 04], et pour la stéréovision, notamment dans le cas de la stéréovision multi-oculaire comme dans [Nakamura 96, Satoh 96, Gong 02, Yao 03, Ju 04]. Avant d'exposer notre approche, nous allons effectuer un bref état de l'art de ces méthodes en nous restreignant à celles qui sont utilisables dans le cas de la stéréovision binoculaire.

5.1.2 Plan

Nous allons tout d'abord répertorier les méthodes qui prennent en compte le problème des occultations en les classant en quatre catégories : les méthodes avec prétraitements, les méthodes locales, les méthodes globales avec un coût d'occultation, les méthodes locales à passages multiples et les méthodes avec affinement. Puis, nous exposerons notre approche qui s'appuie sur une méthode locale à passages multiples.

5.2 Méthodes avec prétraitements

Une première approche, simple et assez courante [Hoff 89, Sankar Kishore 01, Ghafoor 03], consiste à utiliser un masque des occultations. Une première étude des images permet de déterminer les pixels occultés et de construire un masque des occultations. Les méthodes diffèrent par leur façon de le calculer mais, généralement, elles s'appuient sur la même hypothèse : les occultations se situent au niveau des contours de type « marche ». En effet, dans la plupart des cas, les discontinuités de profondeur responsables des occultations, se trouvent aux frontières des objets. Le principe est donc de détecter les contours dans l'image.

Les techniques habituelles sont :

- le seuillage de la norme du vecteur gradient (avec les opérateurs de Sobel, Prewitt ou Frei-Chen [Pratt 78, p. 503], par exemple) ;
- la détection des passages par zéro de la dérivée seconde dans la direction du gradient (en utilisant la méthode SDEF de Shen et Castan [Shen 92], par exemple).

Une seconde méthode qui tend à se développer est d'utiliser une segmentation initiale [Tsai 99, Lee 02, Zhang 02, Dante 03, Sun 03, Cucchiara 04, Wei 04, Zhang 04, Deng 05]. Les pixels proches des frontières des régions de l'image sont considérés comme occultés. La méthode de Deng *et al.* [Deng 05] est différente. Elle utilise le principe suivant :

Une discontinuité dans une image implique une occultation dans l'autre image (on suppose qu'il y a eu une étape d'initialisation des correspondances).

5.3 Méthodes locales

Une autre manière de détecter les occultations est d'utiliser les transformations locales proposées dans [Kaneko 02, Kaneko 03] et présentées aux paragraphes 3.3.4.3 et 3.3.4.4 ou celles proposées dans [Zabih 94] présentées au paragraphe 3.3.4.5, ou les méthodes locales à base de statistiques robustes de Lan [Lan 97] et présentées au paragraphe 3.3.5.2. Banks et Bennamoun [Banks 01] utilisent aussi la transformation de Zabih et Woodfill [Zabih 94] : en utilisant une mesure de dissimilarité, à chaque recherche de correspondance, s'il existe plusieurs minima locaux, alors la contrainte de rang est appliquée pour qu'il n'y ait plus d'ambiguïté. Enfin, Chen *et al.* [Chen 03] utilisent des mesures robustes s'appuyant sur des M-estimateurs (cf. paragraphe 3.3.5.11).

5.4 Méthodes globales

Les méthodes globales qui prennent en compte le problème des occultations ajoutent une contrainte d'occultation dans le coût des contraintes (défini dans le paragraphe 1.9.1 par les équations (1.16) et (1.17)). Nous pouvons distinguer les différents coûts suivants :

- *Utilisation de la limite du gradient de disparité* – Barnard [Barnard 89] utilise la limite du gradient de disparité pour prendre en compte les occultations, c'est-à-dire que le coût de lissage

qui intervient dans l'équation (1.17) est donné par :

$$E_{\text{lissage}}^c(\mathbf{p}_g^{i,j}, \mathbf{p}_g^{i',j'}) = G_d(\mathbf{p}_g^{i,j}, \mathbf{p}_d^{i,v}, \mathbf{p}_g^{i',j'}, \mathbf{p}_d^{i',v'}), \quad (5.1)$$

avec G_d défini par l'équation (1.10), $\mathbf{p}_d^{i,v} = \mathbf{p}_g^{i,j} + d(\mathbf{p}_g^{i,j})$ et $\mathbf{p}_d^{i',v'} = \mathbf{p}_g^{i',j'} + d(\mathbf{p}_g^{i',j'})$.

- *Utilisation d'un coût de pénalité* – Dans [Belhumeur 92, Luo 95, Ishikawa 98, Torr 98, Shafique 03, Agrawal 04, Torr 04, Bleyer 04, Deng 05], les auteurs utilisent un terme de pénalité (défini dans le paragraphe 1.9.3.2 par l'équation (1.27)). Belhumeur et Mumford [Belhumeur 92] appliquent la contrainte de symétrie. Ishikawa et Geiger [Luo 95, Ishikawa 98] exploitent la contrainte d'unicité pour attribuer des pénalités. Agrawal et Davis [Agrawal 04] s'appuient sur une comparaison avec les disparités des pixels voisins. Torr et Criminisi [Torr 04] utilisent un coût de pénalité proportionnel au saut de disparité entre deux pixels consécutifs sur la droite épipolaire. Bleyer et Gelautz [Bleyer 04] attribuent la pénalité lorsque le pixel considéré a été préalablement détecté comme occulté ou proche d'une discontinuité (une segmentation initiale sert comme détection des occultations et des discontinuités).

- *Utilisation d'un coût de lissage* – Dans [Intille 94, Dhond 95, Geiger 95, Bobick 99, Kang 01, Brockers 04a, Brockers 04b, Deng 05], les auteurs ajoutent un coût de lissage. Intille et Bobick [Intille 94, Bobick 99] utilisent le principe suivant :

Un pixel qui possède un de ses quatre voisins avec une disparité différente a un coût de lissage élevé. Ce coût de lissage est moins important si ce pixel est un point d'intérêt (coins, points contours).

Geiger *et al.* [Geiger 95] étudient les disparités le long de la droite épipolaire. À chaque saut de disparité, le pixel correspondant possède un coût de lissage proportionnel au saut de disparité. Dans [Dhond 95, Kang 01, Brockers 04a, Brockers 04b, Deng 05], les auteurs utilisent un coût de lissage dont la forme suit celle de l'équation (1.26), c'est-à-dire qu'ils attribuent un coût de lissage qui est proportionnel à la différence de disparité entre le pixel étudié et les pixels de la zone d'agrégation.

- *Modification du coût de voisinage* – Dans [Kolmogorov 01, Kolmogorov 02, Gong 04], les auteurs utilisent un coût de voisinage donné par :

$$E_{\text{voisinage}}^c(\mathbf{p}_g^{i,j}) = \min_{\mathbf{p}_g^{i',j'} \in \text{ZAC}(\mathbf{p}_g^{i,j})} \|d(\mathbf{p}_g^{i',j'}) - d(\mathbf{p}_g^{i,j})\|. \quad (5.2)$$

- *Utilisation d'un modèle de données aberrantes* – Hasler *et al.* [Hasler 03] proposent un modèle des données aberrantes et des données correctes. Ces modèles sont utilisés au cours de la mise en correspondance pour prédire le risque d'erreur et conditionner les calculs. Les auteurs font l'hypothèse suivante : en supposant que l'erreur engendrée par des données aberrantes est semblable à l'erreur produite en comparant deux régions aléatoirement sélectionnées dans les deux images étudiées, un modèle de données aberrantes peut être construit en s'appuyant sur le contenu des deux images.

5.5 Méthodes locales à passages multiples

Nous pouvons distinguer les méthodes qui utilisent des fenêtres adaptatives, des méthodes qui utilisent des mesures de corrélation pondérées et des méthodes hybrides.

5.5.1 Fenêtres adaptatives

L'utilisation de fenêtres adaptatives est assez populaire [Okutomi 92a, Kanade 94, Lotti 94, Geiger 95, Deverny 97, Fusiello 97a, Paparoditis 98, Scharstein 98, Garcia 01a, Park 01,

Hirschmüller 02, Kostková 02, Zhang 02, Veksler 03, Yoon 05]. Dans la méthode de Lotti et Giraudon [Lotti 94], la taille des fenêtres est déterminée à partir d'une carte de contours : la fenêtre ne doit pas contenir de contour. Dans les méthodes de Devernay [Devernay 97, chap. 3] et de Garcia [Garcia 01a, chap. 2 et 3], une carte de disparités initiale (obtenue par une méthode par corrélation) est utilisée pour estimer l'ensemble des paramètres des transformations locales des fenêtres de corrélation. Une deuxième mise en correspondance est réalisée en prenant en compte ces transformations. Le principe des autres méthodes est de faire varier la taille et la forme des fenêtres de corrélation en fonction de la variation locale des niveaux de gris ainsi que de la disparité. Pour établir la taille de la fenêtre, il est nécessaire d'estimer la disparité ainsi qu'une fonction de coût liée à cette disparité. Cette technique est décrite par l'algorithme 5.1. La mise à jour effectuée à l'étape A de cet algorithme peut être réalisée : en tenant compte d'une mesure d'incertitude [Okutomi 92a, Kanade 94], en observant les disparités des pixels voisins [Hirschmüller 02, Kostková 02, Park 01], en testant seulement un certain nombre de fenêtres [Fusiello 97a, Veksler 03]. Ces méthodes sont souvent coûteuses en temps de calcul mais la version qui a été proposée par Fusiello *et al.* [Fusiello 97a] (seulement neuf formes de fenêtres sont testées) est beaucoup plus rapide que les autres méthodes.

Pour chaque passage faire

2.1 Prétraiter les images

2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$

2.3 Pour chaque $\mathcal{S}^{i,j}$ faire

2.3.1 Tant que le correspondant de $\mathbf{p}_g^{i,j}$ n'est pas fiable faire

A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire

Si passage > 0 alors mettre à jour la forme de la fenêtre

Calculer $\text{Mes}(\mathbf{f}_g, \mathbf{f}_d)$

F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA

2.4 Affiner les résultats

Les différents passages :

Passage 0 -- Estimation de la disparité initiale

Passage 1 -- Estimation de la disparité finale en utilisant la méthode des fenêtres adaptatives

ALGO. 5.1 – *Mise en correspondance par corrélation avec des fenêtres adaptatives* – Il s'agit d'une instance de l'algorithme 1.2, à passages multiples, avec un premier passage qui permet d'obtenir la disparité initiale. L'exemple donné ici est sans approche multirésolution (pour simplifier l'écriture), mais cet algorithme peut aussi être combiné à une approche multirésolution.

5.5.2 Mesures pondérées

Il existe de nombreuses méthodes avec des mesures de corrélation pondérées. Parmi ces méthodes, celle de Zoghلامي *et al.* [Zoghلامي 96] s'appuie sur le calcul de deux cartes de disparités : la première carte de disparités permet de calculer les poids qui vont être attribués à chaque pixel et la deuxième carte est calculée en utilisant une mesure de corrélation pondérée. Nous détaillons cette technique au paragraphe 3.3.5.2.

5.5.3 Méthodes hybrides

Ces méthodes consistent en l'application de plusieurs méthodes en séquence :

- *Méthode avec étude de disparités sur la droite épipolaire* – Weng *et al.* [Weng 92] effectuent une estimation initiale des disparités. Sur chaque droite épipolaire, les variations de ces disparités sont étudiées car les auteurs font l'hypothèse suivante : si le sens de variation des disparités change, cela implique une occultation dans l'image gauche ou dans l'image droite. Par exemple, pour les occultations de l'image gauche vers l'image droite, ils prennent en compte les disparités calculées de l'image droite vers l'image gauche. À l'endroit où les disparités commencent à diminuer, cela implique qu'il y a une occultation dans l'image gauche.
- *Méthodes locales coopératives* – Jones et Malik [Jones 92] utilisent une méthode locale coopérative. Après la première étape, pour chaque étape suivante, la mesure de corrélation utilisée intègre un coût de pénalité qui dépend des disparités des pixels voisins. Mayers [Mayer 03] effectue une première mise en correspondance avec la mesure NCC. Grâce aux résultats obtenus, un algorithme coopératif défini comme dans [Zitnick 00] (cf. paragraphe 1.11.2) est appliqué avec une mesure de corrélation qui combine les mesures SAD et NCC. Cette technique est plus robuste dans les zones des occultations. Park *et al.* [Park 03] s'appuient sur le même principe mais la mesure utilisée est différente. Elle prend en compte la variance des disparités dans le voisinage.
- *Méthodes avec fenêtres adaptatives* – Dans [Kostková 02, Kostková 03], après la première mise en correspondance, une partie des pixels est sélectionnée pour définir les composants de disparités. Ils correspondent à des régions où la disparité est la même pour tous les pixels. Pour cela, le voisinage de chaque pixel est examiné (deux méthodes sont utilisées, une méthode qui examine les huit voisins et une méthode qui peut prendre en compte jusqu'à vingt voisins). La deuxième étape prend en compte ces composants pour recalculer la mesure de corrélation, la zone d'agrégation respectant la forme des composants de disparités (cf. chapitre 3). Eklund et Farag [Eklund 03] appliquent un seuil sur les scores de corrélation obtenus lors de la première étape. Ils utilisent une mesure de dissimilarité, c'est pourquoi, c'est seulement pour les pixels qui se trouvent au dessus de ce seuil que la mise en correspondance va être effectuée une deuxième fois. Pour la deuxième mise en correspondance, des formes de fenêtres de corrélation variables sont utilisées (fenêtres adaptatives, cf. paragraphe 5.5.1).

5.6 Méthodes avec affinement

Ce sont les méthodes les plus populaires, elles permettent de détecter les occultations et nous avons répertorié les techniques suivantes :

- *Utilisation de la contrainte de symétrie* – Dans [Little 92, Faugeras 93, Fua 93, Arcara 00, Egnal 00, Mühlmann 01, Alvarez 02, Lin 02], les auteurs exploitent la contrainte de symétrie : tout pixel ne respectant pas la contrainte de symétrie est considéré comme occulté. C'est une des techniques les plus employées pour détecter les pixels occultés.
- *Utilisation de l'ambiguïté* – Dans [Fua 93, Manduchi 99, Silva 00], les auteurs se placent dans le cadre de la mise en correspondance par corrélation et suggèrent d'utiliser une mesure qui évalue l'ambiguïté d'une correspondance. Ils utilisent une mesure de similarité et étudient la courbe des scores obtenue sur la zone de recherche : moins le pic principal se distingue, plus la mesure est ambiguë et au delà d'un seuil, le pixel est considéré comme occulté (les mesures utilisées sont similaires à celles présentées au paragraphe 3.6.1).
- *Utilisation de la contrainte d'ordre* – Dans [Dhond 95, Koch 98, Manduchi 99, Egnal 00, Zitnick 00, Gong 03], les auteurs utilisent la contrainte d'ordre de la même manière que la

contrainte de symétrie.

- *Utilisation de la contrainte d'unicité* – Dans [Koch 98, Mühlmann 01, Di Stefano 02, Zhang 02, Gong 03, Kostková 03], les auteurs exploitent la contrainte d'unicité. Tout pixel ne respectant pas la contrainte d'unicité est considéré comme occulté.
- *Étude des disparités voisines* – Egnal et Wildes [Egnal 00] calculent un histogramme local des disparités pour un pixel donné. Si cet histogramme contient deux modes, alors le pixel est considéré comme appartenant à une zone des occultations. Maier *et al.* [Maier 03] exploitent aussi cette technique. Mayer [Mayer 03] observe la disparité de chaque pixel, sur une même droite épipolaire, et la compare avec celle du pixel précédent. Si les deux disparités comparées ne semblent pas compatibles, le pixel observé est considéré comme occulté.
- *Seuillage des scores de corrélation* – Dans [Szeliski 02, Kong 03, Woetzel 04], les auteurs s'appuient sur l'hypothèse suivante : tous les pixels qui possèdent un score de corrélation en dessous d'un seuil sont considérés comme appartenant à une zone des occultations.
- *Utilisation d'un intervalle de variation* – El-Sonbaty et Ismail [El-Sonbaty 03] fixent un intervalle de variation pour les disparités (en prenant en considération seulement un échantillon de la carte de disparités), puis ils examinent toute l'image : tout pixel ayant une disparité n'appartenant pas à cet intervalle de variation est considéré comme occulté.

5.7 Méthodes hybrides utilisant deux mesures de corrélation

5.7.1 Zones des occultations considérées

Les zones des occultations que nous considérons dans ce travail correspondent à celles introduites dans le paragraphe 2.9.4. Le but de cette étude est de proposer une méthode qui améliore les résultats dans les zones d'influence des occultations (zones grisées sur les images (a) et (e) de la figure 2.12). L'approche que nous avons retenue et mise en œuvre [Chambon 04b, Chambon 04c] s'appuie sur l'utilisation de deux mesures de corrélation :

- Une mesure des familles CROISÉE ou CLASSIQUE qui est utilisée dans les zones sans occultation. Nous avons testé NCC, ZNCC et SAD.
- Une mesure de la famille ROBUSTE qui est utilisée dans la zone totale des occultations. Nous avons testé ME_m , LTP_2 et $SMPD_2$.

C'est à partir des résultats que nous avons obtenus avec les mesures de corrélation que nous avons choisi d'utiliser celles-ci (cf. paragraphe 3.6). La difficulté de cette approche réside dans la tâche délicate qui consiste à déterminer la zone totale des occultations.

Les algorithmes de mise en correspondance que nous proposons sont de trois formes différentes, selon le moment où la détection de la zone totale des occultations est réalisée. Nous notons d , d_{cla} et d_{rob} les fonctions de disparité obtenues respectivement avec les méthodes proposées, la mesure de corrélation classique, notée Mes_{cla} , et la mesure de corrélation robuste, notée Mes_{rob} . De manière générale nous utilisons « cla » en indice lorsqu'il s'agit d'un calcul effectué avec la mesure de corrélation des familles CROISÉE ou CLASSIQUE et « rob » en indice lorsqu'il s'agit de la famille ROBUSTE.

5.7.2 Méthodes utilisant les contours

Le principe de ces méthodes, cf. algorithme 5.2, est d'effectuer une détection de contours et de considérer les pixels détectés comme des pixels de la zone totale des occultations. Ensuite, nous utilisons la mesure de la famille ROBUSTE dans cette zone et l'autre mesure dans le reste de l'image. Pour le

prétraitement, c'est-à-dire l'étape 2.1 de l'algorithme 5.2, l'équation (2.13) devient :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} \text{ correspond à un point contour} \\ 0 & \text{sinon.} \end{cases} \quad (5.3)$$

Dans ce rapport, nous n'avons considéré que des contours de type « marche » et différents filtres ont été utilisés pour les détecter. Nous distinguons les deux méthodes suivantes :

- Méthode CONTOUR-SOBEL – Un seuillage de la norme du vecteur gradient est calculé par les masque de Sobel.
- Méthode CONTOUR-SDEF – Les contours sont détectés par passage par zéro de la dérivée seconde en utilisant la méthode SDEF de Shen et Castan [Shen 92].

Pour chaque passage faire

2.1 Pour chaque $\mathbf{p}_g^{i,j}$ faire calculer $\text{PO}(\mathbf{p}_g^{i,j})$, cf. équation (5.3)

2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$

2.3 Pour chaque $\mathcal{S}^{i,j}$ faire

A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire calculer $\text{Mes}(\mathbf{F}_g, \mathbf{F}_d)$

$$\text{Mes}(\mathbf{F}_g, \mathbf{F}_d) = \begin{cases} \text{Mes}_{\text{cla}}(\mathbf{F}_g, \mathbf{F}_d) & \text{si } \text{PO}(\mathbf{p}_g^{i,j}) = 0 \\ \text{Mes}_{\text{rob}}(\mathbf{F}_g, \mathbf{F}_d) & \text{sinon} \end{cases} \quad (5.4)$$

F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA

2.4 Si passage l alors appliquer la contrainte de symétrie puis calculer les disparités au sous-pixel

Les différents passages :

Passage 0 -- Mise en correspondance gauche -> droite

Passage 1 -- Mise en correspondance droite -> gauche

ALGO. 5.2 – Méthode CONTOUR – Détection de la zone totale des occultations avant la mise en correspondance en utilisant les contours (prétraitement). Cet algorithme est une instance de l'algorithme 1.2.

5.7.3 Méthodes utilisant le principe des mesures de corrélation pondérée

Le principe de cette méthode, cf. algorithme 5.3, est de déterminer au moment du calcul du score de corrélation, la mesure à utiliser. Cette détermination s'effectue en étudiant l'ensemble des pixels dans les fenêtres de corrélation. Voici la solution que nous avons envisagée :

À chaque calcul du score de corrélation, un poids (déterminé grâce à un outil des statistiques ordinales ou robustes) est attribué à chaque pixel du voisinage du pixel étudié. Suivant les valeurs de ces poids, nous utilisons soit la mesure classique, soit la mesure robuste.

Les deux méthodes que nous avons proposées sont :

- Méthode CORRELATION-KANEKO – Nous utilisons la méthode ordinale de Kaneko *et al.* [Kaneko 02].
- Méthode CORRELATION-LAN – Nous utilisons LMS, comme dans [Lan 97] (nous utilisons cette approche sur les niveaux de gris car nous ne l'avons pas généralisée à la couleur).

Pour chaque passage faire

2.2 $\mathcal{S}^{i,j} = \{\mathbf{p}_g^{i,j}\}$

2.3 Pour chaque $\mathcal{S}^{i,j}$ faire

A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire // Calculer $\text{Mes}(\mathbf{F}_g, \mathbf{F}_d)$

A.1. Évaluer $\text{PO}(\mathbf{p}_g^{i,j})$ (méthode de Kaneko ou de Lan)

A.2. Calculer le score

$$\text{Mes}(\mathbf{F}_g, \mathbf{F}_d) = \begin{cases} \text{Mes}_{\text{cla}}(\mathbf{F}_g, \mathbf{F}_d) & \text{si } \text{PO}(\mathbf{p}_g^{i,j}) = 0 \\ \text{Mes}_{\text{rob}}(\mathbf{F}_g, \mathbf{F}_d) & \text{sinon} \end{cases} \quad (5.5)$$

F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA

2.4 Si passage 1 alors appliquer la contrainte de symétrie puis calculer les disparités au sous-pixel

Les différents passages :

Passage 0 -- Mise en correspondance gauche -> droite

Passage 1 -- Mise en correspondance droite -> gauche

ALGO. 5.3 – Méthode CORRÉLATION – Détection de la zone totale des occultations pendant la mise en correspondance. Cet algorithme est une instance de l'algorithme 1.2. Il ne contient pas de prétraitement (disparition de l'étape 2.1). Le choix de la mesure de corrélation s'effectue au moment du calcul du score de corrélation, étape A.1.

5.7.4 Méthodes avec post-détection

Le principe des méthodes avec post-détection est d'effectuer une première mise en correspondance en utilisant une mesure des familles CROISÉE ou CLASSIQUE, puis de déterminer la zone totale des occultations et enfin d'effectuer une seconde mise en correspondance partielle (sur la zone totale des occultations) en utilisant une mesure de la famille ROBUSTE.

Pour le post-traitement, étape 2.1 de l'algorithme 5.4 (cf. figure 5.1), les solutions que nous avons envisagées pour détecter les pixels qui se trouvent dans la zone totale des occultations sont :

- Méthode POST-SEUIL – Un seuil T_S est appliqué au score de corrélation obtenu avec la mesure de corrélation classique :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \text{Mes}_{\text{cla}}(\mathbf{p}_g^{i,j}) > T_S \\ 0 & \text{sinon.} \end{cases} \quad (5.6)$$

Le seuil T_S dépend de la mesure utilisée.

- Méthode POST-AMBIGUÏTÉ – Il faut appliquer un seuil T_A à la mesure d'ambiguïté définie par De Joinville *et al.* [De Joinville 01] (cf. paragraphe 3.6.1). Nous notons $A_{\text{cla}}(\mathbf{p}_g^{i,j})$ l'ambiguïté obtenue pour le pixel $\mathbf{p}_g^{i,j}$ lorsque la mesure de corrélation classique a été utilisée :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } A_{\text{cla}}(\mathbf{p}_g^{i,j}) > T_A \\ 0 & \text{sinon.} \end{cases} \quad (5.7)$$

Dans cette étude, le seuil T_A correspond à 60% de la valeur maximale de l'ambiguïté.

- Méthode POST-SYMÉTRIE – Nous utilisons la contrainte de symétrie, cf. équation (5.11) (cf. paragraphe 1.6.7) : $\text{PO}(\mathbf{p}_g^{i,j}) = \text{O}(\mathbf{p}_g^{i,j})$.

- Méthode POST-UNICITÉ – La contrainte d’unicité est utilisée (cf. paragraphe 1.6.5) :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} + d_{\text{cla}}(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ et } \exists j' \neq j \mid \mathbf{p}_g^{i,j'} + d_{\text{cla}}(\mathbf{p}_g^{i,j'}) = \mathbf{p}_d^{i,v} \\ 0 & \text{sinon.} \end{cases} \quad (5.8)$$

- Méthode POST-ORDRE – La contrainte d’ordre est utilisée (cf. paragraphe 1.6.6) :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} + d_{\text{cla}}(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ et } \exists j' \neq j \text{ avec } \mathbf{p}_g^{i,j'} + d_{\text{cla}}(\mathbf{p}_g^{i,j'}) = \mathbf{p}_d^{i,v'} \mid \\ & (j - j')(v - v') < 0 \\ 0 & \text{sinon.} \end{cases} \quad (5.9)$$

Pour chaque passage faire

2.1 Si passage 2 ou passage 3 alors

Pour chaque $\mathbf{p}_g^{i,j}$ faire calculer $\text{PO}(\mathbf{p}_g^{i,j})$

2.2 $S^{i,j} = \{\mathbf{p}_g^{i,j}\}$

Si passage 0 ou 1 alors prendre en compte tous les pixels sinon

Si $\text{PO}(\mathbf{p}_g^{i,j}) = 1$ ou $\text{O}(\mathbf{p}_g^{i,j}) = 1$ alors prendre en compte $\mathbf{p}_g^{i,j}$

2.3 Pour chaque $S^{i,j}$ faire

A. Pour chaque $\mathbf{p}_d^{i,v} \in Z_d(\mathbf{p}_g^{i,j})$ faire calculer $\text{Mes}(\mathbf{F}_g, \mathbf{F}_d)$

$$\text{Mes}(\mathbf{F}_g, \mathbf{F}_d) = \begin{cases} \text{Mes}_{\text{cla}}(\mathbf{F}_g, \mathbf{F}_d) & \text{si passage} < 2 \\ \text{Mes}_{\text{rob}}(\mathbf{F}_g, \mathbf{F}_d) & \text{sinon} \end{cases} \quad (5.10)$$

F. Déterminer le correspondant de $\mathbf{p}_g^{i,j}$ selon la méthode WTA

2.4 Si passage 1 alors calculer

$$\text{O}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} + d_{\text{cla}}(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ et } \mathbf{p}_d^{i,v} + d_{\text{cla}}(\mathbf{p}_d^{i,v}) \neq \mathbf{p}_g^{i,j} \\ 0 & \text{sinon} \end{cases} \quad (5.11)$$

Si passage 3 alors appliquer la contrainte de symétrie puis calculer les disparités au sous-pixel

Les différents passages :

Passage 0 -- Mise en correspondance gauche -> droite en utilisant Mes_{cla}

Passage 1 -- Mise en correspondance droite -> gauche en utilisant Mes_{cla}

Passage 2 -- Mise en correspondance gauche -> droite en utilisant Mes_{rob}

Passage 3 -- Mise en correspondance droite -> gauche en utilisant Mes_{rob}

ALGO. 5.4 – Méthode POST-DÉTECTION – Détection de la zone totale des occultations après la mise en correspondance. Cet algorithme effectue quatre passages. Les deux derniers passages consistent en une mise en correspondance partielle. Cet algorithme est une instance de l’algorithme 1.2.

La méthode POST-SYMMÉTRIE utilisant la contrainte de symétrie semble la plus prometteuse car il n’y a pas de seuil à fixer et les cartes de disparités que nous avons obtenues dans [Chambon 03] montrent que la contrainte de symétrie détecte de nombreux pixels comme occultés parmi les pixels proches des pixels réellement occultés. Le but de la méthode proposée est d’améliorer les résultats dans la zone d’influence des occultations, or la contrainte de symétrie détecte des pixels dans la

zone d'influence des occultations mais aussi les pixels véritablement occultés. Ainsi la méthode POST-SYMMÉTRIE peut détériorer les résultats obtenus par la mesure classique. C'est pourquoi, nous proposons d'autres variantes qui utilisent aussi la contrainte de symétrie :

- Méthode POST-MÉDIAN – Le principe est d'utiliser la contrainte de symétrie et d'effectuer un filtrage médian sur la carte de disparités initiale. Les deux règles suivantes sont utilisées :
 - Un pixel occulté dont le nombre de voisins occultés est supérieur à un seuil T_{F_1} est considéré comme appartenant à la zone d'influence des occultations.
 - Un pixel non occulté dont le nombre de voisins occultés est supérieur à un seuil T_{F_2} est considéré comme appartenant à la zone d'influence des occultations.

L'évaluation de $\text{PO}(\mathbf{p}_g^{i,j})$ devient :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \left(\text{O}(\mathbf{p}_g^{i,j}) = 1 \text{ et } \text{V}(\mathbf{p}_g^{i,j}) > T_{F_1} \right) \text{ ou } \left(\text{O}(\mathbf{p}_g^{i,j}) = 0 \text{ et } \text{V}(\mathbf{p}_g^{i,j}) > T_{F_2} \right) \\ 0 & \text{sinon.} \end{cases} \quad (5.12)$$

Dans cette étude, nous avons pris $T_{F_1} = \frac{N_f}{10}$ et $T_{F_2} = \frac{N_f}{7}$.

- Méthode POST-DILATATION – La contrainte de symétrie est appliquée et nous calculons le dilaté, obtenu par dilatation conditionnelle au sens de la morphologie mathématique, par la fenêtre de corrélation, des zones des occultations de la carte de disparités initiale. La dilatation est conditionnée par le nombre de pixels occultés dans la fenêtre de corrélation :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \text{O}(\mathbf{p}_g^{i,j}) = 0 \text{ et } \text{V}(\mathbf{p}_g^{i,j}) > T_D \\ 0 & \text{sinon.} \end{cases} \quad (5.13)$$

Nous avons pris : $T_D = \frac{3N_f}{10}$. Lorsque la carte de disparités initiale possède de nombreux faux négatifs, cet algorithme risque de détecter trop de pixels occultés, c'est pourquoi nous proposons aussi la variante POST-DILATATION-CONTOUR.

- Méthode POST-DILATATION-CONTOUR – La contrainte de symétrie est appliquée et nous calculons le dilaté, obtenu par dilatation conditionnelle au sens de la morphologie mathématique, par la fenêtre de corrélation, des zones des occultations de la carte de disparités initiale. La dilatation est conditionnée par le nombre de pixels occultés et le nombre de points contour dans la fenêtre de corrélation. En prenant en compte les contours de l'image, nous limitons la dilatation des pixels occultés et nous supposons que nous pouvons faire confiance aux pixels détectés comme proches de pixels occultés lorsque ceux-ci sont proches d'un contour :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \text{V}(\mathbf{p}_g^{i,j}) > T_{C_1} \text{ et } \text{PF}(\mathbf{p}_g^{i,j}) > T_{C_2} \\ 0 & \text{sinon,} \end{cases} \quad (5.14)$$

$$\text{avec } \text{PF}(\mathbf{p}_g^{i,j}) = \sum_{p=-N_v}^{N_v} \sum_{q=-N_h}^{N_h} \text{F}(\mathbf{p}_g^{i,j}) \quad (5.15)$$

$$\text{et } \text{F}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \mathbf{p}_g^{i,j} \text{ est un pixel contour} \\ 0 & \text{sinon.} \end{cases} \quad (5.16)$$

Pour les seuils, nous avons pris : $T_{C_1} = \frac{3N_f}{10}$ et $T_{C_2} = \frac{N_f}{20}$.

- Méthode POST-ORDRE-SYMMÉTRIE – Nous combinons la contrainte d'ordre et de symétrie :

$$\text{PO}(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } \left(\mathbf{p}_g^{i,j} + d_{\text{cla}}(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ et } \exists j' \neq j \text{ avec } \mathbf{p}_g^{i,j'} + d_{\text{cla}}(\mathbf{p}_g^{i,j'}) = \mathbf{p}_d^{i,v'} \mid \right. \\ & (j - j')(v - v') < 0 \\ & \text{ou } \left(\mathbf{p}_g^{i,j} + d_{\text{cla}}(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ et } \mathbf{p}_d^{i,v} + d_{\text{cla}}(\mathbf{p}_d^{i,v}) \neq \mathbf{p}_g^{i,j} \right) \\ 0 & \text{sinon.} \end{cases} \quad (5.17)$$

- Méthode POST-AMBIGUÏTÉ-SYMMÉTRIE – Nous combinons l'utilisation d'un seuil sur l'ambiguïté et la contrainte de symétrie :

$$PO(\mathbf{p}_g^{i,j}) = \begin{cases} 1 & \text{si } (A_{\text{cla}}(\mathbf{p}_g^{i,j}) > T_A) \text{ ou } (\mathbf{p}_g^{i,j} + d_{\text{cla}}(\mathbf{p}_g^{i,j}) = \mathbf{p}_d^{i,v} \text{ et } \mathbf{p}_d^{i,v} + d_{\text{cla}}(\mathbf{p}_d^{i,v}) \neq \mathbf{p}_g^{i,j}) \\ 0 & \text{sinon.} \end{cases} \quad (5.18)$$

Ici aussi, le seuil T_A correspond à 60% de la valeur maximale de l'ambiguïté.

5.7.5 Méthodes de fusion de résultats

Enfin, un quatrième type de méthodes est envisagé, la méthode FUSION-DISP qui consiste à calculer les cartes de disparités avec une mesure classique et une mesure robuste et à fusionner les résultats obtenus. Quand un pixel est occulté, nous notons sa disparité **occ**. Pour fusionner les résultats obtenus, les règles retenues sont les suivantes :

- Si les deux mesures de corrélation donnent le même résultat (pixel occulté ou un même correspondant), alors ce résultat est conservé.
- Si une seule des deux mesures détecte le pixel comme occulté, alors les résultats obtenus avec cette mesure dans le voisinage de ce pixel sont examinés. Nous supposons que s'il y a de nombreux pixels détectés comme occultés dans son voisinage alors il y a une forte probabilité pour que ce pixel soit occulté, sinon il s'agit d'un faux négatif. Plus précisément, si plus de la moitié des voisins sont occultés alors le pixel est occulté sinon la disparité obtenue avec l'autre mesure est conservée :

$$d(\mathbf{p}_g^{i,j}) = \begin{cases} \mathbf{occ} & \text{si } \left(O_{\text{cla}}(\mathbf{p}_g^{i,j}) = 1 \text{ et } V_{\text{cla}}(\mathbf{p}_g^{i,j}) > \frac{N_f}{2} \right) \text{ ou} \\ & \left(O_{\text{rob}}(\mathbf{p}_g^{i,j}) = 1 \text{ et } V_{\text{rob}}(\mathbf{p}_g^{i,j}) > \frac{N_f}{2} \right) \\ d_{\text{cla}}(\mathbf{p}_l^{i,j}) & \text{si } O_{\text{rob}}(\mathbf{p}_g^{i,j}) = 1 \\ d_{\text{rob}}(\mathbf{p}_l^{i,j}) & \text{sinon.} \end{cases} \quad (5.19)$$

- Si les deux mesures donnent une disparité différente, nous supposons qu'il y a une forte probabilité pour que ce pixel possède un correspondant mais, en revanche, ce pixel se situe dans une zone difficile à apparier (par exemple, une zone uniforme). De plus, nous pensons que plus il y a de pixels détectés comme occultés dans son voisinage, qui sont probablement des faux négatifs, plus la mesure utilisée est ambiguë, c'est pourquoi nous privilégions la mesure qui a donné le moins de pixels occultés dans le voisinage du pixel :

$$d(\mathbf{p}_g^{i,j}) = \begin{cases} d_{\text{cla}}(\mathbf{p}_l^{i,j}) & \text{si } V_{\text{rob}}(\mathbf{p}_g^{i,j}) > V_{\text{cla}}(\mathbf{p}_g^{i,j}) \\ d_{\text{rob}}(\mathbf{p}_l^{i,j}) & \text{sinon.} \end{cases} \quad (5.20)$$

5.7.6 Récapitulatif des méthodes proposées

Tous les algorithmes proposés ainsi que les différentes variantes sont récapitulés dans le tableau 5.1.

5.8 Résultats expérimentaux

5.8.1 Protocole d'évaluation

5.8.1.1 Images testées

Les images testées sont toutes celles présentées au paragraphe 2.3.4.

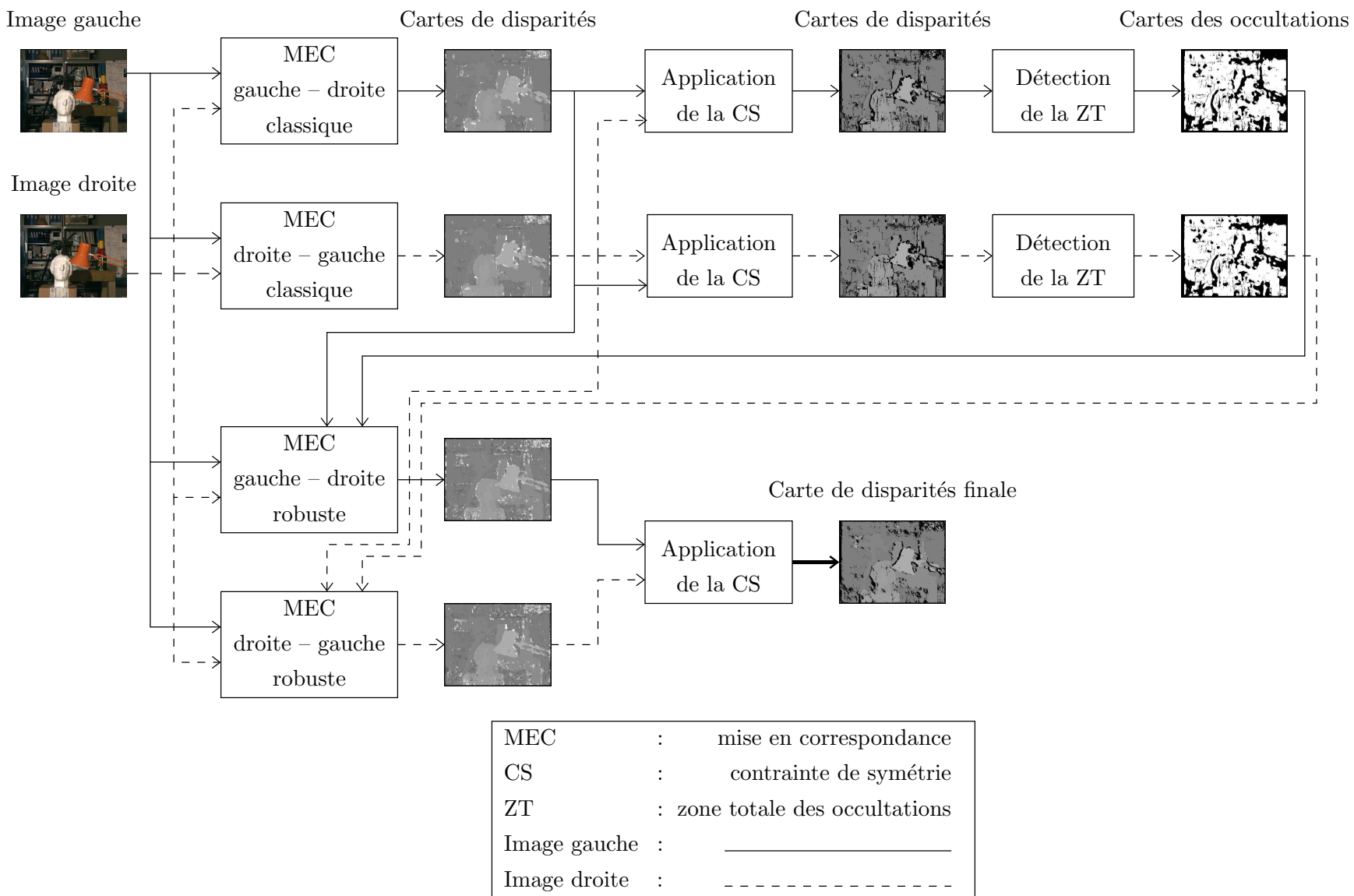


FIG. 5.1 – Méthode POST-DÉTECTION – Détection des occultations à partir d'une carte de disparités initiale.

TYPE	VARIANTE	CARACTÉRISTIQUES
CONTOUR	CONTOUR-SOBEL	Utilisation du filtre de Sobel
	CONTOUR-SDEF	Utilisation de la méthode SDEF
CORRÉLATION	CORRÉLATION-KANEKO	Utilisation de la méthode de Kaneko
	CORRÉLATION-LAN	Utilisation de la méthode de Lan
POST-DÉTECTION	POST-SEUIL	Seuil sur le score de corrélation
	POST-AMBIGUÏTÉ	Seuil sur l'ambiguïté
	POST-SYMÉTRIE	Contrainte de symétrie
	POST-UNICITÉ	Contrainte d'unicité
	POST-ORDRE	Contrainte d'ordre
	POST-MÉDIAN	Contrainte de symétrie suivie d'un filtrage médian
	POST-DILATATION	Contrainte de symétrie suivie d'une dilatation des zones des occultations
	POST-DILATATION-CONTOUR	Contrainte de symétrie suivie d'une dilatation des zones des occultations en prenant en compte les contours
	POST-ORDRE-SYMÉTRIE	Contrainte d'ordre et contrainte de symétrie
POST-AMBIGUÏTÉ-SYMÉTRIE	Seuil sur l'ambiguïté et contrainte de symétrie	
FUSION-DISP	-	Fusion des deux cartes de disparités

TAB. 5.1 – Méthodes proposées pour prendre en compte le problème des occultations – Il y a quatre types de méthodes et quinze variantes différentes.

5.8.1.2 Critères d'évaluation

Le protocole présenté dans le chapitre 2 est utilisé. Cependant, nous avons ajouté deux cartes pour analyser les résultats obtenus :

- *La carte des occultations* – Elle représente l'ensemble des pixels détectés, par la méthode évaluée, comme appartenant à la zone totale des occultations. Les pixels noirs correspondent aux pixels détectés.
- *La carte des occultations correctes* – Elle représente la qualité d'une carte des occultations. Elle met en évidence les vrais négatifs (ce sont les pixels noirs) et les pixels correctement détectés comme proche d'une occultation (ce sont les pixels gris).

5.8.1.3 Méthodes évaluées

Nous avons évalué tous les algorithmes proposés ainsi que les différentes variantes, cf. tableau 5.1. De plus, nous avons comparé les résultats obtenus avec les méthodes hybrides avec ceux obtenus avec la méthode de mise en correspondance élémentaire (algorithme 3.2) présentée au chapitre 3 avec les mesures utilisées, c'est-à-dire, SAD, NCC, ZNCC, ME_m , LTP_2 et $SMPD_2$. À présent, nous appellerons la méthode de mise en correspondance élémentaire, la méthode élémentaire.

5.8.1.4 Synthèse des critères et classement des méthodes

Dans ce chapitre, nous prenons en compte uniquement les critères utilisés au chapitre 3. Afin d'obtenir une présentation la plus claire et la plus lisible possible, dans les tableaux, nous ne présentons les résultats que pour six méthodes : les deux méthodes élémentaires utilisant respectivement ZNCC

et SMPD₂, les méthodes POST-SYMÉTRIE, POST-DILATATION, POST-DILATATION-CONTOUR et FUSION-DISP. Dans les cartes de résultats fournis dans les figures, nous ne donnons les résultats que pour les deux méthodes élémentaires et les méthodes POST-SYMÉTRIE et FUSION-DISP, ces deux dernières étant les plus performantes.

5.8.2 Présentation et analyse des résultats

5.8.2.1 Temps de calcul

Pour les temps de calcul, la méthode CONTOUR-SDEF est la plus rapide, après l'algorithme élémentaire utilisant une mesure classique. La méthode FUSION-DISP est la plus coûteuse en temps de calcul. Tous les autres algorithmes sont moins coûteux que l'algorithme élémentaire avec SMPD₂.

5.8.2.2 Résultats préliminaires

Les premiers résultats obtenus dans [Chambon 04c] nous ont permis de déterminer la meilleure combinaison de mesure. Il s'agit de ZNCC et SMPD₂. C'est la raison pour laquelle nous ne donnons des résultats que pour ces deux mesures. De plus, dans [Chambon 04b], nous avons mis en évidence les résultats suivants :

- Pour la méthode CONTOUR, quel que soit le masque utilisé, les pourcentages de pixels corrects et de pixels corrects près des zones des occultations ne sont pas toujours meilleurs que ceux obtenus par la méthode élémentaire. Cette méthode pose un problème délicat : le choix des seuils. Nous avons déterminé, de manière empirique, les seuils optimaux (pour lesquels les meilleurs résultats sont obtenus) pour chaque image et il est clairement apparu qu'il était très difficile de les déterminer automatiquement.
- Pour la méthode CORRÉLATION, les résultats ne sont pas meilleurs. Le même problème se pose pour le choix des seuils. De plus, la méthode de Lan est beaucoup plus coûteuse en temps de calcul que toutes les autres méthodes.
- Les deux dernières catégories d'algorithmes s'avèrent les plus efficaces. Les résultats sont globalement améliorés par rapport à ceux obtenus avec l'algorithme élémentaire, en particulier en utilisant les techniques fondées sur la contrainte de symétrie : méthodes POST-SYMÉTRIE, POST-DILATATION et POST-DILATATION-CONTOUR. La méthode FUSION-DISP donne aussi de bons pourcentages de pixels corrects et améliore les résultats obtenus avec la méthode élémentaire utilisant une mesure classique, pour tous les critères, excepté le pourcentage de pixels erronés.

Par la suite, nous approfondissons ces premiers résultats, en analysant précisément les résultats obtenus avec les meilleures méthodes : POST-SYMÉTRIE, POST-DILATATION, POST-DILATATION-CONTOUR et FUSION-DISP.

5.8.2.3 Analyse des performances

Les meilleurs résultats sont toujours obtenus avec la méthode FUSION-DISP. La plupart du temps, la méthode élémentaire utilisant SMPD₂ est classée en deuxième position. Cette méthode obtient toujours les meilleurs résultats dans la zone des occultations alors que la méthode FUSION-DISP obtient toujours les meilleurs résultats dans la zone d'influence des occultations et dans la zone des discontinuités. La méthode FUSION-DISP est aussi classée première pour le pourcentage d'appariements corrects. Cette méthode allie les avantages d'une mesure de la famille CROISÉE, c'est-à-dire l'obtention du meilleur pourcentage d'appariements corrects, aux avantages d'une mesure de la famille ROBUSTE, c'est-à-dire l'obtention des meilleurs pourcentages d'appariements corrects dans la zone d'influence des occultations et dans la zone des discontinuités. Seule la méthode POST-DILATATION est moins bien

classée que les méthodes élémentaires. Cette méthode détecte trop de pixels qui ne sont pas dans la zone totale des occultations (car on dilate des zones où il y a des faux négatifs qui ne sont pas proches d'une occultation) et ainsi, au lieu d'améliorer les résultats dans les zones proches des occultations, on dégrade les résultats dans les zones de l'image qui ne présentent pas de difficulté.

Afin d'affiner notre analyse des résultats, nous donnons le détail des résultats obtenus sur quatre couples d'images les plus représentatifs, c'est-à-dire qui possèdent des zones des occultations importantes : *tsukuba*, *map*, *cones* et « livres ».

MÉTHODE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
élémentaire avec ZNCC	4	6	6	3	2	5	3	2	6	5	5
élémentaire avec SMPD ₂	5	4	3	1	1	6	1	1	3	6	2
POST-SYMÉTRIE	2	2	1	6	6	2	4	6	2	2	3
POST-DILATATION	6	5	4	4	5	4	6	5	5	3	6
POST-DILATATION-CONTOUR	3	3	5	2	4	3	5	4	4	4	4
FUSION-DISP	1	1	1	5	3	1	2	3	1	1	1

TAB. 5.2 – Classement des méthodes hybrides de mise en correspondance d'images couleur – Nous avons synthétisé les résultats obtenus sur les quatorze images de tests avec le protocole que nous avons décrit dans le paragraphe 2.9. Nous pouvons remarquer que seule la méthode POST-DILATATION est moins bien classée que la méthode élémentaire utilisant ZNCC. La seule méthode qui donne de meilleurs résultats que la méthode élémentaire utilisant SMPD₂ est la méthode FUSION-DISP. Celle-ci est d'ailleurs classée première pour plus de la moitié des critères, notamment le pourcentage d'appariements corrects dans la zone d'influence des occultations et dans la zone des discontinuités.

Tsukuba – La carte de disparités obtenue avec la méthode FUSION-DISP sur le couple *tsukuba* est beaucoup plus nette et dense (cf. figure 5.2) même si les résultats exposés dans le tableau 5.3 montrent que cette méthode n'obtient pas un meilleur pourcentage d'appariements corrects que la méthode élémentaire utilisant ZNCC. Les résultats sont améliorés de manière considérable dans la zone d'influence des occultations, 6.64%, et dans la zone des discontinuités, 9.85%. Contrairement au classement général, la méthode POST-SYMÉTRIE n'est classée que cinquième. Le pourcentage de faux négatifs est moins important avec la méthode POST-SYM qu'avec les autres algorithmes. Cela vient du fait que seuls les pixels occultés sont réexaminés avec la mesure SMPD₂, donc le pourcentage de faux négatifs ne peut être que diminué. Pour la même raison, la méthode POST-SYM obtient le pourcentage de pixels corrects dans la zone totale des occultations le plus faible et le pourcentage de faux positifs le plus élevé. En effet, seuls les pixels détectés comme occultés, en utilisant la mesure ZNCC, sont examinés et parmi ces pixels, il y a toujours certains pixels qui seront appariés, en utilisant la mesure SMPD₂. La mesure ZNCC obtenant déjà de mauvais résultats pour ces deux critères, ils ne peuvent être que dégradés avec cet algorithme.

Map – Pour ces images, l'objectif recherché est atteint par la méthode FUSION-DISP. En effet, le pourcentage d'appariements corrects dans toute l'image, dans la zone totale des occultations, la zone d'influence des occultations et la zone des discontinuités sont les meilleurs, cf. tableau 5.4 et figure 5.3. L'amélioration du pourcentage d'appariements corrects est cependant faible comparé aux améliorations pour les autres critères. Pour la méthode POST-SYMÉTRIE, nous pouvons faire la même remarque que celle faite pour les images *tsukuba*.

MÉTHODE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
élémentaire avec ZNCC	78.93 (1)	2.72 (6)	1.25 (2)	3.34 (1)	1.69 (4)	12.06 (4)	60.16 (5)	34.28 (4)	68.66 (5)	75.3 (4)	4
élémentaire avec SMPD ₂	72.28 (5)	5.9 (5)	1.46 (1)	3.4 (2)	0.9 (1)	16.07 (6)	75.82 (1)	65.23 (1)	79.3 (2)	74.89 (6)	2
POST-SYMÉTRIE	73.64 (3)	11.35 (2)	0.95 (4)	4.53 (5)	1.77 (6)	7.76 (2)	66.32 (4)	31.36 (6)	77.82 (3)	84.66 (2)	5
POST-DILATATION	71.4 (6)	10.04 (4)	0.81 (5)	3.99 (4)	1.69 (4)	12.06 (4)	60.16 (5)	34.28 (4)	68.66 (5)	75.3 (4)	6
POST-DILATATION- CONTOUR	72.99 (4)	11.07 (3)	0.75 (6)	3.75 (3)	1.31 (2)	10.13 (3)	67.63 (3)	48.98 (2)	73.76 (4)	76.88 (3)	3
FUSION-DISP	74.24 (2)	12.3 (1)	0.96 (3)	4.62 (6)	1.54 (3)	6.33 (1)	74.62 (2)	40.17 (3)	85.94 (1)	85.15 (1)	1

TAB. 5.3 – Résultats obtenus avec le couple d’images tsukuba – La méthode FUSION-DISP est première. Celle-ci n’arrive cependant pas à obtenir un pourcentage d’appariements corrects aussi bon qu’avec la méthode élémentaire utilisant ZNCC. Nous pouvons toutefois remarquer l’amélioration du pourcentage d’appariements acceptés et des pourcentages d’appariements corrects dans la zone d’influence des occultations et dans la zone des discontinuités.

MÉTHODE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
élémentaire avec ZNCC	89.1 (3)	0.05 (6)	0 (1)	0.52 (5)	0.84 (5)	9.5 (3)	70.4 (6)	88.33 (5)	51.21 (6)	70.07 (6)	6
élémentaire avec SMPD ₂	90.13 (2)	0.06 (5)	0 (1)	0.11 (1)	0.42 (1)	9.27 (2)	79.33 (2)	94.15 (1)	63.46 (2)	75.65 (3)	2
POST-SYMÉTRIE	88.72 (4)	0.3 (3)	0 (1)	0.53 (6)	0.88 (6)	9.58 (4)	75.69 (5)	87.78 (6)	62.73 (3)	78.62 (2)	5
POST-DILATATION	88.54 (5)	0.3 (3)	0 (1)	0.36 (2)	0.64 (4)	10.16 (5)	76.17 (4)	91.04 (4)	60.24 (5)	75.4 (4)	4
POST-DILATATION- CONTOUR	88.48 (6)	0.33 (2)	0 (1)	0.36 (2)	0.62 (3)	10.21 (6)	76.45 (3)	91.4 (3)	60.44 (4)	74.91 (5)	3
FUSION-DISP	90.46 (1)	0.36 (1)	0 (1)	0.51 (4)	0.44 (2)	8.24 (1)	80.4 (1)	93.9 (2)	65.94 (1)	81.78 (1)	1

TAB. 5.4 – Résultats obtenus avec le couple d’images map – Pour ces images, la méthode élémentaire utilisant ZNCC est la moins bien classée. La méthode FUSION-DISP est la mieux classée pour la plupart des critères.

Cones – Ces images possèdent la plus grande zone totale des occultations. Bien que la méthode élémentaire utilisant SMPD₂ soit meilleure que celle utilisant ZNCC, la méthode FUSION-DISP est toujours la mieux classée, cf. tableau 5.5. Les améliorations sont nettement visibles si on observe les cartes de disparités, cf. figure 5.4. Notons aussi pour ces images que la méthode POST-DILATATION est classée deuxième. Cela vient du fait qu’elle se comporte bien dans la zone d’influence des occultations et dans la zone des discontinuités. De même, elle fournit moins de faux négatifs qu’avec les autres images.

Livres – Pour ces images, les améliorations sont moins visibles, cf. tableau 5.6 et figure 5.5. Cela vient du fait que ces images sont plus difficiles à apparier car elles possèdent des zones peu texturées. La méthode FUSION-DISP reste cependant la mieux classée et obtient le meilleur pourcentage d'appariements corrects.

MÉTHODE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
élémentaire avec ZNCC	81.22 (5)	1.32 (5)	0.62 (4)	2.7 (3)	4.06 (2)	10.08 (6)	70.09 (5)	70.97 (2)	69.47 (6)	78.11 (6)	5
élémentaire avec SMAD	85.86 (1)	0.46 (6)	0.2 (6)	1.22 (1)	2.91 (1)	9.07 (4)	77.4 (1)	79.2 (1)	76.14 (3)	78.87 (5)	2
POST-SYMMÉTRIE	82.58 (4)	2.01 (1)	0.67 (1)	3.32 (6)	5.51 (6)	5.92 (2)	71.29 (3)	60.68 (6)	78.71 (2)	89.12 (2)	3
POST-DILATATION	80.02 (6)	1.97 (2)	0.66 (2)	3.09 (4)	5.18 (5)	9.08 (5)	67.35 (6)	63 (5)	70.4 (5)	81.49 (4)	5
POST-DILATATION- CONTOUR	83.36 (3)	1.64 (4)	0.49 (5)	2.37 (2)	4.65 (3)	7.49 (3)	70.82 (4)	66.77 (3)	73.65 (4)	82.68 (3)	4
FUSION-DISP	85.23 (2)	1.96 (3)	0.65 (3)	3.2 (5)	4.95 (4)	4.02 (1)	75.79 (2)	64.66 (4)	83.57 (1)	90.48 (1)	1

TAB. 5.5 – Résultats obtenus avec le couple d'images cones – Pour ces images, la méthode FUSION-DISP est la première mais elle n'obtient pas un pourcentage d'appariements corrects aussi élevé que celui obtenu par la méthode élémentaire utilisant SMPD₂. Nous pouvons toutefois noter que contrairement à la majorité des résultats, pour ces images, la méthode élémentaire utilisant SMPD₂ obtient un pourcentage d'appariements corrects nettement meilleur que celui de la méthode élémentaire utilisant ZNCC.

MÉTHODE	COR	ACC	MAU	ERR	FPO	FNE	ZT	ZO	ZI	ZD	RT
élémentaire avec ZNCC	65.1 (3)	3.83 (5)	1.28 (5)	2.51 (3)	0.39 (2)	26.89 (5)	69.99 (2)	93.2 (2)	31.73 (6)	66.91 (6)	5
élémentaire avec SMAD	63.02 (6)	3.54 (6)	1.28 (5)	2.23 (1)	0.17 (1)	29.76 (6)	72.82 (1)	97 (1)	32.98 (4)	72.53 (4)	4
POST-SYMMÉTRIE	66.69 (2)	6.69 (1)	1.83 (1)	3.51 (6)	1.57 (6)	19.7 (1)	60.55 (4)	72.54 (6)	40.81 (1)	81.47 (1)	2
POST-DILATATION	63.1 (5)	5.74 (2)	1.44 (3)	2.75 (5)	1.51 (5)	25.46 (4)	58.27 (6)	73.67 (5)	32.89 (5)	67.14 (5)	6
POST-DILATATION- CONTOUR	64.34 (4)	5.42 (4)	1.48 (2)	2.34 (2)	1.39 (4)	25.04 (3)	60.14 (5)	75.83 (4)	34.28 (2)	72.68 (3)	3
FUSION-DISP	67.46 (1)	5.74 (2)	1.44 (3)	2.63 (4)	1.26 (3)	21.45 (2)	61.41 (3)	77.96 (3)	34.12 (3)	81.17 (2)	1

TAB. 5.6 – Résultats obtenus avec le couple d'images « livres » – Contrairement à tous les résultats obtenus avec les autres images, c'est la méthode POST-SYMMÉTRIE qui obtient les meilleurs résultats dans la zone d'influence des occultations. Celle-ci est ainsi mieux classée que la méthode élémentaire utilisant SMPD₂.

5.9 Conclusion

Ce chapitre nous a permis de présenter un bref état de l'art sur les techniques employées dans la littérature pour prendre en compte les occultations. Après les résultats obtenus aux chapitre 3 et 4, nous avons choisi comme stratégie, pour prendre en compte les occultations, l'utilisation de deux mesures de corrélation : une mesure performante dans les zones sans occultation et une mesure performante dans les zones des occultations. Nous avons alors proposé quatre catégories de méthodes. Tous les tests que nous avons effectués nous ont montré que la méthode la plus efficace, mais la plus coûteuse, est la méthode FUSION-DISP. Elle consiste à fusionner les deux cartes de disparités obtenues avec les deux méthodes élémentaires utilisant respectivement ZNCC et SMPD₂.

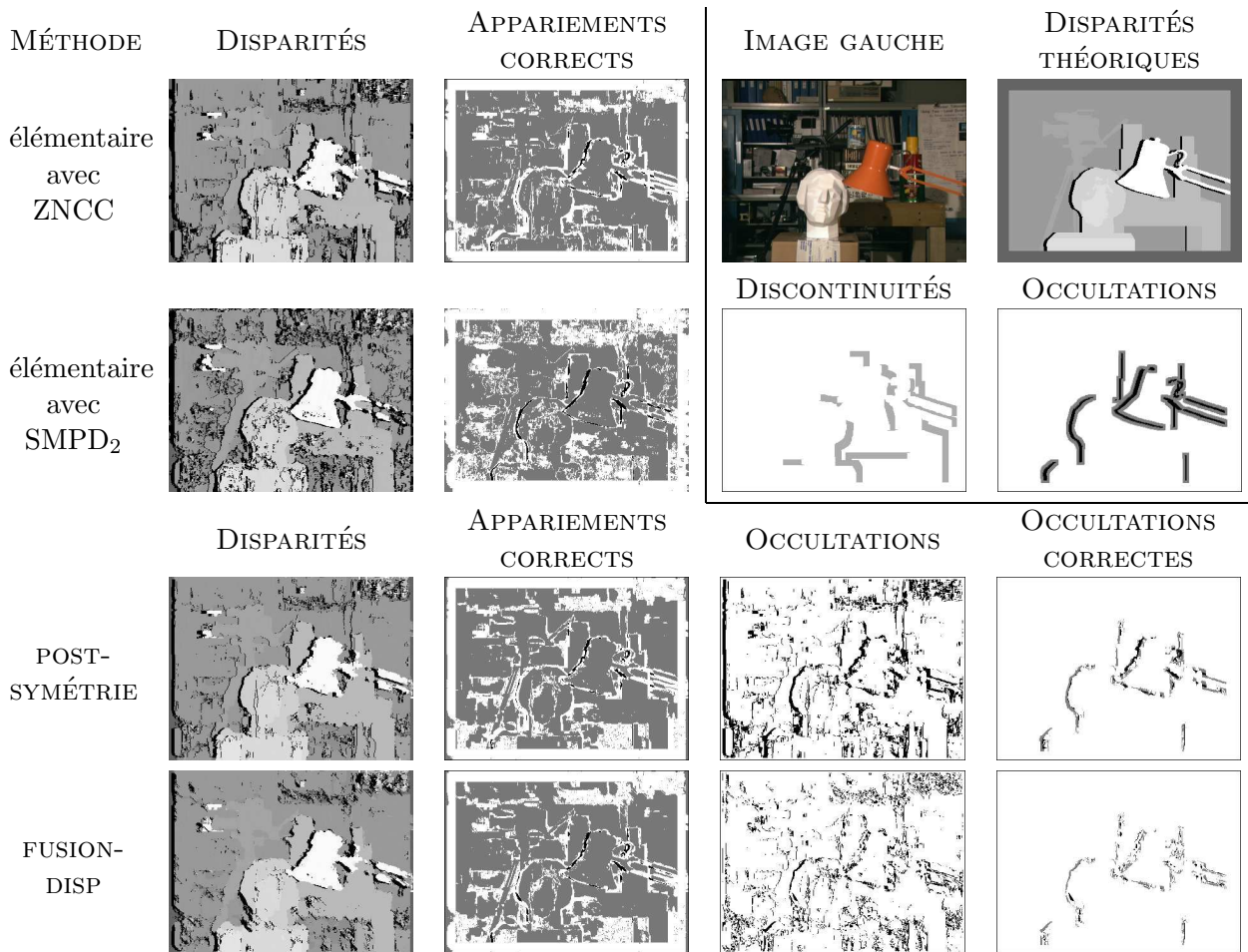


FIG. 5.2 – Cartes de résultats obtenus avec le couple d'images tsukuba – La carte de disparités obtenue avec la méthode FUSION-DISP est la plus nette bien que la méthode POST-SYMÉTRIE obtienne une carte des occultations correctes plus dense que la méthode FUSION-DISP (la zone totale des occultations représente 9.81% de l'image).

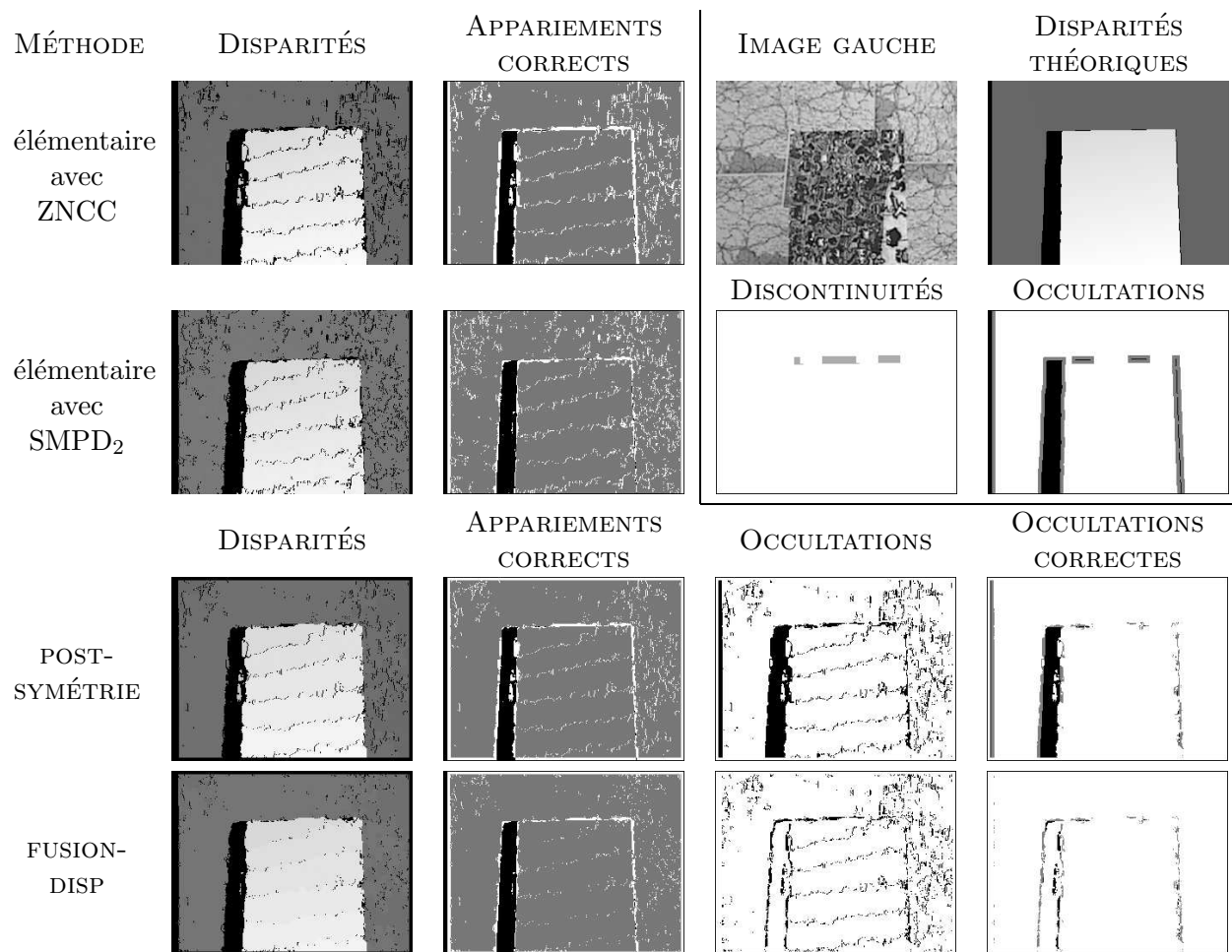


FIG. 5.3 – Cartes de résultats obtenus avec le couple d'images map – La carte d'appariements corrects obtenue avec la méthode FUSION-DISP est la plus dense. Pour ces images, la zone totale des occultations occupe 12.13% de l'image.

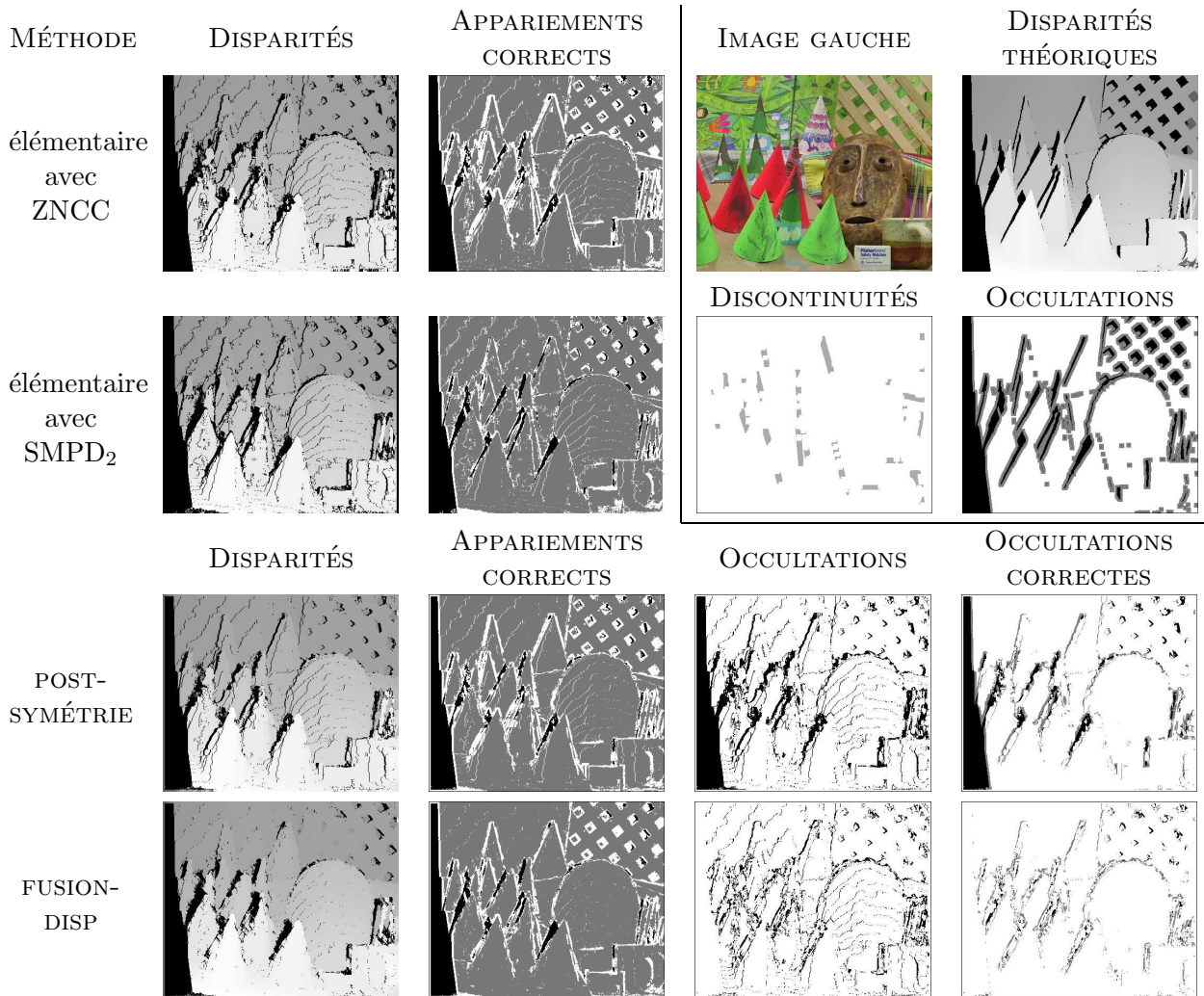


FIG. 5.4 – Cartes de résultats obtenus avec le couple d'images cones – Ces images sont celles qui contiennent la plus grande zone totale des occultations (33% de l'image). La méthode FUSION-DISP obtient la carte de disparités la plus nette et la carte d'appariements corrects la plus dense.

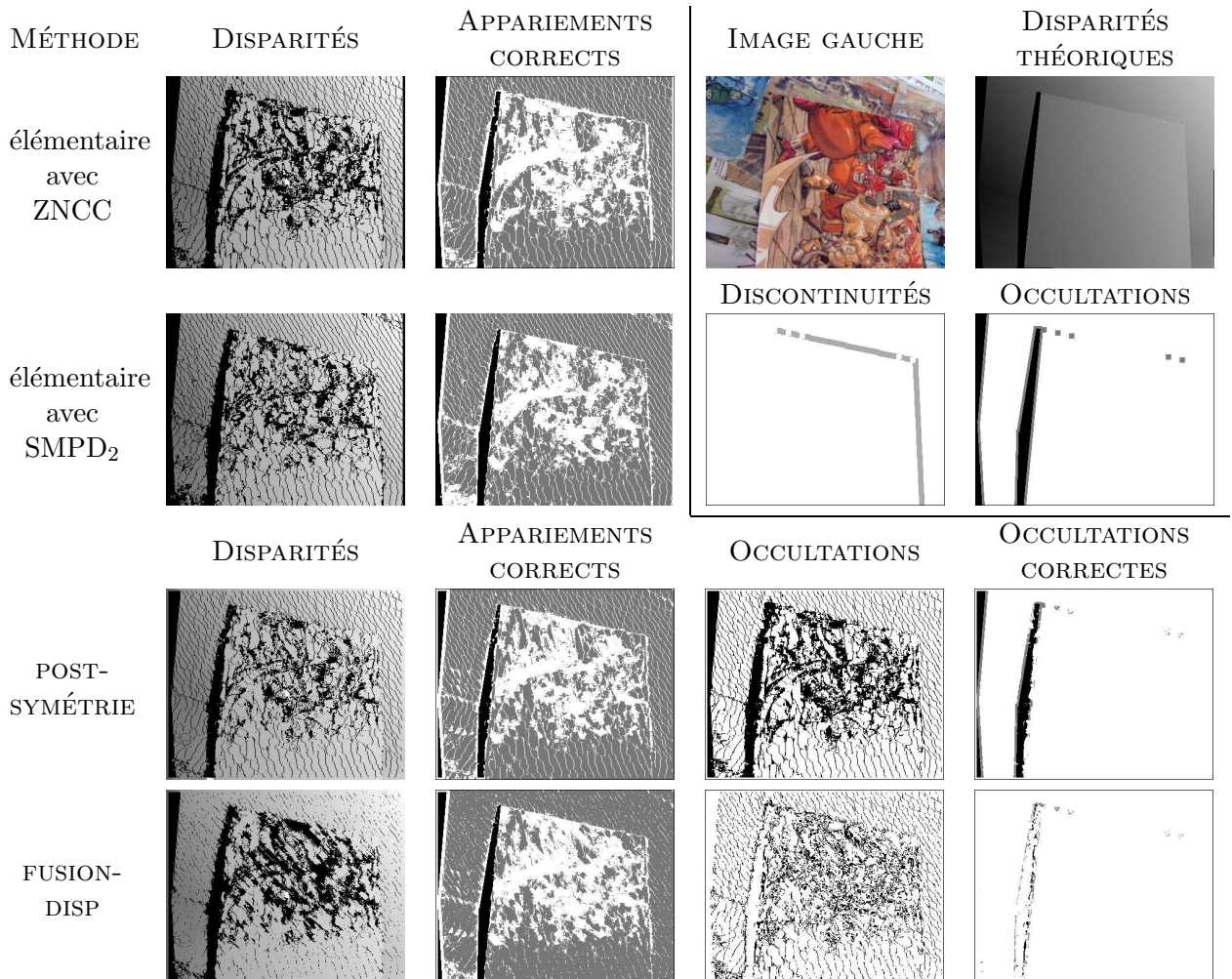


FIG. 5.5 – Cartes de résultats obtenus avec le couple d'images « livres » – Pour ces images, les améliorations sont moins visibles. Il y a cependant une zone totale des occultations qui représente seulement 8.36% et le premier plan manque « localement » de texture, c'est-à-dire que si on considère une zone de la taille de la fenêtre de corrélation, elle est peu texturée.

Conclusion

Dans ce mémoire, nous avons abordé la mise en correspondance stéréoscopique de pixels en utilisant des images couleur et en prenant en compte le problème des occultations. Nous avons tenté d'apporter des contributions à différents niveaux, c'est-à-dire en présentant une bibliographie la plus complète possible, en l'analysant et en la synthétisant, en présentant un nouveau protocole d'évaluation et de comparaison et en proposant de nouvelles techniques à base de corrélation qui prennent en compte le problème des occultations en utilisant des images en couleur.

Bibliographie

Les publications décrites dans ce mémoire peuvent être divisées en plusieurs catégories :

- des publications relatives à la mise en correspondance stéréoscopique des pixels sans restriction quant à la méthode utilisée ;
- des publications abordant le problème de l'évaluation des performances en vision par ordinateur et, essentiellement, de l'évaluation des méthodes de mise en correspondance stéréoscopique de pixels ;
- des publications sur la mise en correspondance à base de corrélation ;
- des publications où des images couleur sont prises en compte en vision par ordinateur ;
- des publications traitant le problème des occultations dans le cas de la mise en correspondance stéréoscopique.

Certaines publications appartiennent à plusieurs catégories. Ce classement des publications est accessible sur la page web suivante : <http://www.irit.fr/~Sylvie.Chambon/these.html>.

Analyse et synthèse bibliographique

Plusieurs analyses de cette bibliographie sont présentées dans chaque chapitre :

- *Chapitre 1* – Il s'agit d'un état de l'art que nous avons souhaité le plus exhaustif et le plus précis possible. C'est pourquoi, nous avons proposé une décomposition en éléments constituants. Cet état de l'art nous a permis de distinguer quatre catégories : les méthodes locales, les méthodes globales, les méthodes mixtes et les méthodes à multiples passages. Toutes les références significatives sont regroupées dans le tableau 1.3. Notre analyse nous permet de dégager les méthodes les plus populaires, c'est-à-dire les méthodes locales.
- *Chapitre 2* – Nous avons répertorié toutes les publications les plus significatives dans le domaine de l'évaluation et de la comparaison des performances des méthodes de mise en correspondance stéréoscopique de pixels. Nous avons remarqué qu'il y a peu de publications dans ce domaine. Selon nous, une seule de ces publications propose un protocole qui a permis une évaluation de grande ampleur. Nous avons décrit ce protocole et nous avons analysé les résultats que les auteurs de cette publication ont présentés sur leur site web.

- *Chapitre 3* – Nous avons classé la plupart des mesures de corrélation existantes dans la littérature en cinq familles. Ces mesures sont récapitulées dans les tableaux 3.3 à 3.7.
- *Chapitre 4* – La recherche bibliographique que nous avons effectuée sur la prise en compte d’images couleur pour la mise en correspondance stéréoscopique de pixels montre qu’il y a très peu de publications significatives dans ce domaine. Dans ce chapitre, les publications citées concernent surtout les travaux effectués sur les systèmes de représentation de la couleur.
- *Chapitre 5* – Les références de ce chapitre concernent toutes les publications qui prennent en compte les occultations dans le cadre de la mise en correspondance de pixels. L’analyse de ces publications nous a permis de distinguer cinq catégories : les méthodes avec prétraitement, les méthodes locales, les méthodes globales, les méthodes locales à multiples passages et les méthodes avec affinement.

Protocole d’évaluation et de comparaison

Nous avons proposé un protocole d’évaluation et de comparaison différent de ceux proposés dans la littérature car il est adapté aux méthodes à base de corrélation et aux aspects que nous abordons : la couleur et les occultations. C’est pourquoi, nos contributions concernent :

- *Les images testées* – Nous avons proposé de nouvelles images de tests : deux couples d’images de synthèse et un couple d’images réelles. De plus, nous fournissons un outil qui permet d’obtenir des données de référence pour des couples d’images réelles de scènes polyédriques.
- *Les critères utilisés* – Tout d’abord, nous avons distingué différentes zones des occultations et une zone des discontinuités, pour permettre une évaluation fine quant au comportement des méthodes face aux occultations. Par la suite, nous avons différencié les zones chromatiques des zones achromatiques pour évaluer le comportement des méthodes lorsque l’on prend en compte des images couleur. Puis, nous avons évalué des critères spécifiques aux méthodes de mise en correspondance à base de corrélation : l’ambiguïté et l’imprécision.

Mesures de corrélation

Pour prendre en compte le problème des occultations, nous avons proposé dix-sept mesures de corrélation s’appuyant sur des outils de statistiques robustes. Notre protocole d’évaluation nous a permis de démontrer l’efficacité de ces mesures de corrélation dans les zones des occultations que nous considérons. Cependant, les résultats obtenus ont mis en évidence que ces nouvelles mesures n’obtiennent pas les meilleurs résultats dans les autres zones de l’image.

Méthodes utilisant des images couleur

Pour prendre en compte les images couleur, nous avons étudié les différents systèmes de représentation de la couleur, nous avons généralisé la plupart des mesures de corrélation et nous avons proposé trois types de stratégies de généralisation à la couleur de la mise en correspondance à base de corrélation. D’une part, notre protocole d’évaluation confirme l’hypothèse que l’utilisation de la couleur améliore toujours les résultats par rapport à une méthode qui utilise des images en niveaux de gris. D’autre part, il s’avère que la méthode la plus efficace consiste à effectuer le calcul du score de corrélation sur chaque composante du système XYZ puis à fusionner ces résultats en utilisant l’opérateur minimum. Cette méthode améliore les résultats par rapport à une méthode classique utilisant les niveaux de gris dans la zone chromatique.

Méthodes hybrides

Pour prendre en compte les limites des mesures de corrélation robustes que nous avons proposées, nous avons présenté quatre nouveaux types de méthodes hybrides à base de détection des contours, de corrélation pondérée, de post-détection des occultations et de fusion de cartes de disparités. La méthode qui est validée par notre protocole, c'est-à-dire pour laquelle on obtient toujours les meilleurs résultats, est la méthode qui consiste à fusionner deux cartes de disparités obtenues à partir d'une mesure classique et d'une mesure robuste.

Perspectives

Nous pouvons distinguer deux types de perspectives par rapport au travail qui a été réalisé : celles qui concernent l'amélioration des méthodes proposées et celles qui portent sur l'intégration des méthodes proposées dans un cadre plus complet.

Amélioration des méthodes proposées

Face aux résultats obtenus, nous avons distingué plusieurs limitations aux méthodes proposées : les temps de calculs sont importants, les résultats des méthodes dans les zones non texturées ne sont pas satisfaisants et le problème des occultations est partiellement résolu. Nos projets concernant l'amélioration de nos méthodes sont donc les suivants :

- réduire les temps de calculs en utilisant des techniques d'accélération comme celles que nous avons brièvement évoquées au chapitre 3 ;
- améliorer les résultats dans les zones non texturées en détectant ces zones et en appliquant une autre méthode dans ces zones (par exemple, une méthode de croissance de germes) ;
- améliorer la prise en compte de la couleur en utilisant une méthode à base de couleur uniquement dans les zones chromatiques de l'image (en utilisant la méthode de détection des zones achromatiques que nous avons employée pour notre évaluation) ;
- permettre une meilleure détection des occultations dans les méthodes hybrides avec post-détection (en prenant en compte, par exemple, les résultats obtenus par une segmentation) ;
- améliorer nos méthodes hybrides en utilisant plus de deux mesures (en prenant en compte la mesure de corrélation des champs de gradients, par exemple).

Intégration des méthodes proposées

Nous souhaitons intégrer nos méthodes dans des méthodes plus complètes :

- *Une méthode de mise en correspondance multirésolution* – Ceci permettrait, de réduire les temps de calcul. La difficulté des méthodes multirésolution est la propagation des résultats au niveau des occultations. En utilisant nos méthodes qui sont plus résistantes aux occultations, nous pourrions proposer une méthode multirésolution qui résisterait bien aux occultations.
- *Une méthode de mise en correspondance par croissance de germes* – Cette méthode permet à partir de correspondances fiables de propager les résultats dans leur voisinage jusqu'à obtenir une mise en correspondance dense. Nous pourrions utiliser les mesures de corrélation que nous avons proposées pour la détection de ces germes et la propagation des résultats.
- *Une méthode globale de mise en correspondance* – L'utilisation d'une méthode globale permet de prendre en compte simultanément les différentes structures des images. Nos méthodes pourraient être utilisées pour fournir une initialisation à ces méthodes globales et les mesures de corrélation que nous avons proposées pourraient être intégrées au calcul du coût global.

Annexe

Ce paragraphe présente tous les couples d'images que nous avons testés pour valider les approches. Dans les tableaux suivants, nous donnons les couples d'images ainsi que :

- *La carte de disparités* – Plus le pixel est clair, plus la disparité est élevée et les pixels noirs sont les pixels de l'image gauche occultés dans l'image droite.
- *La carte des occultations* – Les pixels noirs sont les pixels de l'image gauche occultés dans l'image droite. Les pixels gris correspondent aux pixels proches des pixels occultés (dilatation morphologique de la zone des occultations en utilisant comme élément structurant la zone d'agrégation).
- *La carte des discontinuités* – Les pixels en gris correspondent aux pixels proches d'une discontinuité de profondeur (les discontinuités sont détectées en utilisant la carte de disparités).
- *La zone chromatique* – Les pixels blancs sont les pixels qui se trouvent dans une zone chromatique. Le stéréogramme aléatoire et le couple *map* sont en niveaux de gris.

NOM	IMAGE GAUCHE	IMAGE DROITE	CARTE DE DISPARITÉS	CARTES DES OCCULTATIONS	CARTES DES DISCONTINUITÉS	ZONE CHROMATIQUE
Stéréogramme aléatoire						
Murs						
Journaux						

Couples d'images de synthèse proposés et utilisés dans le protocole d'évaluation.

NOM	IMAGE GAUCHE	IMAGE DROITE	CARTE DE DISPARITÉS	CARTES DES OCCULTATIONS	CARTES DES DISCONTINUITÉS	ZONE CHROMATIQUE
Livres						

Couple d'images réelles proposés et utilisés dans le protocole d'évaluation.

NOM	IMAGE GAUCHE	IMAGE DROITE	CARTE DE DISPARITÉS	CARTES DES OCCULTATIONS	CARTES DES DISCONTINUITÉS	ZONE CHROMATIQUE
<i>Tsukuba</i>						
<i>Map</i>						
<i>Sawtooth</i>						
<i>Venus</i>						
<i>Bull</i>						
<i>Poster</i>						
<i>Barn 1</i>						
<i>Barn 2</i>						
<i>Cones</i>						
<i>Teddy</i>						

Couple d'images réelles présentés par Scharstein et Szeliski et utilisés dans le protocole d'évaluation.

Bibliographie

- [Adelson 84] E. H. ADELSON, C. H. ANDERSON, J. R. BERGEN, P. J. BURT et J. M. OGDEN. *Pyramid methods in image processing*. *Radio Corporation of America, RCA Engineer*, 29(6):33–41, novembre 1984.
- [Aggarwal 88] J. K. AGGARWAL et N. NANDHAKUMAR. *On the Computation of Motion from sequences of Images – A Review*. *Proceedings of the IEEE*, 76(8):917–935, août 1988.
- [Agrawal 04] M. AGRAWAL et L. DAVIS. *Window-based, discontinuity preserving stereo*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 66–73, Washington, États-Unis, juin–juillet 2004.
- [Alvarez 00] L. ALVAREZ, R. DERICHE, J. SÁNCHEZ et J. WEICKERT. Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach. Rapport de recherche RR-3874, Institut National de Recherche en Informatique et en Automatique, INRIA, janvier 2000.
- [Alvarez 02] L. ALVAREZ, R. DERICHE, T. PAPADOPOULOS et J. SÁNCHEZ. *Symmetrical Dense Optical Flow Estimation with Occlusions Detection*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 1, pages 721–735, Copenhagen, Danemark, mai 2002.
- [Anandan 89] P. ANANDAN. *A Computational Framework and an Algorithm for the Measurement of Visual motion*. *International Journal of Computer Vision, IJCV*, 2(3):283–310, janvier 1989.
- [Arcara 00] P. ARCARA, L. DI STEFANO, S. MATTOCIA, C. MELCHIORRI et G. VASSURA. *Perception of Depth Information by Means of a Wire-Actuated Haptic Interface*. Dans proceedings of International Conference on Robotic and Automation, ICRA, San Francisco, États-Unis, avril 2000.
- [Armangué 03] X. ARMANGUÉ et J. SALVI. *Overall view regarding fundamental matrix estimation*. *International Journal of Image and Vision Computing, IVC*, 21(2):205–220, février 2003.
- [Aschwanden 92] P. ASCHWANDEN et W. GUGGENBÜL. *Experimental results from a comparative study on correlation type registration algorithms*. Dans W. FÖRSTNER et S. RUWIEDEL, éditeurs, *Robust computer vision: Quality of Vision Algorithms*, pages 268–282. Wichmann, Karlsruhe, Allemagne, mars 1992.
- [Ayache 89] N. AYACHE. *Vision stéréoscopique et perception multisensorielle – Applications à la robotique mobile*. Science informatique. InterEditions, Paris, France, mai 1989.
- [Baillard 00] C. BAILLARD et O. DISSARD. *A stereo matching algorithm for urban digital elevation models*. *Photogrammetric Engineering and Remote Sensing*, 66(9):1119–1128, 2000.
- [Banks 01] J. BANKS et M. BENNAMOUN. *Reliability Analysis of the Rank Transform for Stereo Matching*. *IEEE Transactions on Systems, Man and Cybernetics, SMC*, 31(6):870–880, décembre 2001.
- [Barnard 89] S. T. BARNARD. *Stochastic Stereo Matching over Scale*. *International Journal of Computer Vision, IJCV*, 3(1):17–32, mai 1989.
- [Bazbaz 04] Camille BAZBAZ. *Sur le bout de la langue*. Dans *Sur le bout de la langue*, 2004.

- [Belhumeur 92] P. N. BELHUMEUR et D. MUMFORD. *A Bayesian Treatment of the Stereo Correspondence Problem Using Half-Occluded Regions*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 506–512, Urbana-Champaign, États-Unis, juin 1992.
- [Belli 00] T. BELLI, M. CORD et S. PHILIPP-FOLIGUET. *Colour contribution for stereo image matching*. Dans proceedings of International Conference on Color in Graphics and Image Processing, CGIP, pages 317–322, Saint-Étienne, France, octobre 2000.
- [Bellman 62] R. E. BELLMAN et S. E. DREYFUS. *La programmation dynamique et ses applications*. Dunod, Paris, France, 1962.
- [Benhimane 04] S. BENHIMANE et E. MALIS. *Mise en correspondance d'images à différentes résolutions à l'aide d'invariants aux paramètres intrinsèques*. Dans actes du Congrès AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA, volume 2, pages 585–594, Toulouse, France, janvier 2004.
- [Bhat 98] D. N. BHAT et S. K. NAYAR. *Ordinal Measures for Image Correspondence*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 20(4):415–423, avril 1998.
- [Birchfield 98] S. BIRCHFIELD et C. TOMASI. *Depth Discontinuities by Pixel-to-Pixel Stereo*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, pages 1073–1080, Bombai, Inde, janvier 1998.
- [Birchfield 99] S. BIRCHFIELD et C. TOMASI. *Multiway Cut for Stereo and Motion with Slanted Surfaces*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 1, pages 489–495, Kerkyra, Grèce, septembre 1999.
- [Bischof 94] H. BISCHOF, E. BERTIN et P. BERTOLLINE. *Voronoi Pyramids and Hopfield Networks*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 330–333, Jérusalem, Israël, octobre 1994.
- [Black 93] M. J. BLACK et P. ANANDAN. *A Framework for the Robust Estimation of Optical Flow*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, pages 231–236, Berlin, Allemagne, mai 1993.
- [Black 94] M. J. BLACK et A. JEPSON. *Estimating Multiple Independent Motions in Segmented Images using Parametric Models with Local Deformations*. Dans IEEE Workshop on Motion of Non-Rigid and Articulated Objects, Austin, États-Unis, novembre 1994.
- [Black 96] M. J. BLACK et A. RANGARAJAN. *On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision*. *International Journal of Computer Vision, IJCV*, 19(1):57–91, juillet 1996.
- [Blake 98] A. BLAKE et M. ISARD. *Active contours: The application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion*. Springer-Verlag, New-York, États-Unis, 1998.
- [Bleyer 04] M. BLEYER et M. GELAUTZ. *A layered stereo algorithm using image segmentation and global visibility constraints*. Dans IEEE Conference Proceedings of International Conference on Image Processing, ICIP, volume 5, pages 2997–3000, Singapour, octobre 2004.
- [Bleyer 05] M. BLEYER et M. GELAUTZ. *Graph-based surface reconstruction from stereo pairs using image segmentation*. Dans Videometrics VIII, volume SPIE-5665, pages 288–299, San José, États-Unis, janvier 2005.
- [Bobick 99] A. F. BOBICK et S. S. INTILLE. *Large Occlusion Stereo*. *International Journal of Computer Vision, IJCV*, 33(3):181–200, septembre 1999.
- [Bocquillon 04] B. BOCQUILLON. *Obtention de la vérité terrain pour la mise en correspondance stéréoscopique*. Rapport de DEA, Université Paul Sabatier, UPS, Toulouse, France, juin 2004.

- [Bocquillon 05a] B. BOCQUILLON, S. CHAMBON et A. CROUZIL. *Segmentation semi-automatique en plans pour la génération de cartes denses de disparités*. Dans actes du congrès francophone de Vision par Ordinateur, ORASIS, Fournol, France, mai 2005. CD-ROM.
- [Bocquillon 05b] B. BOCQUILLON, S. CHAMBON et A. CROUZIL. *Segmentation semi-automatique en plans pour la génération de cartes denses de disparités*. Rapport de recherche 2005-23-R, Institut de Recherche en Informatique de Toulouse, IRIT, Université Paul Sabatier, UPS, France, octobre 2005.
- [Bolles 93] R. C. BOLLES, H. H. BAKER et M. J. HANNAH. *The JISCT Stereo Evaluation*. Dans ARPA Image Understanding Workshop, pages 263–274, Washington, États-Unis, avril 1993.
- [Boufama 02] B. S. BOUFAMA et D. O’CONNELL. *Region Segmentation and Matching of Planes in a Pair of Uncalibrated Images*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 631–634, Québec, Canada, août 2002.
- [Bovyryn 03] A. BOVYRIN, V. ERUHIMOV, S. MOLINOV, V. MOSYAGIN et V. PISAREVSKY. *Fast and robust dense stereo correspondence by column segmentation*. Dans IEEE Conference Proceedings of International Conference on Image Processing, ICIP, pages 1033–1036, Barcelone, Espagne, septembre 2003.
- [Boykov 01] Y. BOYKOV, O. VEKSLER et R. ZABIH. *Fast approximate energy minimization via graph cuts*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 23(11):1222–1239, novembre 2001.
- [Braquelaire 97] J. P. BRAQUELAIRE et L. BRUN. *Comparison and optimization of methods of color image quantization*. *IEEE Transactions on Image Processing, IP*, 6(7):1048–1052, juillet 1997.
- [Brigger 99] P. BRIGGER, F. MÜLLER, K. ILLGNER et M. UNSER. *Centered Pyramids*. *IEEE Transactions on Image Processing, IP*, 8(9):1254–1264, septembre 1999.
- [Brockers 04a] R. BROCKERS, M. HUND et B. MERTSCHING. *A Fast Cost Relaxation Stereo Algorithm with Occlusion Detection for Mobile Robot Applications*. Dans proceedings of Vision, Modeling and Visualization Conference, VMVC, pages 47–53, Stanford, États-Unis, novembre 2004.
- [Brockers 04b] R. BROCKERS, M. HUND et B. MERTSCHING. *Fast Stereo Vision for Mobile Robots by Global Minima of Cost Functions*. Dans proceedings of Vision, Modeling and Visualization Conference, VMVC, pages 85–89, Stanford, États-Unis, novembre 2004.
- [Brown 03] M. Z. BROWN, D. BURSCHKA et G. D. HAGER. *Advances in Computational Stereo*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 25(8):993–1008, août 2003.
- [Brun 96] L. BRUN. *Segmentation d’images à base topologique*. Thèse de doctorat, Université de Bordeaux I, France, décembre 1996.
- [Buessler 05] J.-L. BUSSLER, J.-P. URBAN, G. HERMANN et H. KIHLE. *Colour Histogram Algorithms for Visual Robot Control*. *International Journal of Robotics and Automation, IJRA*, 20(2):86–93, 2005.
- [Burt 83] P. J. BURT et E. H. ADELSON. *The Laplacian Pyramid as a Compact Image Code*. *IEEE Transactions on Communications*, 31(4):532–540, avril 1983.
- [Busin 05] L. BUSIN, N. VANDENBROUCKE, L. MACAIRE et J.-G. POSTAIRE. *Color Space Selection for Unsupervised Colour Image Segmentation by Analysis of Connectedness Properties*. *International Journal of Robotics and Automation, IJRA*, 20(2):70–77, 2005.
- [Canny 86] J. F. CANNY. *A computational approach to edge detection*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 8(6):679–698, novembre 1986.
- [Carneiro 04] G. CARNEIRO et A. D. JEPSON. *Pruning Local Feature Correspondences Using Shape Context*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 16–19, Cambridge, Royaume-Uni, août 2004.

- [Carron 95] T. CARRON. *Segmentation d'images couleur dans la base Teinte Luminance Saturation : approche numérique et symbolique*. Thèse de doctorat, Université de Savoie, France, décembre 1995.
- [Chabrier 04] S. CHABRIER, B. EMILE, H. LAURENT, C. ROSENBERGER et P. MARCHÉ. *Unsupervised Evaluation of Image Segmentation – Application to Multi-Spectral Images*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 576–579, Cambridge, Royaume-Uni, août 2004.
- [Chambah 01] M. CHAMBAH. *Analyse et traitement de données chromatiques d'images numérisées à haute résolution. Application à la restauration numérique des couleurs des films cinématographiques*. Thèse de doctorat, Université de La Rochelle, France, décembre 2001.
- [Chambon 82] D. CHAMBON. *Description d'images numériques : Réalisation du module de vision bas-niveau de S.A.C.S.O. (Système d'Analyse et de Compréhension de Scène par Ordinateur)*. Thèse de doctorat, Université Paul Sabatier, UPS, Toulouse, France, octobre 1982.
- [Chambon 02] S. CHAMBON et A. CROUZIL. Évaluation et comparaison de mesures de corrélation robustes aux occultations. Rapport de recherche 2002-34-R, Institut de Recherche en Informatique de Toulouse, IRIT, Université Paul Sabatier, UPS, France, décembre 2002.
- [Chambon 03] S. CHAMBON et A. CROUZIL. *Dense matching using correlation: new measures that are robust near occlusions*. Dans proceedings of British Machine Vision Conference, BMVC, volume 1, pages 143–152, Norwich, Royaume-Uni, septembre 2003.
- [Chambon 04a] S. CHAMBON et A. CROUZIL. *Mesures de corrélation pour des images couleur. Traitement du signal, TS*, 21(6):635–659, 2004.
- [Chambon 04b] S. CHAMBON et A. CROUZIL. Mise en correspondance par corrélation avec détection des zones d'occultation. Rapport de recherche 2004-31-R, Institut de Recherche en Informatique de Toulouse, IRIT, Université Paul Sabatier, UPS, France, décembre 2004.
- [Chambon 04c] S. CHAMBON et A. CROUZIL. *Towards correlation-based matching algorithms that are robust near occlusions*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 20–23, Cambridge, Royaume-Uni, août 2004.
- [Chambon 05] S. CHAMBON et A. CROUZIL. *Colour Correlation-Based Matching. International Journal of Robotics and Automation, IJRA*, 20(2):78–87, 2005.
- [Chanussot 98] J. CHANUSSOT. *Approches vectorielles ou marginales pour le traitement d'images multi-composantes*. Thèse de doctorat, Université de Savoie, France, novembre 1998.
- [Chehata 05] N. CHEHATA. *Interprétation de scènes urbaines à partir d'images satellitaires THR : reconstruction de facettes 3D et optimisation globale 3D. Bulletin d'information scientifique et technique de l'IGN*, (75):29–40, janvier 2005.
- [Chen 99] Q. CHEN et G. MEDIONI. *A Volumetric Stereo Matching Method: Application to Image-Based Modeling*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 1029–1034, Fort Collins, États-Unis, juin 1999.
- [Chen 03] J.-H. CHEN, C.-S. CHEN et Y.-S. CHEN. *Fast Algorithm for Robust Template Matching With M-Estimator. IEEE Transactions on Signal Processing*, 51(1):230–243, janvier 2003.
- [Cheng 03] S.-C. CHENG et S.-C. HSIA. *Fast algorithms for color image processing by principal component analysis. Journal of Visual Communication and Image Representation, JVCIP*, 14(2):184–203, juin 2003.
- [Cie 15.2 86] CIE 15.2. Colorimetry, second edition. Rapport Technique, Commission Internationale de l'Éclairage, Vienne, Autriche, 1986.
- [Cochran 92] S. D. COCHRAN et G. MÉDIONI. *3-D Surface Description from Binocular Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 14(10):981–994, octobre 1992.

- [Collins 96] R. T. COLLINS. *A space-sweep approach to true multi-image matching*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 358–363, San Francisco, États-Unis, juin 1996.
- [Comaniciu 99] D. COMANICIU et P. MEER. *Distribution free decomposition of multivariate data*. *Pattern Analysis and Applications, PAA*, 2(1):22–30, avril 1999.
- [Coquin 02] D. COQUIN, P. BOLON et B. IONESCU. *Dissimilarity Measures in Color Spaces*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 612–615, Québec, Canada, août 2002.
- [Cord 98] M. CORD, N. PAPANICOLAOS et M. JORDAN. *Dense, reliable and depth discontinuity preserving DEM computation from HRV urban stereopairs*. *International Archives of Photogrammetry and Remote Sensing*, 32(2):57–63, 1998.
- [Cox 92] I. J. COX. *Stereo without Disparity Gradient Smoothing: a Bayesian Sensor Fusion Solution*. Dans proceedings of British Machine Vision Conference, BMVC, pages 337–346, Leeds, Royaume-Uni, septembre 1992.
- [Cox 95] G. S. COX. *Template matching and measures of match in image processing*. Rapport Technique, Université de Cape Town, Afrique du Sud, juillet 1995.
- [Cox 96] I. J. COX, S. L. HINGORANI, S. B. RAO et B. M. MAGGS. *A Maximum Likelihood Stereo Algorithm*. *Computer Vision and Image Understanding, CVIU*, 63(3):1271–1281, mai 1996.
- [Criminisi 03] A. CRIMINISI, J. SHOTTON, A. BLAKE, C. ROTHER et P. H. S. TORR. *Efficient dense-stereo and novel-view synthesis for gaze manipulation in one-to-one teleconferencing*. Rapport technique MSR-TR-2003-59, Recherche Microsoft, Cambridge, Royaume-Uni, septembre 2003.
- [Criminisi 05] A. CRIMINISI, J. SHOTTON et A. BLAKE, C. ROTHER et P. H. S. TORR. *Efficient dense-stereo with occlusions and new view synthesis by four state DP for gaze correction*. *International Journal of Computer Vision, IJCV*, 2005. Soumis.
- [Crouzil 96] A. CROUZIL, L. MASSIP-PAIHES et S. CASTAN. *A New Correlation Criterion Based on Gradient Fields Similarity*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 632–636, Vienne, Autriche, août 1996.
- [Crouzil 97] A. CROUZIL. *Perception du relief et du mouvement par analyse d'une séquence stéréoscopique d'images*. Thèse de doctorat, Université Paul Sabatier, UPS, Toulouse, France, septembre 1997.
- [Crouzil 03] A. CROUZIL, X. DESCOMBES et J.-D. DUROU. *A Multiresolution Approach for Shape from Shading Coupling Deterministic and Stochastic Optimization*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 25(11):1416–1421, novembre 2003.
- [Cucchiara 04] R. CUCCHIARA, C. GRANA, G. TARDINI et R. VEZZANI. *Probabilistic People Tracking for Occlusion Handling*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 132–135, Cambridge, Royaume-Uni, août 2004.
- [Cyganek 01] B. CYGANEK et J. BORGOSZ. *A Comparative Study of Performance and Implementation of Some Area-Based Stereo Algorithms*. Dans proceedings of International Conference on Computer Analysis of Images and Patterns, CAIP, pages 709–716, Varsovie, Pologne, septembre 2001.
- [Cyganek 03] B. CYGANEK et J. BORGOSZ. *An Improved Variogram Analysis of the Maximum Expected Disparity in Stereo Images*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 640–645, Göteborg, Suède, juin 2003.
- [Dante 03] A. DANTE, M. BROOKES et A. G. CONSTANTINIDES. *Robust multi-body segmentation*. Dans proceedings of British Machine Vision Conference, BMVC, volume 2, pages 669–678, Norwich, Royaume-Uni, septembre 2003.
- [Davis 81] L. S. DAVIS et A. ROSENFELD. *Cooperating processes for low-level vision: A survey*. *Artificial Intelligence*, 17(1–3):245–263, août 1981.

- [Davis 95] C. Q. DAVIS, Z. Z. KARU et D. M. FREEMAN. *Equivalence of Subpixel Motion Estimators Based on Optical Flow and Block Matching*. Dans IEEE Symposium on Computer Vision, pages 7–12, Coral Gables, États-Unis, novembre 1995.
- [De Joinville 00] O. DE JOINVILLE, H. MAÎTRE, D. PIQUET PELLORCE et M. ROUX. *How to Design DEM Assessment Maps*. Dans proceedings of International Workshop on Pattern Recognition in Remote Sensing, Andorre-La-Vieille, Andorre, septembre 2000.
- [De Joinville 01] O. DE JOINVILLE, G. MAILLET, H. MAÎTRE et M. ROUX. *Évaluation a priori de la qualité d'un MNS*. Dans actes du congrès francophone de Vision par Ordinateur, ORASIS, pages 67–76, Cahors, France, juin 2001.
- [De Saint-Exupéry 43] A. DE SAINT-EXUPÉRY. *Le petit prince*. Gallimard, 1943.
- [Delon 01] J. DELON et B. ROUGÉ. *Le phénomène d'adhérence en stéréoscopie dépend du critère de corrélation*. Dans actes du colloque GRETSI sur le traitement du signal et des images, Toulouse, France, septembre 2001.
- [Deng 05] Y. DENG, Q. YANG, X. LIN et X. TANG. *A Symmetric Patch-Based Correspondence Model for Occlusion Handling*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, Beijing, Chine, octobre 2005. À paraître.
- [Descombes 04] X. DESCOMBES. *Méthodes stochastiques en analyse d'image : des champs de Markov aux processus ponctuels marqués*. Habilitation à diriger des recherches, Université de Nice – Sophia Antipolis, France, février 2004.
- [Devernay 97] F. DEVERNAY. *Vision stéréoscopique et propriétés différentielles des surfaces*. Thèse de doctorat, Institut National Polytechnique, INP, Grenoble, France, février 1997.
- [Devernay 02] F. DEVERNAY, O. BANTICHE et E. COSTE. *Structured light on dynamic scenes using standard stereoscopy algorithms*. Rapport de recherche RR-4477, Institut National de Recherche en Informatique et en Automatique, INRIA, juin 2002.
- [Devernay 05] F. DEVERNAY et O. FAUGERAS. *Shape from Stereo Using Fine Correlation: Method and Error Analysis*. *International Journal of Image and Vision Computing, IVC*, 2005. Soumis.
- [Dhond 95] U. R. DHOND et J. K. AGGARWAL. *Stereo Matching in the Presence of Narrow Occluding Objects Using Dynamic Disparity Search*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 17(7):719–724, juillet 1995.
- [Di Stefano 01] L. DI STEFANO, S. MATTOCIA, G. NERI et D. PICCININI. *Temporal Filtering of Disparity Measurements*. Dans proceedings of International Conference on Image Analysis and Processing, ICIAP, pages 145–150, Palerme, Italie, septembre 2001.
- [Di Stefano 02] L. DI STEFANO, M. MARCHIONNI, S. MATTOCIA et G. NERI. *A Fast Area-Based Stereo Matching Algorithm*. Dans proceedings of International Conference on Vision Interface, VI, pages 146–153, Calgary, Canada, mai 2002.
- [Di Stefano 03] L. DI STEFANO et S. MATTOCIA. *Fast template matching using bounded partial correlation*. *International Journal of Machine Vision and Applications, MVA*, 13(4):213–221, février 2003.
- [Di Zenzo 86] S. DI ZENZO. *A Note on the Gradient of a Multi-Image*. *Computer Vision, Graphics, and Image Processing, CVGIP*, 33(1):116–125, janvier 1986.
- [Duffy 98] N. DUFFY et G. LACEY. *Colour Profiling Using Multiple Colour Spaces*. Dans proceedings of British Machine Vision Conference, BMVC, pages 245–256, Southampton, Royaume-Uni, septembre 1998.
- [Dufournaud 00] Y. DUFOURNAUD, C. SCHMID et R. HORAUD. *Matching Images with Different Resolutions*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 612–618, Hilton Head Island, États-Unis, juin 2000.

- [Egnal 00] G. EGNAL et R. P. WILDES. *Detecting Binocular Half-Occlusions: Empirical Comparisons of Four Approaches*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 2, pages 466–473, Hilton Head Island, États-Unis, juin 2000.
- [Egnal 02] G. EGNAL et R. P. WILDES. *Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 24(8):1127–1133, août 2002.
- [Ehlers 86] M. EHLERS et R. WELCH. *Stereo-Correlation of Landsat-TM Images*. *Photogrammetric Engineering and Remote Sensing*, 53(9):1321–1237, 1986.
- [Eklund 03] M. P. EKLUND et A. A. FARAG. *Robust correspondence methods for stereo vision*. *International Journal of Pattern Recognition and Artificial Intelligence, PRAI*, 17(7):1059–1079, novembre 2003.
- [El-Sonbaty 03] Y. EL-SONBATY et M. A. ISMAIL. *Matching Occluded Objects Invariant to Rotations, Translations, Reflections and Scale Changes*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 836–843, Göteborg, Suède, juin 2003.
- [Faghihi 02] R. FAGHIHI. *Mise en correspondance SPECT-CT par conditions de consistance*. Thèse de doctorat, Université Joseph Fourier, La Tronche, France, novembre 2002.
- [Falkenhagen 97] L. FALKENHAGEN. *Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints*. Dans International Workshop on Synthetic-Natural Hybrid Coding and 3D Imaging, Rhodes, Grèce, septembre 1997.
- [Faugeras 79] O. FAUGERAS. *Digital Color Image Processing Within the Framework of a Human Visual Model*. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(4):380–393, août 1979.
- [Faugeras 92] O. FAUGERAS, P. FUA, B. HOTZ, R. MA, L. ROBERT, M. THONNAT et Z. ZHANG. *Quantitative and Qualitative Comparison of Some Area and Feature-based Stereo Algorithms*. Dans FÖRSTNER et RUWIEDEL, éditeurs, *Robust computer vision: Quality of Vision Algorithms*, pages 1–26. Wichmann, Karlsruhe, Allemagne, mars 1992.
- [Faugeras 93] O. FAUGERAS, B. HOTZ, Z. ZHANG et P. FUA. *Real time correlation-based stereo: Algorithm, implementation and applications*. Rapport de recherche RR-2013, Institut National de Recherche en Informatique et en Automatique, INRIA, août 1993.
- [Ferrari 03] V. FERRARI, T. TUYTELAARS et L. VAN GOOL. *Wide-baseline Multiple-view Correspondences*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 2, pages 718–725, Madison, États-Unis, juin 2003.
- [Fielding 00] G. FIELDING et M. KAM. *Weighted matchings for dense stereo correspondence*. *The Journal of the Pattern Recognition Society, PR*, 33(9):1511–1524, septembre 2000.
- [Fleet 91] D. J. FLEET, A. D. JEPSON et M. R. M. JENKIN. *Phase-Based Disparity Measurement*. *Computer Vision, Graphics, and Image Processing: Image Understanding, CVGIP*, 53(2):198–210, mars 1991.
- [Forney 73] G. D. FORNEY. *The Viterbi algorithm*. *Proceedings of the IEEE*, 61(3):268–278, mars 1973.
- [Forstmann 04] S. FORSTMANN, Y. KANOU, J. OHYA, S. THUERING et A. SCHMITT. *Real-Time Stereo by using Dynamic Programming*. Dans Workshop on real-time 3D sensors and their use, IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 3, pages 29–36, Washington, États-Unis, juin–juillet 2004.
- [Frohlinghaus 96] T. FROHLINGHAUS et J. M. BUHMANN. *Regularizing Phase-based Stereo*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 451–455, Vienne, Autriche, août 1996.

- [Fua 91] P. FUA. *Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities*. Dans proceedings of International Joint Conference on Artificial Intelligence, IJAI, pages 1292–1298, Sydney, Australie, août 1991.
- [Fua 93] P. FUA. *A parallel stereo algorithm that produces dense depth maps and preserves image features*. *International Journal of Machine Vision and Applications*, MVA, 6(1):35–49, janvier 1993.
- [Fuchs 01] F. FUCHS. *Contribution à la reconstruction du bâti en milieu urbain, à l'aide d'images aériennes stéréoscopiques à grande échelle. Étude d'une approche structurelle*. Thèse de doctorat, Université René Descartes – Paris V, France, avril 2001.
- [Fusiello 97a] A. FUSIELLO, V. ROBERTO et E. TRUCCO. *Efficient Stereo with Multiple Windowing*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 858–863, San Juan, Porto Rico, juin 1997.
- [Fusiello 97b] A. FUSIELLO, V. ROBERTO et E. TRUCCO. *Experiments with a New Area-Based Stereo Algorithm*. Dans proceedings of International Conference on Image Analysis and Processing, ICIAP, volume 2, pages 669–676, Florence, Italie, septembre 1997.
- [Fusiello 99] A. FUSIELLO, E. TRUCCO, T. TOMMASINI et V. ROBERTO. *Improving Feature Tracking with Robust Statistics*. *Pattern Analysis and Applications*, PAA, 2(4):312–320, novembre 1999.
- [Fusiello 00] A. FUSIELLO, E. TRUCCO et A. VERRI. *A compact algorithm for rectification of stereo pairs*. *International Journal of Machine Vision and Applications*, MVA, 12(1):16–22, juillet 2000.
- [Garbay 79] C. GARBAY. *Modélisation de la couleur dans le cadre de l'analyse d'images et de son application à la cytologie automatique*. Thèse de doctorat, Institut National Polytechnique, INP, Grenoble, France, décembre 1979.
- [Garbay 81] C. GARBAY, F. BRUGAL et C. CHOQUET. *Application of Colored Image Analysis to Bone Marrow Cell Recognition*. *Analytical and Quantitative Cytology*, 3(4):272–280, juillet 1981.
- [Garcia 01a] D. GARCIA. *Mesures de formes et de champs de déplacements tridimensionnels par stéréo-corrélation d'images*. Thèse de doctorat, École des Mines d'Albi, France, décembre 2001.
- [Garcia 01b] D. GARCIA et J.-J. ORTEU. *3D Deformation Measurement using Stereo-correlation applied to Experimental Mechanics*. Dans International Symposium on Deformation Measurements, pages 50–60, Orange, États-Unis, mars 2001.
- [Garcia 01c] R. GARCIA, X. CUFÍ et J. BATLE. *Detection of Matching in a Sequence of Underwater Images through Texture Analysis*. Dans IEEE Conference Proceedings of International Conference on Image Processing, ICIP, volume 1, pages 361–364, Thessalonique, Grèce, octobre 2001.
- [Garcia 02] D. GARCIA, J.-J. ORTEU et L. PENAZZI. *A Combined Temporal Tracking and Stereo-correlation Technique for Accurate Measurement of 3D Displacements: Application to Sheet Metal Forming*. *Journal of Materials Processing Technology*, 2002(125–126):736–742, septembre 2002.
- [Geiger 95] D. GEIGER, B. LADENDORF et A. YUILLE. *Occlusions and Binocular Stereo*. Dans IEEE Conference Proceedings of International Conference on Image Processing, ICIP, volume 14, pages 211–226, Washington, États-Unis, octobre 1995.
- [Geman 84] S. GEMAN et D. GEMAN. *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI, 6(6):721–741, novembre 1984.
- [Gennert 88] M. A. GENNERT. *Brightness-Based Stereo Matching*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, pages 138–143, Tampa, États-Unis, décembre 1988.
- [Georgescu 04] B. GEORGESCU et P. MEER. *Point Matching under Large Image Deformations and Illumination Changes*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI, 26(6):674–688, juin 2004.

- [Gevers 98] T. GEVERS, A. W. M. SMEULDERS et H. STOKMAN. *Photometric Invariant Region Detection*. Dans proceedings of British Machine Vision Conference, BMVC, pages 578–589, Southampton, Royaume-Uni, septembre 1998.
- [Gevers 03] T. GEVERS. *Reflectance-based Classification of Color Edges*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 856–861, Nice, France, octobre 2003.
- [Ghafoor 03] A. GHAFOR, R. NAVEED IQBAL et S. SHOAD KHAN. *Image Matching Using Distance Transform*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 654–660, Göteborg, Suède, juin 2003.
- [Giachetti 00] A. GIACHETTI. *Matching techniques to compute image motion*. *International Journal of Image and Vision Computing, IVC*, 18(3):245–258, février 2000.
- [Gong 02] M. GONG et Y.-H. YANG. *Genetic-Based Stereo Algorithm and Disparity Map Evaluation*. *International Journal of Computer Vision, IJCV*, 47(1–3):63–77, avril 2002.
- [Gong 03] M. GONG et Y.-H. YANG. *Fast Stereo Matching Using Reliability-Based Dynamic Programming and Consistency Constraints*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 1, pages 610–617, Nice, France, octobre 2003.
- [Gong 04] M. GONG. *Motion Estimation using Dynamic Programming with Selective Path Search*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 4, pages 203–206, Cambridge, Royaume-Uni, août 2004.
- [Gong 05] M. GONG et Y.-H. YANG. *Near Real-time Reliable Stereo Matching Using Programmable Graphics Hardware*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 924–931, San Diego, États-Unis, juin 2005.
- [Gottesfeld Brown 92] L. GOTTESFELD BROWN. *A Survey of Image Registration Techniques*. *ACM Computing Surveys*, 24(4):325–376, décembre 1992.
- [Gouet 98] V. GOUET, P. MONTESINOS et D. PEL. *A Fast Matching Method for Color Uncalibrated Images using Differential Invariants*. Dans proceedings of British Machine Vision Conference, BMVC, pages 367–376, Southampton, Royaume-Uni, septembre 1998.
- [Gouet 00] V. GOUET, P. MONTESINOS, R. DERICHE et D. PELÉ. *Évaluation de détecteurs de points d'intérêt pour la couleur*. Dans actes du Congrès AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA, pages 257–266, Paris, France, février 2000.
- [Goulermas 03] J. Y. GOULERMAS et P. LIATSIS. *A Collective-Based Adaptive Symbiotic Model for Surface Reconstruction in Area-Based Stereo*. *IEEE Transactions on Evolutionary Computation*, 7(5):482–502, octobre 2003.
- [Gribkov 03] I. V. GRIBKOV, P. P. KOL'TSOV, A. A. KRAVCHENKO, A.S. KUTSAEV, V. K. NIKOLAEV et A. V. ZAKHAROV. *PICASSO—The System for Evaluating Edge-Detection Algorithms*. *Pattern Recognition and Image Analysis, PRIA*, 13(2):208–210, octobre 2003.
- [Gros 97] P. GROS, G. MACLEAN, R. DELON, R. MOHR, C. SCHMID et G. MISTLER. *Utilisation de la couleur pour l'appariement et l'indexation d'images*. Rapport de recherche RR-3269, Institut National de Recherche en Informatique et en Automatique, INRIA, septembre 1997.
- [Guisser 00] L. GUISSER, R. PAYRISSAT et S. CASTAN. *PGSD: an accurate 3D vision system using a projected grid for surface descriptions*. *International Journal of Image and Vision Computing, IVC*, 18(6–7):463–491, mai 2000.
- [Gutiérrez 04] S. GUTIÉRREZ et J. LUIS MARROQUÍN. *Robust approach for disparity estimation in stereo vision*. *International Journal of Image and Vision Computing, IVC*, 22(3):183–195, mars 2004.

- [Gökstorp 95] M. GÖKSTORP et C.-J. WESTELIUS. *Multiresolution Differential-based Disparity Estimation*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 67–76, Uppsala, Suède, juin 1995.
- [Han 01] K.-P. HAN, K.-W. SONG, E.-Y. CHUNG, S.-J. CHO et Y.-H. HA. *Stereo matching using genetic algorithm with adaptive chromosomes*. *The Journal of the Pattern Recognition Society, PR*, 34(9):1729–1740, septembre 2001.
- [Hanbury 03] A. HANBURY. *A 3D-Polar Coordinate Colour Representation Well Adapted to Image Analysis*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 804–811, Göteborg, Suède, juin 2003.
- [Hansen 03] M. E. HANSEN et J. M. CARSTENSEN. *Color-Based Image Retrieval from High-Similarity Image Databases*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 1098–1105, Göteborg, Suède, juin 2003.
- [Hardeberg 99] J. Y. HARDEBERG. *Acquisition and reproduction of colour images: colorimetric and multispectral approaches*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, ENST, Paris, France, janvier 1999.
- [Harris 88] C. HARRIS et M. STEPHENS. *A Combined Corner and Edge Detector*. Dans Alvey Vision Conference, pages 147–151, Manchester, Royaume-Uni, janvier 1988.
- [Hartley 99] R. I. HARTLEY. *Theory and Practice of Projective Rectification*. *International Journal of Computer Vision, IJCV*, 35(2):115–127, novembre–décembre 1999.
- [Hartley 04] R. I. HARTLEY et A. ZISSERMAN. *Multiple view geometry in computer vision*. Cambridge University Press, deuxième édition, 2004.
- [Hasler 03] D. HASLER, L. SBAIZ, S. SÜSTRUNK et M. VETTERLI. *Outlier Modelling in Image Matching*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 25(3):301–315, mars 2003.
- [Haxhimusa 02] Y. HAXHIMUSA et W. G. KROPATSCH. *Paths Lengths in Stochastic Graph Image Pyramid*. Dans proceedings of Workshop of the Austrian Association for Pattern Recognition, AAPR, pages 79–86, Graz, Autriche, septembre 2002.
- [Heisele 00] B. HEISELE. *Motion-Based Object Detection and Tracking in Color Sequences*. Dans proceedings of Asian Conference on Computer Vision, ACCV, volume 2, pages 1028–1033, Taïpeh, Taiwan, janvier 2000.
- [Hermosillo Valadez 02] G. HERMOSILLO VALADEZ. *Variational Methods for Multimodal Image Matching*. Thèse de doctorat, Université de Nice – Sophia Antipolis, France, mai 2002.
- [Hirschmüller 02] H. HIRSCHMÜLLER, P. R. INNOCENT et J. GARIBALDI. *Real-Time Correlation-Based Vision with Reduced Border Errors*. *International Journal of Computer Vision, IJCV*, 47(1–3):229–246, avril–juin 2002.
- [Hoff 89] W. HOFF et N. AHUJA. *Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 11(2):121–136, février 1989.
- [Holland 75] J. HOLLAND. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and, artificial intelligence*. Massachusetts Institute of Technology, MIT Press, 1975.
- [Hong 04] L. HONG et G. CHEN. *Segment-Based Stereo Matching Using Graph Cuts*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 74–81, Washington, États-Unis, juin–juillet 2004.
- [Hopfield 82] J. J. HOPFIELD. *Neural networks and physical systems with emergent collective computational abilities*. *Proceedings of the National Academy of Sciences*, 79(8):2254–2258, avril 1982.

- [Horaud 89] R. HORAUD et T. SKORDAS. *Stereo Correspondence Through Feature Grouping and Maximal Cliques*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI, 11(11):1168–1180, novembre 1989.
- [Horaud 93] R. HORAUD et O. MONGA. *Vision par ordinateur, outils fondamentaux*. Traité des nouvelles technologies, série Informatique. Hermès, Paris, France, 1993.
- [Horowitz 96] F. HOROWITZ, D. BONE et P. VEDKAMP. *Karhunen-Loeve Based Iterated Function System Encodings*. Dans *International Picture Coding Symposium*, volume 2, pages 409–413, Melbourne, Australie, mars 1996.
- [Hu 93] J. HU et P. SIY. *Stereo Correspondence through Multiple Constraint Neural Networks*. Dans *IEEE International Conference on Neural Networks*, volume 2, pages 332–342, San Francisco, États-Unis, mars 1993.
- [Huber 81] P. J. HUBER. *Robust statistics*. J. Wiley & Sons, New-York, États-Unis, 1981.
- [Intille 94] S. S. INTILLE et A. F. BOBICK. *Disparity-Space Images and Large Occlusion Stereo*. Dans *proceedings of European Conference on Computer Vision, ECCV*, volume 1, pages 179–186, Stockholm, Suède, mai 1994.
- [Irani 95] M. IRANI, P. ANANDAN et S. HSU. *Mosaic based representations of video sequences and their applications*. Dans *IEEE Conference Proceedings of International Conference on Computer Vision, ICCV*, pages 605–611, Cambridge, États-Unis, juin 1995.
- [Irani 02] M. IRANI, P. ANANDAN et M. COHEN. *Direct recovery of Planar-Parallax from Multiple Frames*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI, 24(11):1528–1534, novembre 2002.
- [Ishikawa 98] H. ISHIKAWA et D. GEIGER. *Occlusions, Discontinuities, and Epipolar Lines in Stereo*. Dans *proceedings of European Conference on Computer Vision, ECCV*, volume 1, pages 232–248, Fribourg, Allemagne, juin 1998.
- [Ishikawa 00] H. ISHIKAWA. *Global Optimization Using Embedded Graphs*. PhD thesis, Université de New-York, États-Unis, mai 2000.
- [Jaffré 03] G. JAFFRÉ et A. CROUZIL. *Non-Rigid Object Localization from Color Model using Mean Shift*. Dans *IEEE Conference Proceedings of International Conference on Image Processing, ICIP*, volume 3, pages 317–320, Barcelone, Espagne, septembre 2003.
- [Jagmohan 04] A. JAGMOHAN, M. SINGH et N. AHUJA. *Dense stereo matching using kernel maximum likelihood estimation*. Dans *IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR*, volume 3, pages 28–31, Cambridge, Royaume-Uni, août 2004.
- [Jahn 01] H. JAHN. *Binocular Stereo Matching by Local Attraction*. Dans *proceedings of International Conference on Computer Analysis of Images and Patterns, CAIP*, pages 676–683, Varsovie, Pologne, septembre 2001.
- [Jawahar 02] C. V. JAWAHAR et P. J. NARAYANAN. *Generalised correlation for multi-feature correspondence*. *The Journal of the Pattern Recognition Society, PR*, 35(6):1303–1313, juin 2002.
- [Jibrini 04] H. JIBRINI, M. PIERROT-DESEILLIGNY, N. PAPARODITIS et H. MAÎTRE. *Détermination d'une surface polyédrique continue optimale à partir d'un fouillis de plans*. Dans *actes du Congrès AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA*, volume 1, pages 175–183, Toulouse, France, janvier 2004.
- [Jolion 03] J. M. JOLION. *Stochastic Pyramid Revisited*. *Pattern Recognition Letters, PRL*, 24(8):1335–1342, mai 2003.
- [Jones 92] D. G. JONES et J. MALIK. *A Computational Framework for Determining Stereo Correspondence from a Set of Linear Spatial Filters*. *International Journal of Image and Vision Computing, IVC*, 10(10):699–708, décembre 1992.

- [Jones 97] G. A. JONES. *Constraint, Optimization, and Hierarchy: Reviewing Stereoscopic Correspondence of Complex Features*. *International Journal of Computer Vision, IJCV*, 65(1):57–58, janvier 1997.
- [Ju 04] X. JU, T. BOYLING, J. P. SIEBERT, N. MCFARLANE, J. WU et R. TILLET. *Integration of Range Images in a Multi-View Stereo System*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 4, pages 280–283, Cambridge, Royaume-Uni, août 2004.
- [Jung 01] I.-K. JUNG et S. LACROIX. *A Robust Interest Points Matching Algorithm*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 538–543, Vancouver, Canada, juillet 2001.
- [Jurie 01] F. JURIE. *Reconnaissance d'objets volumiques par mise en correspondance d'indices visuels*. *Traitement du signal, TS*, 18(5–6):321–345, 2001.
- [Kanade 94] T. KANADE et M. OKUTOMI. *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 16(9):920–932, septembre 1994.
- [Kanade 96] T. KANADE, A. YOSHIDA, K. ODA, H. KANO et M. TANAKA. *A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 196–202, San Francisco, États-Unis, juin 1996.
- [Kanatani 99] K. KANATANI et N. OHTA. *Accuracy Bounds and Optimal Computation of Homography for Image Mosaicing Applications*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 1, pages 73–78, Kerkyra, Grèce, septembre 1999.
- [Kaneko 02] S. KANEKO, I. MURASE et S. IGARASHI. *Robust Image registration by Increment Sign Correlation*. *The Journal of the Pattern Recognition Society, PR*, 35(10):2223–2234, octobre 2002.
- [Kaneko 03] S. KANEKO, Y. SATOH et S. IGARASHI. *Using selective correlation coefficient for robust image registration*. *The Journal of the Pattern Recognition Society, PR*, 36(5):1165–1173, mai 2003.
- [Kang 95] S. B. KANG, C. L. ZITNICK et T. KANADE. *A multibaseline stereo system with active illumination and real-time image acquisition*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 1, pages 88–93, Boston, États-Unis, juin 1995.
- [Kang 01] S. B. KANG, R. SZELISKI et J. CHAI. *Handling Occlusions in Dense Multi-view Stereo*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 103–110, Kauai, États-Unis, décembre 2001.
- [Kaplan 04] A. KAPLAN et E. RIVLIN. *Robust feature matching across widely separated color images*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 2, pages 136–139, Cambridge, Royaume-Uni, août 2004.
- [Kawanishi 04] T. KAWANISHI, T. KUROZUMI, K. KASHINO et S. TAKAGI. *A Fast Template Matching Algorithm with Adaptive Skipping Using Inner-Subtemplates' Distances*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 654–657, Cambridge, Royaume-Uni, août 2004.
- [Kim 05] C. KIM, K. M. LEE, B. T. CHOI et S. U. LEE. *A Dense Stereo Matching Using Two-Pass Dynamic Programming with Generalized Ground Control Points*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 2, pages 1075–1082, San Diego, États-Unis, juin 2005.
- [Klette 00] R. KLETTE, H. S. STIEHL, M. A. VIERGEVER et K. L. VINCKEN, éditeurs. *Performance characterization in computer vision*, volume 17 de *Computational Imaging and Vision*. Kluwer Academic Publishers, Dordrecht, Pays-Bas, 2000.

- [Koch 96] R. KOCH. *Surface Segmentation and Modeling of 3-D Polygonal Objects from Stereoscopic Image Pairs*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 233–237, Vienne, Autriche, juin 1996.
- [Koch 98] R. KOCH, M. POLLEFEYS et L. VAN GOOL. *Multi Viewpoint Stereo from Uncalibrated Video Sequences*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 1, pages 55–71, Fribourg, Allemagne, juin 1998.
- [Kolmogorov 01] V. KOLMOGOROV et R. ZABIH. *Computing Visual Correspondence with Occlusions using Graph Cuts*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 508–515, Vancouver, Canada, juillet 2001.
- [Kolmogorov 02] V. KOLMOGOROV et R. ZABIH. *Multi-camera Scene Reconstruction via Graph Cuts*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 2, pages 82–96, Copenhague, Danemark, mai 2002.
- [Kong 03] D. KONG, H. TAO et H. H. GONZALEZ-BANOS. *Sparse IBR Using Range Space Rendering*. Dans proceedings of British Machine Vision Conference, BMVC, volume 1, pages 181–190, Norwich, Royaume-Uni, septembre 2003.
- [Koschan 93] A. KOSCHAN. *Dense Stereo Correspondence Using Polychromatic Block Matching*. Dans proceedings of International Conference on Computer Analysis of Images and Patterns, CAIP, volume 719 de *Lecture Notes in Computer Science*, pages 538–542, Budapest, Hongrie, septembre 1993.
- [Koschan 96] A. KOSCHAN. *Using perceptual attributes to obtain dense depth maps*. Dans IEEE Conference Proceedings of Southwest Symposium on Image Analysis and Interpretation, pages 155–159, San Antonio, États-Unis, avril 1996.
- [Koschan 97] A. KOSCHAN et V. RODEHORST. *Dense Depth Maps by Active Color Illumination and Image Pyramids*. Dans F. SOLINA, W. G. KROPATSCH, R. KLETTE et R. BAJCSY, éditeurs, *Advances in Computer Vision*, pages 137–148. Springer-Verlag, New-York, octobre 1997.
- [Kostková 02] J. KOSTKOVÁ et R. ŠÁRA. *Stable Matching Based on Disparity Components*. Dans proceedings of Computer Vision Winter Workshop, pages 140–148, Bad Ausse, Autriche, février 2002.
- [Kostková 03] J. KOSTKOVÁ et R. ŠÁRA. *Stratified Dense Matching for Stereopsis in Complex Scenes*. Dans proceedings of British Machine Vision Conference, BMVC, volume 1, pages 339–348, Norwich, Royaume-Uni, septembre 2003.
- [Kropatsch 95] W. G. KROPATSCH. *Building Irregular Pyramids by Dual Graph Contraction*. *IEE Proceedings of Vision, Image and Signal Processing, VISP*, 142(6):366–374, décembre 1995.
- [Ku 01] J. S. KU, K. M. LEE et S. U. LEE. *Multi-image matching for a general motion stereo camera model*. *The Journal of the Pattern Recognition Society, PR*, 34(9):1701–1712, septembre 2001.
- [Lai 00] S. LAI. *Robust Image Matching Under Partial Occlusion and Spatially Varying Illumination Change*. *Computer Vision and Image Understanding, CVIU*, 78(1):84–98, avril 2000.
- [Lambert 02] P. LAMBERT. *Études méthodologiques du filtrage et de la segmentation d'images multicomposantes*. Habilitation à diriger des recherches, Université de Savoie, France, juillet 2002.
- [Lan 97] Z. D. LAN. *Méthodes robustes en vision: application aux appariements visuels*. Thèse de doctorat, Institut National Polytechnique, INP, Grenoble, France, mai 1997.
- [Lantagne 03] M. LANTAGNE, M. PARIZEAU et R. BERGEVIN. *VIP: Vision tool for comparing Images of People*. *Vision Interface*, 2003.
- [Laurence 98] L. LAURENCE. *Navigation sous-marine référencée terrain par mise en correspondance de cartes bathymétriques*. Thèse de doctorat, Université de Nice – Sophia Antipolis, France, janvier 1998.

- [Le Besnerais 98] G. LE BESNERAIS et H. ORIOT. *Disparity estimation for high resolution stereoscopic reconstruction using GNC approach*. Dans IEEE Conference Proceedings of International Conference on Image Processing, ICIP, volume 2, pages 594–597, Chicago, États-Unis, octobre 1998.
- [Leclerc 00] Y. G. LECLERC, Q. LUONG et P. FUA. *Measuring the Self-Consistency of Stereo Algorithms*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 1, pages 282–298, Dublin, Irlande, juin–juillet 2000.
- [Lee 91] H.-C. LEE et D. R. COK. *Detecting Boundaries in a Vector Field*. *IEEE Transactions on Signal Processing*, 39(5):1181–1194, mai 1991.
- [Lee 02] S. H. LEE, Y. KANATSUGU et J.-I. PARK. *MAP-Based Stochastic Diffusion for Stereo Matching and Line Fields Estimation*. *International Journal of Computer Vision, IJCV*, 47(1–3):195–218, avril 2002.
- [Lefèvre 96] V. LEFÈVRE, Y. POLLET, S. PHILIPP et S. BRUNESSAUX. *Un système multi-agents pour la fusion de données en analyse d'images*. *Traitement du signal, TS*, 13(1):100–111, janvier 1996.
- [Leung 04] C. LEUNG, B. APPLETON, B. C. LOVELL et C. SUN. *An Energy Minimisation Approach to Stereo-Temporal Dense Reconstruction*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 4, pages 72–75, Cambridge, Royaume-Uni, août 2004.
- [Levesque 04] D. LEVESQUE et F. DESCHÊNES. *Sparse Scene Structure Recovery from Atmospheric Degradation*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 84–87, Cambridge, Royaume-Uni, août 2004.
- [Levine 85] M. D. LEVINE. *Vision in man and machine*. McGraw-Hill Book Company, 1985.
- [Levy 00] A. LEVY et M. LINDENBAUM. *Sequential Karhunen-Loeve Basis Extraction and its Application to Images*. *IEEE Transactions on Image Processing, IP*, 9(8):1371–1374, août 2000.
- [Lezoray 03] O. LEZORAY, A. ELMOATAZ et H. CARDOT. *A color object recognition scheme: application to cellular sorting*. *International Journal of Machine Vision and Applications, MVA*, 14(3):166–176, juillet 2003.
- [Lhuillier 00] M. LHUILLIER et L. QUAN. *Robust Dense Matching Using Local and Global Geometric Constraints*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 968–972, Barcelone, Espagne, septembre 2000.
- [Lhuillier 04] M. LHUILLIER et L. QUAN. *Reconstruction quasi-dense de modèles 3D à partir d'une séquence d'images*. Dans actes du Congrès AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle, RFIA, volume 2, pages 895–904, Toulouse, France, janvier 2004.
- [Lin 02] M. H. LIN. *Surfaces with Occlusions from Layered Stereo*. PhD thesis, Université de Stanford, États-Unis, 2002.
- [Little 92] J. LITTLE. *Accurate early detection of discontinuities*. Dans proceedings of International Conference on Vision Interface, VI, pages 97–102, Vancouver, Canada, mai 1992.
- [Lotti 94] J. L. LOTTI et G. GIRAUDON. *Correlation algorithm with adaptive window for aerial image in stereo vision*. Dans proceedings of European Symposium on Satellite Remote Sensing, pages 2315–2325, Rome, Italie, septembre 1994.
- [Lundqvist 03] R. LUNDQVIST, E. BENGTTSSON et L. THURFJELL. *A Combined Intensity and Gradient-Based Similarity Criterion for Interindividual SPECT Brain Scan Registration*. *EURASIP Journal on Applied Signal Processing*, 2003(5):461–469, avril 2003.
- [Luo 95] A. LUO et H. BURKHARDT. *An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions*. *International Journal of Computer Vision, IJCV*, 15(3):171–188, juillet 1995.

- [Maier 03] D. MAIER, A. RÖSSLE, J. HESSER et R. MÄNNER. *Dense Disparity Maps Respecting Occlusions and Object Separation Using Partial Differential Equations*. Dans *Digital Image Computing: Techniques and Applications*, pages 613–622, Sydney, Australie, décembre 2003.
- [Mainmone 96] M. W. MAINMONE et S. A. SHAFER. *A Taxonomy for Stereo Computer Vision Experiments*. Dans *proceedings of European Conference on Computer Vision, ECCV, Workshop on Performance Characteristics of Vision Algorithms*, pages 59–79, Cambridge, Royaume-Uni, avril 1996.
- [Manduchi 99] R. MANDUCHI et C. TOMASI. *Distinctiveness Maps for Image Matching*. Dans *proceedings of International Conference on Image Analysis and Processing, ICIAP*, pages 26–31, Venise, Italie, septembre 1999.
- [Marr 76] D. MARR et T. POGGIO. *Cooperative Computation of Stereo Disparity*. *Science*, 194(4262):283–287, 1976.
- [Matas 02] J. MATAS, O. CHUM, M. URBAN et T. PAJDLA. *Robust Wide Baseline Stereo from Maximally Stable Extremal Regions*. Dans *proceedings of British Machine Vision Conference, BMVC*, volume 1, pages 384–393, Cardiff, Royaume-Uni, septembre 2002.
- [Matthews 03] I. MATTHEWS, T. ISHIKAWA et S. BAKER. *The Template Update Problem*. Dans *proceedings of British Machine Vision Conference, BMVC*, volume 2, pages 649–658, Norwich, Royaume-Uni, septembre 2003.
- [Mayer 03] H. MAYER. *Analysis of Means to Improve Cooperative Disparity Estimation*. Dans *proceedings of ISPRS Conference on Photogrammetric Image Analysis, PIA*, pages 25–31, Munich, Allemagne, septembre 2003.
- [Mayoral 04] R. MAYORAL et M. AURNHAMMER. *Evaluation of correspondence errors for stereo*. Dans *IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR*, volume 4, pages 104–107, Cambridge, Royaume-Uni, août 2004.
- [McCane 01] B. MCCANE, K. NOVINS, D. CRANNITCH et B. GALVIN. *On Benchmarking Optical Flow*. *Computer Vision and Image Understanding, CVIU*, 84(1):126–143, octobre 2001.
- [Meer 00] P. MEER, C. V. STEWART et D. E. TYLER. *Robust Computer Vision: An Interdisciplinary Challenge*. *Computer Vision and Image Understanding, CVIU*, 78(1):1–7, juillet 2000.
- [Megyesi 04] Z. MEGYESI et D. CHETVERIKOV. *Affine Propagation for Surface Reconstruction in Wide Baseline Stereo*. Dans *IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR*, volume 4, pages 76–79, Cambridge, Royaume-Uni, août 2004.
- [Menard 97] C. MENARD et W. G. KROPATSCH. *Adaptive Stereo Matching in Correlation Scale-Space*. Dans *proceedings of International Conference on Image Analysis and Processing, ICIAP*, volume 2, pages 677–684, Florence, Italie, septembre 1997.
- [Messer 04] K. MESSER, J. KITTLER, M. SADEGHI, M. HAMOUZ, A. KOSTIN, F. CARDINAUX, S. MARCEL, S. BENGIO, C. SANDERSON, N. POH, Y. RODRIGUEZ, J. CZYZ, L. VANDENDORPE, C. MCCOOL, S. LOWTHER, S. SRIDHARAN, V. CHANDRAN, R. PARADES PALACIOS, E. VIDAL, L. BAI, L. SHEN, Y. WANG, Y.-H. CHIANG, H.-C. LIU, Y.-P. HUNG, A. HEINRICHS, M. MÜLLER, A. TEWES, C. VON DER MALSBERG, R. WÜRTZ, Z. WANG, F. XUE, Y. MA, Q. YANG, C. FANG, X. DING, S. LUCEY, R. GOSS et H. SCHNEIDERMAN. *Face Authentication Test on the BANCA Database*. Dans *IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR*, volume 4, pages 523–532, Cambridge, Royaume-Uni, août 2004.
- [Metropolis 53] N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH et A. H. TELLER. *Equation of State Calculations by Fast Computing Machines*. *The Journal of Chemical Physics*, 21(6):1087–1092, juin 1953.

- [Mohan 89] R. MOHAN, G. MEDIONI et R. NEVATIA. *Stereo Error Detection, Correction, and Evaluation*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 11(2):113–120, février 1989.
- [Montanvert 91] A. MONTANVERT, P. MEER et A. ROSENFELD. *Hierarchical Image Analysis Using Irregular Tessellations*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 13(4):307–316, avril 1991.
- [Moravec 80] H. MORAVEC. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD thesis, Université de Carnegie Mellon, CMU, Pittsburgh, États-Unis, septembre 1980.
- [Mulligan 00] J. MULLIGAN et K. DANILIDIS. *Predicting Disparity Windows for Real-Time Stereo*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 1, pages 220–235, Dublin, Irlande, juin–juillet 2000.
- [Mulligan 01] J. MULLIGAN, V. ISLER et K. DANILIDIS. *Performance of Stereo for Tele-presence*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 558–565, Vancouver, Canada, juillet 2001.
- [Muñoz Barrutia 00] A. MUÑOZ BARRUTIA, T. BLU et M. UNSER. *Non-Euclidean Pyramids*. Dans proceedings of Mathematical Imaging: Wavelet Applications in Signal and Image Processing, volume SPIE–4119, pages 710–720, San Diego, États-Unis, juillet 2000.
- [Mühlmann 01] K. MÜHLMANN, D. MAIER, J. HESSER et R. MÄNNER. *Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation*. Dans IEEE Conference Proceedings of Workshop on Stereo and Multi-Baseline Vision, SMBV, pages 30–36, Kauai, États-Unis, juin 2001.
- [Nack 75] M. L. NACK. *Temporal registration of multispectral digital satellite images using their edge images*. Dans AAS/AIAA/Astrodynamics Specialist Conference, Nassau, Bahamas, juillet 1975. Papier AAS75–104.
- [Nakamura 96] Y. NAKAMURA, T. MATSUURA, K. SATOH et Y. OHTA. *Occlusion Detectable Stereo – Occlusion Patterns in Camera Matrix*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 371–378, San Francisco, États-Unis, juin 1996.
- [Nasrabadi 92a] N. M. NASRABADI. *A Stereo Vision Technique Using Curve-Segments and Relaxation Matching*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 14(5):566–572, mai 1992.
- [Nasrabadi 92b] N. M. NASRABADI et C. Y. CHOO. *Hopfield Network for Stereo Vision Correspondence*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 3(1):5–13, janvier 1992.
- [Nishihara 84] H. K. NISHIHARA. PRISM, a practical real-time imaging stereo matcher. Rapport de recherche A. I. Memo 780, Massachusetts Institute of Technology, MIT, États-Unis, 1984.
- [Nishihara 88] H. K. NISHIHARA et P. A. CROSSLEY. *Measuring Photographic Overlay Accuracy and Critical Dimensions by Correlating Binarized Laplacian of Gaussian Convolutions*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 10(1):17–30, janvier 1988.
- [Ohta 80] Y.-I. OHTA, T. KANADE et T. SAKAI. *Color Information for Region Segmentation*. *Computer Graphics and Image Processing, CGIP*, 13(3):222–241, juillet 1980.
- [Ohta 85] Y.-I. OHTA et T. KANADE. *Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 7(2):139–154, mars 1985.
- [Okutomi 92a] M. OKUTOMI et T. KANADE. *A Locally Adaptive Window For Signal Matching*. *International Journal of Computer Vision, IJCV*, 7(2):143–162, janvier 1992.

- [Okutomi 92b] M. OKUTOMI et G. TOMITA. *Color Stereo Matching and Its Application to 3-D Measurement of Optic Nerve Head*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 509–513, La Haye, Hollande, septembre 1992.
- [Okutomi 93] M. OKUTOMI et T. KANADE. *A Multiple-Baseline Stereo*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 15(4):358–363, avril 1993.
- [Orteu 03] J.-J. ORTEU, D. GARCIA et M. DEVY. *Mesure de formes et de champs de déplacements tridimensionnels par stéréo-corrélation d'images : applications en mécanique expérimentale*. *Photoniques – La Revue de la Société Française d'Optique*, (10):34–43, juin 2003.
- [Pajares 03] G. PAJARES et J. M. DE LA CRUZ. *Stereovision matching through support vector machines*. *Pattern Recognition Letters, PRL*, 24(15):2575–2583, novembre 2003.
- [Pan 96] H.-P. PAN. *General Stereo Matching Using Symmetric Complex Wavelets*. Dans Wavelet Applications in Signal and Image Processing IV, volume SPIE–2825, pages 697–721, Denver, États-Unis, août 1996.
- [Paparoditis 98] N. PAPANODITIS, M. CORD, M. JORDAN et J.-P. COCQUEREZ. *Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery*. *Computer Vision and Image Understanding, CVIU*, 72(2):122–142, novembre 1998.
- [Paris 02] S. PARIS et F. SILLION. *Optimisation à base de flot de graphe pour l'acquisition d'informations 3D à partir de séquences d'images*. Dans Actes des 15^{ème} journées de l'AFIG, Association Française d'Informatique Graphique, pages 165–182, Villeurbanne, France, décembre 2002.
- [Park 01] C. S. PARK et H. W. PARK. *A robust stereo disparity estimation using adaptive window search and dynamic programming search*. *The Journal of the Pattern Recognition Society, PR*, 34(12):2573–2576, décembre 2001.
- [Park 03] S. Y. PARK, J. S. KIM, S. H. LEE, N. I. CHO, G.-S. LEE et C.-H. AHN. *Composition of High-Resolution Stereo Images from Hybrid Stereo System*. Dans proceedings of International Workshop on Advanced Image Technology, WAIT, Nagasaki, Japon, janvier 2003.
- [Pearl 88] J. PEARL. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers, San Francisco, États-Unis, 1988.
- [Peleg 90] S. PELEG et G. RON. *Nonlinear Multiresolution: A Shape-from-shading Example*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 12(12):1206–1210, décembre 1990.
- [Perwass 01] C. B. U. PERWASS et G. SOMMER. *A Fuzzy Logic Algorithm for Dense Image Point Matching*. Dans proceedings of International Conference on Vision Interface, VI, pages 39–47, Ottawa, Canada, juin 2001.
- [Petrakis 02] E. G. M. PETRAKIS, A. DIPLAROS et E. MILIOS. *Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 24(11):1501–1516, novembre 2002.
- [Pollefeys 02] M. POLLEFEYS, F. VERBIEST et L. V. GOOL. *Surviving dominant planes in uncalibrated structure and motion recovery*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 2, pages 837–851, Copenhague, Danemark, mai 2002.
- [Porter 03] S. PORTER, M. MIRMEHDI et B. THOMAS. *Video Indexing using Motion Estimation*. Dans proceedings of British Machine Vision Conference, BMVC, volume 2, pages 659–668, Norwich, Royaume-Uni, septembre 2003.
- [Pratt 78] W. K. PRATT. *Digital image processing*. Wiley-Interscience Publication, New-York, États-Unis, 1978.
- [Pritchett 98] P. PRITCHETT et A. ZISSERMAN. *Matching and Reconstruction from Widely Separated Views*. Dans proceedings of European Workshop on 3D Structure from Multiple Images of Large

- Scale Environments, SMILE, volume 1506 de *Lecture Notes in Computer Science*, pages 78–92, Fribourg, Allemagne, juin 1998.
- [Pugeault 03] N. PUGEAULT et N. KRÜLGER. *Multi-Modal Matching Applied to Stereo*. Dans proceedings of British Machine Vision Conference, BMVC, volume 1, pages 271–280, Norwich, Royaume-Uni, septembre 2003.
- [Puzicha 97] J. PUZICHA, T. HOFMANN et J. M. BUHMANN. *Non-parametric Similarity Measures for Unsupervised Texture Segmentation and Image Retrieval*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 267–272, San Juan, Porto Rico, juin 1997.
- [Pétillot 05] I. PÉTILLOT. Reconstruction du relief d'un document à partir de ses contours. Rapport de DEA, Université Paul Sabatier, UPS, Toulouse, France, juin 2005.
- [Qian 02] G. QIAN, S. SURAL et S. PRAMANIK. *A comparative analysis of two distance measures in color image databases*. Dans IEEE Conference Proceedings of International Conference on Image Processing, ICIP, volume 1, pages 401–404, Rochester, États-Unis, septembre 2002.
- [Rao 03] C. RAO, A. GRITAI et M. SHAH. *View-invariant Alignment and Matching of Video Sequences*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 939–945, Nice, France, octobre 2003.
- [Reddy 04] V. S. K. REDDY et S. SENGUPTA. *A New Predictive Full-Search Block Motion Estimation*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 4, pages 721–724, Cambridge, Royaume-Uni, août 2004.
- [Rital 04] S. RITAL, H. CHERFI et D. ABOUTAJDINE. *A vectorial Edge detector using spatiocolorimetric neighborhood hypergraph and perceptual color distance*. Dans International Conference on Complex Systems, Intelligence and Modern Technology, pages 305–310, Cherbourg, France, septembre 2004.
- [Robert 96] L. ROBERT et R. DERICHE. *Dense Depth Map Reconstruction: A Minimization and Regularization Approach which Preserves Discontinuities*. Dans proceedings of European Conference on Computer Vision, ECCV, volume A, pages 439–451, Cambridge, Royaume-Uni, avril 1996.
- [Roche 01] A. ROCHE, X. PENNEC, G. MALANDAIN et N. AYACHE. *Rigid Registration of 3-D Ultrasound with MR Images: A New approach Combining Intensity and Gradient Information*. *IEEE Transactions on Medical Imaging*, 20(10):1038–1049, octobre 2001.
- [Rogelj 03] P. ROGELJ, S. KOVAČIČ et J. C. GEE. *Point similarity measures for non-rigid registration of multi-modal data*. *Computer Vision and Image Understanding, CVIU*, 92(1):112–140, octobre 2003.
- [Rousseeuw 87] P. J. ROUSSEEUW et A. M. LEROY. Robust regression and outlier detection. J. Wiley & Sons, New-York, États-Unis, 1987.
- [Rousseeuw 92] P. J. ROUSSEEUW et C. CROUX. *L₁-Statistical Analysis and Related Methods*. Dans Y. DODGE, éditeur, *Explicit Scale Estimators with High Breakdown Point*, pages 77–92. Elsevier, Amsterdam, Hollande, 1992.
- [Roy 98] S. ROY et I. J. COX. *A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, pages 492–499, Bombai, Inde, janvier 1998.
- [Rubner 01] Y. RUBNER, J. PUZICHA, C. TOMASI et J. M. BUHMANN. *Empirical Evaluation of Dissimilarity Measures for Color and Texture*. *Computer Vision and Image Understanding, CVIU*, 84(1):25–43, octobre 2001.
- [Russ 99] T. D. RUSS et A. P. REEVES. *High Accuracy Depth Measurement using Multi-view Stereo*. Dans proceedings of International Conference on Vision Interface, VI, pages 103–110, Trois-Rivières, Canada, mai 1999.

- [Rziza 00] M. RZIZA, A. TAMTAOUI, L. MORIN et D. ABOUTAJDINE. *Estimation and segmentation of a dense disparity map for 3D reconstruction*. Dans IEEE Conference Proceedings of International Conference on Acoustics, Speech and Signal Processing, ICASSP, Istamboul, Turquie, mars 2000.
- [Rziza 01] M. RZIZA, D. ABOUTAJDINE, L. MORIN et A. TAMTAOUI. *Schéma multirésolution d'estimation d'un champ de disparités dense sous contrainte épipolaire pour les images bruitées*. Dans actes du colloque GRETSI sur le traitement du signal et des images, Toulouse, France, septembre 2001.
- [Saito 95] H. SAITO et M. MORI. *Application of genetic algorithms to stereo matching of images*. *Pattern Recognition Letters, PRL*, 16(8):815–821, août 1995.
- [Sankar Kishore 00] M. SANKAR KISHORE et K. VEERABHADRA RAO. *A study of correlation technique on pyramid processed images*. *Sādhanā, Academy Proceedings in Engineering Sciences*, 25(1):37–43, février 2000.
- [Sankar Kishore 01] M. SANKAR KISHORE et K. VEERABHADRA RAO. *Robust correlation tracker*. *Sādhanā, Academy Proceedings in Engineering Sciences*, 26(3):227–236, juin 2001.
- [Satoh 96] K. SATOH et Y. OHTA. *Occlusion Detectable Stereo – Systematic Comparison of Detection Algorithms*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 280–286, Vienne, Autriche, août 1996.
- [Scharstein 98] D. SCHARSTEIN et R. SZELISKI. *Stereo Matching with Non-Linear Diffusion*. *International Journal of Computer Vision, IJCV*, 28(2):155–174, juin 1998.
- [Scharstein 02] D. SCHARSTEIN et R. SZELISKI. *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*. *International Journal of Computer Vision, IJCV*, 47(1):7–42, avril 2002.
- [Scharstein 03] D. SCHARSTEIN et R. SZELISKI. *High-Accuracy Stereo Depth Maps Using Structured Light*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 195–202, Madison, États-Unis, juin 2003.
- [Schindler 03] K. SCHINDLER. *Generalized Use of Homographies for Piecewise Planar Reconstruction*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 470–476, Göteborg, Suède, juin 2003.
- [Schlögl 04] T. SCHLÖGL, C. BELEZNAI, M. WINTER et H. BISCHOF. *Performance Evaluation Metrics for Motion Detection and Tracking*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 4, pages 519–522, Cambridge, Royaume-Uni, août 2004.
- [Schlüns 00] K. SCHLÜNS et A. KOSCHAN. *Global and Local Highlight Analysis In Color Images*. Dans proceedings of International Conference on Color in Graphics and Image Processing, CGIP, pages 300–304, Saint-Étienne, France, octobre 2000.
- [Schmid 95] C. SCHMID et R. MOHR. *Matching by local invariants*. Rapport de recherche RR-2644, Institut National de Recherche en Informatique et en Automatique, INRIA, août 1995.
- [Schmid 97] C. SCHMID et A. ZISSERMAN. *Automatic Line Matching across Views*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 666–671, San Juan, Porto Rico, juin 1997.
- [Schreier 00] H. W. SCHREIER, J. R. BRAASCH et M. A. SUTTON. *Systematic errors in digital image correlation caused by intensity interpolation*. *Optical engineering*, 39(11):2915–2921, novembre 2000.
- [Seitz 89] P. SEITZ. *Using Local Orientational Information as Image Primitive for Robust Object Recognition*. Dans Visual Communication and Image Processing IV, volume SPIE-1199, pages 1630–1639, 1989.
- [Semani 05] D. SEMANI, M. CHAMBAH et P. COURTELLEMONT. *Processing of Underwater Colour Images Applied to Live Aquarium Videos*. *International Journal of Robotics and Automation, IJRA*, 20(2):123–130, 2005.

- [Shafique 03] K. SHAFIQUE et M. SHAH. *A Non-Iterative Greedy Algorithm for Multi-frame Point Correspondence*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 1, pages 111–115, Nice, France, octobre 2003.
- [Shao 02] J. SHAO. *Generation of Temporally Consistent Multiple Virtual Camera Views from Stereoscopic Image Sequences*. *International Journal of Computer Vision, IJCV*, 47(2):171–180, avril 2002.
- [Sharma 97] G. SHARMA et H. J. TRUSELL. *Digital Color Imaging*. *IEEE Transactions on Image Processing, IP*, 6(7):901–932, juillet 1997.
- [Shekarforoush 95] H. SHEKARFOROUSH, M. BERTHOD et J. ZERUBIA. Subpixel image registration by estimating the polyphase decomposition of the cross power spectrum. Rapport de recherche RR-2707, Institut National de Recherche en Informatique et en Automatique, INRIA, Sophia-Antipolis, France, novembre 1995.
- [Shen 92] J. SHEN et S. CASTAN. *An optimal linear operator for step edge detection*. *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing, CVGIP*, 24(2):112–133, mars 1992.
- [Shi 94] J. SHI et C. TOMASI. *Good features to track*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 593–600, Seattle, États-Unis, juin 1994.
- [Shimizu 05] M. SHIMIZU et M. OKUTOMI. *Sub-Pixel Estimation Error Cancellation on Area-Based Matching*. *International Journal of Computer Vision, IJCV*, 63(3):207–224, juillet 2005.
- [Silva 00] C. SILVA et J. SANTOS-VICTOR. *Intrinsic Images for Dense Stereo Matching with Occlusions*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 1, pages 100–114, Dublin, Irlande, juin–juillet 2000.
- [Simonetto 05] E. SIMONETTO et A. SOADANE. *Colour Image Watermarking Using a Visual Sub-Band Decomposition*. *International Journal of Robotics and Automation, IJRA*, 20(2):101–108, 2005.
- [Škerl 04] D. ŠKERL, B. LIKAR et F. PERNUŠ. *Evaluation of Nine Similarity Measures used in Rigid Registration*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 794–797, Cambridge, Royaume-Uni, août 2004.
- [Smith 96] P. W. SMITH et N. NANDHAKUMAR. *An Improved Power Cepstrum Based Stereo Correspondence Method for Textured Scenes*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 18(3):338–348, mars 1996.
- [Smith 97] S. SMITH et J. BRADY. *SUSAN - A New Approach to Low Level Image Processing*. *International Journal of Computer Vision, IJCV*, 23(1):45–78, mai 1997.
- [Smith 98] P. SMITH, D. SINCLAIR, R. CIPOLLA et K. WOOD. *Effective Corner Matching*. Dans proceedings of British Machine Vision Conference, BMVC, pages 545–556, Southampton, Royaume-Uni, septembre 1998.
- [Smolka 03] B. SMOLKA. *On the Similarity Based Impulsive Noise Removal Technique for Multi-channel Images*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 1066–1073, Göteborg, Suède, juin 2003.
- [Starink 95] J. P. P. STARINK et E. BACKER. *Finding point correspondences using simulated annealing*. *The Journal of the Pattern Recognition Society, PR*, 28(2):231–240, février 1995.
- [Stock 03] C. STOCK et A. PINZ. *Similarity Measure for Corner Redetection*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 133–139, Göteborg, Suède, juin 2003.
- [Stricker 95] M. STRICKER et M. ORENGO. *Similarity of Color Images*. Dans Storage and Retrieval for Image and Video Databases, volume SPIE-2420, pages 381–392, San José, États-Unis, février 1995.
- [Sun 97] C. SUN. *A Fast Stereo Matching Method*. Dans proceedings of Digital Image Computing: Techniques and Applications, DICTA, pages 95–100, Auckland, Nouvelle-Zélande, décembre 1997.

- [Sun 02] C. SUN. *Fast optical flow using 3D shortest path techniques*. *International Journal of Image and Vision Computing, IVC*, 20(13–14):981–991, décembre 2002.
- [Sun 03] J. SUN. *Stereo Matching Using Belief Propagation*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 25(7):787–800, décembre 2003.
- [Sun 05] J. SUN, Y. LI, S. B. KANG et H.-Y. SHUM. *Symmetric Stereo Matching for Occlusion Handling*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 2, pages 399–406, San Diego, États-Unis, juin 2005.
- [Swain 91] M. J. SWAIN et D. H. BALLARD. *Color Indexing*. *International Journal of Computer Vision, IJCV*, 7(1):11–32, novembre 1991.
- [Szeliski 99a] R. SZELISKI. *Prediction Error as a Quality Metric for Motion and Stereo*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 781–788, Kerkyra, Grèce, septembre 1999.
- [Szeliski 99b] R. SZELISKI et R. ZABIH. *An experimental Comparison of Stereo Algorithms*. Dans ICCV Workshop on Vision Algorithms: Theory and Practice, volume 1883 de *Lecture Notes in Computer Science*, pages 1–19, Kerkyra, Grèce, septembre 1999.
- [Szeliski 02] R. SZELISKI et D. SCHARSTEIN. *Symmetric Sub-Pixel Stereo Matching*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 2, pages 525–540, Copenhague, Danemark, mai 2002.
- [Szeliski 04] R. SZELISKI. *Sampling the Disparity Space Image*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 26(3):419–425, mars 2004.
- [Šára 97] R. ŠÁRA et R. BAJCSY. *On Occluding Contour Artifacts in Stereo Vision*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 852–857, San Juan, Porto Rico, juin 1997.
- [Séchan 91] Renaud SÉCHAN. *Dans ton sac*. Dans Marchands de cailloux, 1991.
- [Tao 00] H. TAO et H. S. SAWHNEY. *Global Matching Criterion and Color Segmentation Based Stereo*. Dans IEEE Conference Proceedings of Workshop on Applications of Computer Vision, WACV, pages 246–253, Palm Springs, États-Unis, décembre 2000.
- [Tao 01] H. TAO, H. S. SAWHNEY et R. KUMAR. *A Global Matching Framework for Stereo Computation*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 532–539, Vancouver, Canada, juillet 2001.
- [Tappen 03] M. F. TAPPEN et W. T. FREEMAN. *Comparison of Graph Cuts with Belief Propagation for Stereo, using Identical MRF Parameters*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 900–907, Nice, France, octobre 2003.
- [Tesser 03] H. TESSER et T. TROUT. *A Note on Evaluation of Image Recognition Systems*. Dans proceedings of Scandinavian Conference on Image Analysis, SCIA, pages 60–66, Göteborg, Suède, juin 2003.
- [Tian 86] Q. TIAN et M. N. HUHS. *Algorithms for Subpixel Registration*. *Computer Vision, Graphics, and Image Processing: Image Understanding, CVGIP*, 35(2):220–233, août 1986.
- [Torr 98] P. H. S. TORR et A. ZISSERMAN. *Robust Computation and Parametrization of Multiple View Relations*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, pages 727–732, Bombay, Inde, janvier 1998.
- [Torr 04] P. H. S. TORR et A. CRIMINISI. *Dense stereo using pivoted dynamic programming*. *International Journal of Image and Vision Computing, IVC*, 22(10):795–806, septembre 2004.
- [Triantafyllidis 00] G. A. TRIANTAFYLIDIS et M. G. STRINTZIS. *Occlusion and Visible Background and Foreground Areas in Stereo: A Bayesian Approach*. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(4):563–575, juillet 2000.

- [Trujillo 04] M. TRUJILLO et E. IZQUIERDO. *A robust correlation measure for correspondence estimation*. Dans proceedings of International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT, pages 155–162, Thessalonique, Grèce, septembre 2004.
- [Trémeau 03] A. TRÉMEAU et P. COLANTONI. *Visualisation 3D dédiée à l'analyse des images couleur*. Dans actes du congrès francophone de Vision par Ordinateur, ORASIS, pages 261–268, Gérardmer, France, mai 2003.
- [Trémeau 04] A. TRÉMEAU, C. FERNANDEZ-MALOIGNE et P. BONTON, éditeurs. *Image numérique couleur – de l'acquisition au traitement*. Dunod, janvier 2004.
- [Tsai 99] C. TSAI et A. K. KATSAGGELOS. *Dense Disparity Estimation with a Divide-and-Conquer Disparity Space Image Technique*. *IEEE Transactions on Multimedia*, 1(1):18–29, mars 1999.
- [Tsai 03] D.-M. TSAI, C.-T. LIN et J.-F. CHEN. *The evaluation of normalized cross correlations for defect detection*. *Pattern Recognition Letters, PRL*, 24(15):2525–2535, novembre 2003.
- [Tseng 92] D.-C. TSENG et C.-H. CHANG. *Color Segmentation Using Perceptual Attributes*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 228–231, La Haye, Hollande, septembre 1992.
- [Tuytelaars 00] T. TUYTELAARS et L. VAN GOOL. *Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions*. Dans proceedings of British Machine Vision Conference, BMVC, volume 1, pages 412–422, Bristol, Royaume-Uni, septembre 2000.
- [Uenohara 97] M. UENOHARA et T. KANADE. *Use of Fourier and Karhunen-Loeve Decomposition for Fast Pattern Matching With a Large Set of Templates*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 19(8):891–898, août 1997.
- [Ullah 01] F. ULLAH, S. KANEKO et S. IGARASHI. *Orientation Code Matching For Robust Object Search*. *Proceedings of IEICE Transactions on Information and Systems*, E-84-D(8):999–1006, mars 2001.
- [Unnikrishnan 05] R. UNNIKRISHNAN et M. HEBERT. *Measures of Similarity*. Dans IEEE Conference Proceedings of Workshop on Applications of Computer Vision, WACV, volume 1, pages 394–400, Breckenridge, États-Unis, janvier 2005.
- [Vandenbroucke 00] N. VANDENBROUCKE. *Segmentation d'images couleur par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'images de football*. Thèse de doctorat, Université des sciences et technologies de Lille 1, France, décembre 2000.
- [Veksler 99] O. VEKSLER. *Efficient graph-based energy minimization methods in computer vision*. PhD thesis, Université de Cornell, Ithaca, États-Unis, août 1999.
- [Veksler 01] O. VEKSLER. *Dense Features for Semi-Dense Stereo Correspondence*. Dans IEEE Conference Proceedings of Workshop on Stereo and Multi-Baseline Vision, SMBV, pages 149–157, Kauai, États-Unis, juin 2001.
- [Veksler 03] O. VEKSLER. *Fast Variable Window for Stereo Correspondence using Integral Images*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 565–570, Madison, États-Unis, juin 2003.
- [Veksler 05] O. VEKSLER. *Stereo Correspondence by Dynamic Programming on a Tree*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 2, pages 384–390, San Diego, États-Unis, juin 2005.
- [Viterbi 67] A. J. VITERBI. *Error bounds for convolutional codes and an asymptotically optimum decoding algorithm*. *IEEE Transactions on Information Theory*, 13(2):260–269, avril 1967.
- [Wang 92] Y. WANG et D. WIENS. *Optimal, robust R-estimators and test statistics in the linear model*. *Statistics and Probability Letters, SPL*, 14:179–188, juin 1992.

- [Wang 93] J. Y. A. WANG et E. H. ADELSON. *Layered representation for motion analysis*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 361–366, New-York, États-Unis, juin 1993.
- [Wang 99] J. WANG et C. HSIAO. *On Disparity Matching in Stereo Vision via a Neural Network Framework*. *National Science Council*, 23(5):665–678, mars 1999.
- [Wei 98] Q.-Q. WEI, W. BRAUER et G. HIRZINGER. *Intensity- and Gradient-Based Stereo Matching Using Hierarchical Gaussian Basis Functions*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 20(11):1143–1160, novembre 1998.
- [Wei 04] Y. WEI et L. QUAN. *Region-based Progressive Stereo Matching*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 1, pages 106–113, Washington, États-Unis, juin–juillet 2004.
- [Weiss 01] Y. WEISS et W. T. FREEMAN. *On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs*. *IEEE Transactions on Information Theory*, 47(2):736–744, février 2001.
- [Weng 92] J. WENG, N. AHUJA et T. S. HUANG. *Matching Two Perspective Views*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 14(8):806–825, juillet 1992.
- [Wesolkowski 99] S. WESOLKOWSKI et E. JERNIGAN. *Color edge detection in RGB using jointly euclidean distance and vector angle*. Dans proceedings of International Conference on Vision Interface, VI, pages 9–16, Trois-Rivières, Canada, mai 1999.
- [Whitted 80] T. WHITTED. *An improved illumination model for shaded display*. *Communications of the ACM*, 23(6):343–349, 1980.
- [Wiens 94] D. WIENS et J. ZHOU. *Bounded-influence rank estimation in the linear model*. *The Canadian Journal of Statistics*, 22(2):233–245, 1994.
- [Withagen 04] P. WITHAGEN, K. SCHUTTE et F. GROEN. *Probabilistic Classification between Foreground Objects and Background*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 31–34, Cambridge, Royaume-Uni, août 2004.
- [Woetzel 04] J. WOETZEL et R. KOCH. *Multi-Camera Real-Time Depth Estimation with Discontinuity Handling on PC Graphics Hardware*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 1, pages 741–744, Cambridge, Royaume-Uni, août 2004.
- [Xiao 03] J. XIAO et M. SHAH. *Two-Frame Wide Baseline Matching*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 1, pages 603–609, Nice, France, octobre 2003.
- [Xu 97] G. XU. *A Unified Approach to Image Matching and Segmentation in Stereo, Motion, and Object Recognition via Recovery of Epipolar Geometry*. *Videre: Journal of Computer Vision Research*, 1(1):22–54, 1997.
- [Yang 93] Y. YANG, A. YUILLE et J. LU. *Local, Global, and Multilevel Stereo Matching*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, pages 274–279, New-York, États-Unis, juin 1993.
- [Yao 03] A. YAO et A. CALWAY. *Dense 3-D Structure from Image Sequences Using Probabilistic Depth Carving*. Dans proceedings of British Machine Vision Conference, BMVC, volume 1, pages 211–220, Norwich, Royaume-Uni, septembre 2003.
- [Yedidia 05] J. S. YEDIDIA, W. T. FREEMAN et Y. WEISS. *Constructing free-energy approximations and generalized belief propagation algorithms*. *IEEE Transactions on Information Theory*, 51(7):2282–2312, juillet 2005.

- [Yoon 05] K.-J. YOON et I.-S. KWEON. *Locally Adaptive Support-Weight Approach for Visual Correspondence Search*. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 2, pages 924–931, San Diego, États-Unis, juin 2005.
- [Zabih 94] R. ZABIH et J. WOODFILL. *Non-parametric Local Transforms for Computing Visual Correspondence*. Dans proceedings of European Conference on Computer Vision, ECCV, pages 151–158, Stockholm, Suède, mai 1994.
- [Zhang 93] Z. ZHANG. Le problème de la mise en correspondance : l'état de l'art. Rapport de recherche RR-2146, Institut National de Recherche en Informatique et en Automatique, INRIA, décembre 1993.
- [Zhang 95] Z. ZHANG. Parameter estimation techniques: A tutorial with application to conic fitting. Rapport de recherche RR-2676, Institut National de Recherche en Informatique et en Automatique, INRIA, octobre 1995.
- [Zhang 96] Z. ZHANG. Determining the epipolar geometry and its uncertainty: A review. Rapport de recherche RR-2927, Institut National de Recherche en Informatique et en Automatique, INRIA, juillet 1996.
- [Zhang 00] Z. ZHANG et Y. SHAN. *A Progressive Scheme for Stereo Matching*. Dans proceedings of European Workshop on 3D Structure from Multiple Images of Large-Scale Environments, SMILE, volume 2018 de *Lecture Notes in Computer Science*, pages 68–85, Dublin, Irlande, juillet 2000.
- [Zhang 02] Y. ZHANG et C. KAMBHAMETTU. *Stereo Matching with Segmentation-based Cooperation*. Dans proceedings of European Conference on Computer Vision, ECCV, volume 2, pages 556–571, Copenhague, Danemark, mai 2002.
- [Zhang 04] X. ZHANG et Y. LIU. *Point Pattern Matching for Articulated or Multiple Objects*. Dans IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR, volume 3, pages 630–633, Cambridge, Royaume-Uni, août 2004.
- [Zickler 03] T. E. ZICKLER, J. HO, D. J. KRIEGMAN, J. PONCE et P. N. BELHUMEUR. *Binocular Helmholtz Stereopsis*. Dans IEEE Conference Proceedings of International Conference on Computer Vision, ICCV, volume 2, pages 1411–1417, Nice, France, octobre 2003.
- [Zitnick 00] C. L. ZITNICK et T. KANADE. *A cooperative Algorithm for Stereo Matching and Occlusion Detection*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 22(7):675–684, juillet 2000.
- [Zoghلامي 96] I. ZOGLAMI, O. FAUGERAS et R. DERICHE. *Traitement des occlusions pour la modification d'objet plan dans une séquence d'image*. Dans actes du congrès francophone de Vision par Ordinateur, ORASIS, pages 93–103, Clermont-Ferrand, France, mai 1996.