

SPIM

Thèse de Doctorat



école doctorale **sciences pour l'ingénieur et microtechniques**
UNIVERSITÉ DE BOURGOGNE

Thèse présentée par

Bastien BILLIOT

Pour obtenir le titre de

Docteur de l'Université de Bourgogne

Spécialité: **Instrumentation et Information de l'Image**

Conception d'un dispositif d'acquisition d'images agronomiques 3D en extérieur et développement des traitements associés pour la détection et la reconnaissance de plantes et de maladies

Thèse soutenue le 20 novembre 2013 devant le jury composé de

Frédéric Morain-Nicolier	Professeur - Université de Reims	Rapporteur
David Rousseau	Professeur - Université de Lyon	Rapporteur
Frédéric Baret	Directeur de Recherches - INRA Avignon	Examineur
Mustapha Kardouchi	Professeur - Université de Moncton	Examineur
Gabriel Bloch	Directeur adjoint - Nicéphore Cité	Examineur
Pierre Gouton	Professeur - Université de Bourgogne	Directeur
Frédéric Cointault	MCF-HDR - Agrosup Dijon	Co-Directeur



Résumé

Dans le cadre de l'acquisition de l'information de profondeur de scènes texturées, un processus d'estimation de la profondeur basé sur la méthode de reconstruction 3D « Shape from Focus » est présenté dans ce manuscrit. Les deux étapes fondamentales de cette approche sont l'acquisition de la séquence d'images de la scène par sectionnement optique et l'évaluation de la netteté locale pour chaque pixel des images acquises. Deux systèmes d'acquisition de cette séquence d'images sont présentés ainsi que les traitements permettant d'exploiter celle-ci pour la suite du processus d'estimation de la profondeur. L'étape d'évaluation de la netteté des pixels passe par la comparaison des différents opérateurs de mesure de netteté. En plus des opérateurs usuels, deux nouveaux opérateurs basés sur les descripteurs généralisés de Fourier sont proposés. Une méthode nouvelle et originale de comparaison est développée et permet une analyse approfondie de la robustesse à différents paramètres des divers opérateurs. Afin de proposer une automatisation du processus de reconstruction, deux méthodes d'évaluation automatique de la netteté sont détaillées. Finalement, le processus complet de reconstruction est appliqué à des scènes agronomiques, mais également à une problématique du domaine de l'analyse de défaillances de circuits intégrés afin d'élargir les domaines d'utilisation.

Mots clés : Reconstruction 3D ; Shape from Focus ; Mesure de netteté ; Descripteurs généralisés de Fourier ; Evaluation de robustesse.

Abstract

In the context of the acquisition of depth information for textured scenes, a depth estimation process based on a 3D reconstruction method called "shape from focus" is proposed in this thesis. The two crucial steps of this approach are the image sequence acquisition of the scene by optical sectioning and the local sharpness evaluation for each pixel of the acquired images. Two acquisition systems have been developed and are presented as well as different image processing techniques that enable the image exploitation for the depth estimation process. The pixel sharpness evaluation requires comparison of different focus measure operators in order to determine the most appropriate ones. In addition to the usual focus measure operators, two news operators based on generalized Fourier descriptors are presented. A new and original comparison method is developed and provides a further analysis of the robustness to various parameters of the focus measure operators. In order to provide an automatic version of the reconstruction process, two automatic sharpness evaluation methods are detailed. Finally, the whole reconstruction process is applied to agronomic scenes, but also to a problematic in failure analysis domain aiming to expand to other applications.

Keywords : 3D reconstruction ; Shape from Focus ; Focus measure ; Generalized Fourier descriptors ; Robustness evaluation.

Remerciements

Bien que mener un doctorat demande beaucoup d'autonomie et d'investissement, son accomplissement est, lui, dû à de multiples rencontres, échanges et collaborations. C'est pourquoi je tiens à remercier dans cette partie les personnes sans qui ces travaux n'auraient pas été tels qu'ils sont aujourd'hui.

Tout d'abord, un grand merci à Pierre Gouton et Frédéric Cointault, les directeurs de ces travaux. Merci à toi, Pierre, pour la confiance que tu m'accordes maintenant depuis plusieurs années et bien avant ce doctorat mais également pour tes qualités humaines et ton naturel. Merci Fred pour ton soutien, ta disponibilité et tes nombreux conseils tout au long de ce projet. Nos nombreux échanges tant professionnels qu'amicaux ont joué beaucoup dans le plaisir que j'ai eu à mener ces recherches durant trois ans et probablement dans nos futurs projets communs.

Je tiens à exprimer ma gratitude à David Rousseau et Frédéric Morain-Nicolier d'avoir accepté d'être les rapporteurs de cette thèse. Leurs conseils et questions m'ont permis de prendre du recul sur ce projet et d'en améliorer le contenu. Merci à Frédéric Baret d'avoir accepté la présidence du jury de soutenance et d'avoir examiné mes travaux, tout comme Gabriel Bloch et Mustapha Kardouchi.

Merci beaucoup à mon cher ami expatrié en Thaïlande, Romuald, pour toutes nos discussions plus orientées gastronomie que recherche la plupart du temps, mais surtout pour tes conseils et remarques sur mes travaux. Ton rôle de rapporteur et correcteur officieux m'a grandement aidé à améliorer ce manuscrit.

Je remercie vivement toutes les personnes qui ont contribué de près ou de loin à ce doctorat. Je pense tout particulièrement à Samuel, notamment pour l'application de mes travaux à ton domaine de recherche, Rachid et Christine pour votre aide et vos conseils en statistique, Maxime pour tes corrections, Jean-Claude pour mes débuts dans la peau d'un enseignant, Alex pour être toujours disponible pour résoudre un problème informatique, Claude-Vivien pour ton aide, Richard pour la partie conception et Patricia pour ton soutien dans toute cette complexité administrative.

Merci également à tous mes collègues avec qui j'ai eu la chance de passer ces trois années que ce soit à Agrosup ou à Nicéphore Cité. Avec un merci particulier à Bilal, Sofija, Houda, Simeng, Marion, Lucile, Ludo, Fred Massart, Benjamin, Isabelle, Vanessa, Jean-Claude et toutes les autres personnes que je n'ai pas citées mais qui se reconnaîtront.

Un remerciement particulier pour Tom, ami et compagnon d'infortune lors de ces 3 années, qui a énormément contribué à rendre toutes ces journées géniales. Cependant, je te remercie moins pour tous ces tests et retests de restaurants qui ont entraîné une approche du quintal très préoccupante :-)
même si j'avoue avoir un peu participé aussi à toutes ces tentations. . .

Je remercie affectueusement l'ensemble de mes proches sans qui tout ceci ne serait jamais arrivé. Merci donc à toute ma famille et mes amis d'être eux, tout simplement, et de m'avoir toujours soutenu.

Enfin, comment ne pas remercier ma femme, Laurence, pour les années passées et pour toutes celles à venir. Merci pour l'attention portée à mon travail, ton amour et ton soutien sans faille dans tout ce que j'entreprends.

Table des matières

Résumé	iii
Abstract	v
Remerciements	vii
Table des matières	ix
Table des figures	xiii
Liste des tableaux	xix
Introduction	1
Contexte et problématiques	1
Objectifs	5
Contributions & structure de la thèse	5
1 État de l'art	7
1.1 Généralités	7
1.2 Techniques non optiques	10
1.2.1 Avec contact	10
1.2.1.1 CMMs	10

TABLE DES MATIÈRES

1.2.1.2	Découpe	11
1.2.2	Sans contact	12
1.2.2.1	CT-Scan	12
1.2.2.2	IRM	12
1.2.2.3	SONAR / RADAR	13
1.3	Techniques optiques	13
1.3.1	Vision active	13
1.3.1.1	Triangulation active	14
1.3.1.2	Télémétrie	19
1.3.2	Vision passive	20
1.3.2.1	Triangulation passive	20
1.3.2.2	Shape from Texture	21
1.3.2.3	Shape from Shading	22
1.3.2.4	Shape from Motion	23
1.3.2.5	Shape from Silhouette	24
1.3.2.6	Shape from Defocus	24
1.3.2.7	Shape from Focus	26
1.3.2.8	Confocal stereo	28
1.3.2.9	Systèmes plénoptiques	29
1.4	Conclusion	30
2	Systèmes d'acquisition	33
2.1	Généralités optiques	33
2.1.1	Formation géométrique des images	34
2.1.2	Profondeur de champ	38
2.1.3	Acquisition d'une séquence d'images	43

2.2	Système expérimental	44
2.2.1	Vue d'ensemble	44
2.2.2	Partie optique	45
2.2.3	Partie électronique	47
2.2.4	Contraintes du système expérimental	51
2.3	Système final	52
2.4	Calibrage	54
2.4.1	Correction des aberrations	54
2.4.2	Correction du grossissement	57
2.4.3	Étalonnage de l'objectif motorisé	61
2.5	Procédure d'acquisition	63
2.6	Conclusion	64
3	Evaluation de la netteté	67
3.1	État de l'art des opérateurs de mesure de netteté	69
3.1.1	Mesures différentielles	70
3.1.2	Mesures statistiques	72
3.1.3	Mesures fréquentielles	73
3.1.4	Autres types de mesures	74
3.2	Mesure de netteté basée sur les descripteurs généralisés de Fourier	75
3.3	Approximation	79
3.4	Comparaison	80
3.4.1	Méthodes de comparaison usuelles	80
3.4.2	Méthode de comparaison adaptée	82
3.4.3	Analyse du comportement des opérateurs	87
3.5	Evaluation automatique	92
3.5.1	Méthode non supervisée par fusion d'opérateurs	92

TABLE DES MATIÈRES

3.5.2	Méthode supervisée par analyse de variance	95
3.6	Conclusion	101
4	Applications	103
4.1	Application au dénombrement des épis de blé	106
4.1.1	Acquisitions par le système expérimental	107
4.1.2	Acquisitions par le système final	108
4.2	Application à l'étude de croissance de plante	114
4.3	Application à l'analyse de défaillances de circuits intégrés . . .	120
4.4	Conclusion	125
	Conclusion générale et perspectives	127
	Conclusion générale	127
	Perspectives	131
	Publications	135
	Annexe	137
	Bibliographie	139

Table des figures

0.0.1	Fonctions essentielles de la gestion optimisée de parcelles.	2
1.1.1	Classification des méthodes de reconstruction 3D.	9
1.2.1	Machine de mesure de coordonnées [AER, 2013].	11
1.3.1	Triangulation active par projection de point laser.	14
1.3.2	Triangulation active par projection de ligne laser.	15
1.3.3	Motifs de projection basés sur le code binaire [Posdamer et Altschuler, 1982].	16
1.3.4	Motifs codés spatialement. a. De Bruijn [Pajes et Forest, 2004]. b. Codage M-arrays [Vuylsteke et Oosterlinck, 1990]. c. Codage couleur [Pribanić et al., 2009]. d. Codage non formel [Maruyama et Abe, 1993].	18
1.3.5	Motifs codage direct. a. Codage direct niveau de gris [Carrihill et Hummel, 1985]. b. Codage direct couleur [Tajima et Iwakawa, 1990].	19
1.3.6	Shape from Texture. a. Image originale. b. Champ de normales. c. Surface reconstruite à partir du champ de normales. d. Carte de profondeur. [Clerc et Mallat, 2002].	22
1.3.7	Principe d'acquisition « Shape from Motion ».	23
1.3.8	Schéma représentatif du « Visual Hull ».	25
1.3.9	Exemple de reconstruction 3D par « Shape from Defocus ». [Favaro et Soatto, 2002]	26

TABLE DES FIGURES

1.3.10 Principe du « Shape from Focus ».	27
1.3.11 Exemple de systèmes plénoptiques. a. Caméra Raytrix. b. Appareil Lytro.	30
2.1.1 Schéma d'un appareil sténopé.	34
2.1.2 Loi de Snell-Descartes.	35
2.1.3 Schéma de formation d'une image par une lentille mince convergente.	36
2.1.4 Schéma illustrant la profondeur de champ.	39
2.1.5 Exemples d'ouvertures relatives.	41
2.2.1 Schématisation du procédé d'acquisition.	45
2.2.2 Ensemble constituant la partie optique.	46
2.2.3 Couple moteur nécessaire au module de guidage en fonction de la charge axiale entraînée.	48
2.2.4 Vue d'ensemble du système.	50
2.3.1 Objectif et caméra utilisés.	53
2.4.1 Mire de calibrage et projection des positions utilisées.	56
2.4.2 Distorsions radiales des objectifs utilisés.	57
2.4.3 Grossissement relatif en fonction de la focale et de la distance de mise au point [Willson et Shafer, 1991].	59
2.4.4 Schématisation du grossissement entre la 1 ^{ère} et la $n^{i\text{ème}}$ image.	61
2.4.5 Distance de mise au point en fonction de la position du moteur.	62
2.4.6 Mesure de netteté pour la 1 ^{ère} position moteur.	63
2.4.7 Profondeur de champ en fonction de la distance de mise au point.	64
3.0.1 Schéma du principe de mesure de netteté d'un pixel de la séquence d'images.	69

3.2.1	Schématisation de l'acquisition du descripteur généralisé de Fourier.	76
3.2.2	Obtention du vecteur des scores factoriels.	78
3.3.1	Approximation gaussienne.	79
3.4.1	Effet des niveaux de bruit proposés par [Malik et Choi, 2008] sur la valeur de corrélation entre l'image originale et l'image bruitée.	83
3.4.2	Effet de niveaux de bruit induisant une évolution constante de la valeur de corrélation entre l'image originale et l'image bruitée.	84
3.4.3	Exemple de cartes de profondeur obtenues : a. Voisinage 64×64 sans bruit b. Voisinage 64×64 et bruit de 10%.	85
3.4.4	Evolution de l'erreur en fonction des doublets de paramètres (voisinage 64×64 et 4×4 ; bruit 0% et 20%).	88
3.4.5	Evolution de l'erreur en fonction des doublets de paramètres (voisinage 64×64 et 8×8 ; bruit 0% et 10%).	89
3.5.1	Processus de fusion d'opérateurs de mesure de netteté.	93
3.5.2	Courbe de mesure de netteté correspondant à la projection des scores factoriels.	94
3.5.3	Courbe d'erreur moyenne minimale en fonction des doublets de paramètres.	99
4.0.1	Schéma fonctionnel du processus complet d'estimation de la profondeur.	104
4.1.1	Extrait d'images de la séquence issue du système expérimental.	108
4.1.2	Reconstruction 3D par le système expérimental. a. Carte de profondeur. b. Image nette. c. Projection en fausses couleurs. d. Projection avec placage de texture.	109

TABLE DES FIGURES

4.1.3	Extrait d'images de la séquence d'épis de blé issue du système final.	111
4.1.4	Cartes de profondeur et projection 3D d'une scène d'épis de blé pour un voisinage de 9×9 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$	112
4.1.5	Cartes de profondeur et projection 3D d'une scène d'épis de blé pour un voisinage de 33×33 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$	113
4.1.6	Superposition de l'image nette à la carte de profondeur. a. FM_{tenvar} et voisinage 9×9 . b. FM_{DGFsom} et voisinage 9×9 . c. $FM_{entropy}$ et voisinage 9×9 . d. FM_{tenvar} et voisinage 33×33 . e. FM_{DGFsom} et voisinage 33×33 . f. $FM_{entropy}$ et voisinage 33×33	115
4.2.1	Extrait d'images de la séquence de la plante Millepertuis issue du système final.	116
4.2.2	Cartes de profondeur et projection 3D scène Millepertuis pour un voisinage de 9×9 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$	117
4.2.3	Cartes de profondeur et projection 3D scène Millepertuis pour un voisinage de 33×33 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$	118
4.2.4	Superposition de l'image nette à la carte de profondeur. a. FM_{tenvar} et voisinage 9×9 . b. FM_{DGFsom} et voisinage 9×9 . c. $FM_{entropy}$ et voisinage 9×9 . d. FM_{tenvar} et voisinage 33×33 . e. FM_{DGFsom} et voisinage 33×33 . f. $FM_{entropy}$ et voisinage 33×33	119
4.3.1	Extrait d'images de la séquence du microcontrôleur.	122

4.3.2	Images de profondeur de champ étendue d'une partie de microcontrôleur. a. FM_{tenvar} et voisinage 9×9 . b. $FM_{DGF_{som}}$ et voisinage 9×9 . c. $FM_{entropy}$ et voisinage 9×9 . d. FM_{tenvar} et voisinage 33×33 . e. $FM_{DGF_{som}}$ et voisinage 33×33 . f. $FM_{entropy}$ et voisinage 33×33	123
4.4.1	Base de texture VisTex	137

Liste des tableaux

2.1	Champ de vue (Largeur et Hauteur) et profondeur de champ (PdC) en fonction du type d'objectif (valeurs en mm).	43
3.1	Matrice d'expériences, effets et interaction pour 2 niveaux de paramètres.	86
3.2	Synthèse des résultats des plans d'expériences (voisinage 64×64 et 4×4 ; bruit 0% et 20%).	88
3.3	Synthèse des résultats des plans d'expériences (voisinage 64×64 et 8×8 ; bruit 0% et 10%).	89
3.4	Synthèse des groupes issus de l'analyse de variance. Du meilleur groupe au moins bon : Bleu, Vert, Rouge, Orange, Bleu clair, Noir, Blanc.	97
3.5	Synthèse du groupe d'opérateurs les plus efficaces issus de l'analyse de variance.	98
3.6	Groupes de doublets de comportement similaire. Les couleurs identiques symbolisent les mêmes comportements. Le blanc est utilisé pour les doublets n'appartenant à aucun groupe. . .	100
3.7	Classement général des opérateurs de mesure de netteté (du meilleur groupe A au moins bon F).	101

Introduction

Contexte et problématiques

Initialement, la gestion des cultures était effectuée de manière globale, la parcelle étant considérée homogène. Cependant, si cela était valable à une époque où les parcelles agricoles étaient relativement petites et cultivées en fonction des caractéristiques et de la richesse des sols, la taille toujours plus importante des parcelles et les pressions environnementales ont entraîné une remise en question de ce postulat. En effet, l'existence d'une variabilité intra-parcellaire a permis la mise en évidence de la nature hétérogène des parcelles. Dès lors, la gestion des parcelles nécessite la connaissance et la caractérisation de celles-ci afin de tendre vers une culture optimale. Le besoin d'informations nécessaire à une bonne caractérisation des parcelles a fait émerger le concept d'agriculture de précision dans les années 90. L'objectif de cette nouvelle vision de l'agriculture est une optimisation de la gestion d'une parcelle d'un point de vue agronomique, environnemental et économique dans le but de fournir une gestion sur mesure de chaque parcelle. Ceci permet, par exemple, de minimiser les surdosages d'intrants (l'ensemble des produits n'étant pas présent dans le sol naturellement comme les produits phytosanitaires et les engrais) entraînant un gain économique (moins de produits utilisés), écologique (moins de pertes par le sol) et sanitaire (moins de contact entre l'opérateur et les produits). Une expression couramment utilisée et représentant très bien la philosophie sous-jacente de l'agriculture de précision est : « *la bonne intervention au bon endroit et au bon moment* » [Zwaenepoel et al., 1997]. Bien entendu, beaucoup d'autres objectifs sont apparus au cours du temps avec l'évolution des

moyens technologiques comme le suivi de l'évolution des cultures, l'évaluation du rendement ou encore la détection de maladies et des infestations.

Afin de définir le concept de « précision », [McBratney, 2000] propose un diagramme fonctionnel composé de 5 éléments essentiels (figure 0.0.1). Ses travaux portent sur la viticulture de précision qui fait cependant partie intégrante de l'agriculture de précision au sens large.

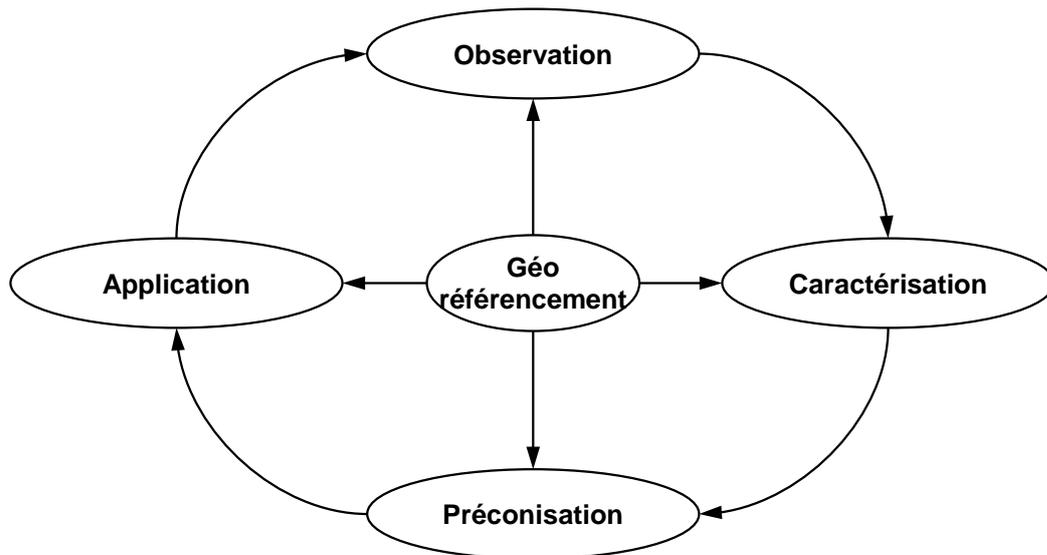


Figure 0.0.1: Fonctions essentielles de la gestion optimisée de parcelles.

Le géoréférencement constitue l'élément central permettant de lier l'ensemble des autres fonctions à une même localisation. La partie observation correspond à l'étape d'acquisition de données alimentant par la suite la fonction de caractérisation. Cette dernière permet de transformer ces données en informations exploitables par un expert du domaine agronomique. Ces informations induisent ensuite des préconisations et conseils quant à la gestion de la parcelle analysée. Tenant compte de ces conseils, l'opérateur peut adapter sa gestion de parcelle comme, par exemple, appliquer un traitement propre à une zone spécifique.

Un des outils s'étant rapidement imposé comme adapté aux problématiques de l'agriculture de précision est la vision artificielle. Egalement appelée vision par ordinateur, cette notion comprend l'ensemble des étapes permettant une

interprétation des images d'un objet ou d'une scène analysée. Cette étape débute par l'acquisition de la scène par un capteur, son optique ainsi que l'illuminant adapté. Puis vient le traitement des images éliminant l'information inutile et mettant en évidence les données recherchées. Ces données sont extraites des images par un processus d'analyse qui permet d'alimenter la phase décisionnelle transformant ces données en informations utiles.

Naturellement, on retrouve une analogie entre ces processus de vision par ordinateur et les étapes d'observation et caractérisation en agriculture de précision. Ainsi, ces étapes, autrefois réalisées visuellement et manuellement, sont automatisées par le processus adapté en vision artificielle. La rapidité, la répétabilité ainsi que l'objectivité liées à l'évaluation automatique de l'information s'opposent aux aspects fastidieux et subjectifs de l'évaluation manuelle.

Parmi les procédés d'acquisition retenus en agriculture de précision, on distingue principalement deux types de vecteurs d'acquisition : la télédétection et la proxidétection. Ces deux modes d'acquisition sont complémentaires en agriculture de précision car les données acquises par leur biais sont souvent différentes et les informations à obtenir également. Les systèmes de télédétection tels que les satellites, avions ou, plus récemment, drones permettent l'obtention de données à une échelle plus ou moins éloignée de la parcelle. Ainsi, les informations extraites caractérisent la parcelle dans son ensemble pour réaliser des cartographies. Ces dernières mettent en évidence la variabilité existante au sein d'un champ comme, par exemple, les variations de types de sols, l'état du couvert végétal, l'humidité des sols, le stress des plantes cultivées ou encore l'identification du type de culture. Les systèmes de proxidétection, quant à eux, permettent une acquisition de données au plus près de la plante afin de recueillir des données plus fines que celles obtenues par télédétection. Les avantages de ces systèmes sont multiples : meilleure résolution d'images, meilleure temporalité (acquisition à la demande à n'importe quel moment) mais accès simplifié et coût moindre. De tels systèmes s'orientent principalement vers la détection de maladies, le suivi de croissance et l'étude de la structure du couvert.

Depuis quelques années, l'agriculture de précision a connu une forte évolution de ses procédés d'acquisition de données. Là où précédemment les ac-

quisitions étaient réalisées essentiellement sur le terrain, la démocratisation de la culture en environnement contrôlé a entraîné la conception de nouveaux types d'outils de proxidtection : les plateformes de phénotypage. L'objectif principal de ces plateformes est la caractérisation des phénotypes des plantes, ceci afin d'établir le lien entre génotype et phénotype. Le génotype d'un organisme représente son patrimoine génétique déterminant l'ensemble de ses caractéristiques. Le phénotype, quant à lui, correspond uniquement aux caractéristiques observables de cet organisme. L'intérêt de comprendre le lien entre ces deux éléments réside dans le fait que même si le génotype participe en grande partie au phénotype d'un organisme, l'environnement dans lequel celui-ci se développe contribue également à modifier le phénotype. Ainsi, l'étude de ces interactions permet de meilleures compréhensions et caractérisations de la croissance des plantes afin de sélectionner des variétés plus robustes aux différents stress en vue d'une agriculture compétitive et respectueuse de l'environnement.

C'est dans le but d'atteindre ces objectifs que le projet baptisé « PHE-NOME » a vu le jour. Sa vocation est de fédérer l'ensemble des plateformes de phénotypage à l'échelle nationale avec des collaborations au niveau européen. Parmi les différents types de plateforme du projet, on retrouve des plateformes au champ équipées de capteurs et systèmes d'imagerie dédiés à l'observation et la caractérisation des cultures. On retrouve également un véhicule autonome doté de multiples capteurs baptisé « phénomobile » dont la mission est l'analyse complète de microparcelles. Enfin, deux plateformes de phénotypage à haut débit, dont l'objectif est l'analyse de plusieurs centaines de plantes par jour, ont été créées. Au cœur de ces plateformes se trouve un ensemble de systèmes d'acquisition allant du visible à l'infrarouge en passant par l'imagerie de fluorescence.

Cependant, bien que des systèmes classiques d'acquisition en deux dimensions permettent l'obtention de données précises et variées, les limites inhérentes à ce type de système sont perçues. Ces limites sont, entre autres, le manque d'information de profondeur en tout point de la scène analysée. Cette information est nécessaire dans plusieurs applications telles que l'estimation de la croissance des plantes et l'étude de la disposition spatiale des éléments

composant ces plantes (feuilles, tige, fleurs, épis ...). Pour cela, l'utilisation de systèmes capables d'évaluer cette profondeur est nécessaire.

Objectifs

Les objectifs de ces travaux de recherche sont donc le développement d'un système d'acquisition et les traitements associés permettant l'évaluation de la profondeur en tout point de la scène considérée. Si les systèmes d'acquisition 3D sont relativement ancrés actuellement dans d'autres domaines, leur utilisation est récente en agriculture de précision. Les spécificités des scènes agronomiques au sens large sont leur grande diversité et leur aspect complexe du fait de leurs discontinuités de relief contrairement à un objet de surface continue. De plus, le système d'acquisition n'étant pas destiné à une utilisation par des spécialistes du domaine de la vision artificielle, il doit faire l'objet d'une utilisation intuitive et d'une automatisation maximum.

Contributions & structure de la thèse

Ce manuscrit est organisé de façon à refléter la démarche ayant permis d'aboutir au système fonctionnel existant. Ainsi, le premier chapitre est consacré à dresser l'état de l'art des techniques de reconstruction 3D. Un état de l'art exhaustif de ces méthodes n'étant évidemment pas envisageable, ce chapitre présente néanmoins un ensemble relativement large des techniques existantes avec, lorsque cela est possible, un exemple de l'utilisation de la méthode à une problématique agronomique. Les avantages et limitations de chacune des méthodes présentées sont exposés en justifiant le choix de la méthode retenue.

Le second chapitre est dédié à la description des deux systèmes conçus dont le rôle est l'acquisition de la séquence d'images particulière requise pour la mise en œuvre de la méthode de reconstruction 3D développée. La méthode choisie s'articulant autour d'une maîtrise parfaite de certains paramètres optiques, une description de ces notions est réalisée. Les images acquises n'étant pas originellement exemptes de défauts, un calibrage des systèmes d'acquisition

et une correction des images sont nécessaires et détaillés ainsi que la procédure d'acquisition à respecter.

Le troisième chapitre a pour rôle de présenter un aspect crucial de la méthode d'estimation de la profondeur qui repose sur l'évaluation de la netteté de chaque pixel. Un rapide état de l'art des outils permettant l'estimation de la netteté est présenté ainsi que deux nouveaux opérateurs de mesure de netteté développés et basés sur les descripteurs généralisés de Fourier. L'étude du comportement de chaque opérateur de mesure de netteté présenté est réalisée au travers d'une analyse adaptée à notre problématique faisant appel à un plan d'expériences factoriel. Puisqu'un des objectifs est l'automatisation de l'estimation de l'information de profondeur, deux approches d'évaluation automatique de la netteté sont détaillées. De plus, cette analyse du comportement des opérateurs et leur comparaison sont nécessaires afin de choisir l'opérateur adéquat en fonction de la diversité des scènes étudiées et permet de pallier au manque de vérité terrain de l'information de profondeur.

Le quatrième et dernier chapitre est, quant à lui, consacré à la présentation de résultats de reconstruction 3D de scènes agronomiques. De plus, afin d'illustrer la portée des travaux réalisés au cours de ce doctorat, une application au domaine de l'analyse de défaillance de circuits électroniques est présentée.

Enfin, une conclusion générale est exposée ainsi que les perspectives liées à ces travaux de recherche.

Chapitre 1

État de l'art

1.1 Généralités

Depuis quelques années, le terme « 3D » est devenu quasi-omniprésent dans un certain nombre de domaines. C'est le cas pour les téléviseurs, consoles de jeux, imprimantes, cinéma, appareils photo ... etc. On remarque que, en dehors des imprimantes 3D initialement utilisées pour du prototypage rapide et de plus en plus accessibles au grand public, le terme « 3D » indique bien souvent un mode de perception tridimensionnel du média considéré. La volonté d'obtention de l'information tridimensionnelle à l'image de la vision humaine n'est pas récente en vision par ordinateur. On peut citer les travaux de [Marr et Poggio, 1977] sur la théorisation de la vision humaine et une implémentation quatre ans plus tard par [Grimson, 1981].

Ainsi, depuis plus de trente ans, obtenir l'information tridimensionnelle est un axe de recherche vaste qui est retrouvé au travers de différents termes tels que acquisition 3D, reconstruction 3D ou encore numérisation 3D. Aujourd'hui, le besoin de recourir à ces méthodes est présent dans beaucoup de domaines comme la robotique, le contrôle industriel, l'imagerie médicale, la reconstruction environnementale ou encore, l'agronomie au travers notamment de l'agriculture de précision. Bien souvent, ce besoin se manifeste pour pallier au manque d'information induit par la 2D grâce à une interprétation complète de la scène.

Le choix de la méthode de reconstruction 3D peut être imposé par le type de représentation souhaité. On distingue deux grandes familles de représentation 3D : les représentations surfaciques et les représentations volumiques. Pour les premières, on retrouve :

- les cartes de profondeurs : c'est une image en niveaux de gris ou en fausses couleurs où la valeur de chaque pixel correspond à une distance entre le point correspondant dans la scène et le système d'acquisition ;
- les surfels : abréviation de « surface element », ils sont composés d'attributs décrivant un échantillon ponctuel de la surface comme ses coordonnées, sa texture, la normale ... etc ;
- le maillage : forme très répandue de représentation 3D, c'est un ensemble de points définis par leurs coordonnées 3D reliés les uns aux autres par des arêtes.

Les représentations volumiques regroupent :

- les voxels : pour « volumetric element », dénomination d'un pixel tridimensionnel constitué d'une information colorimétrique associée à des coordonnées spatiales ;
- les harmoniques sphériques : ceci permet une représentation fréquentielle du modèle 3D en coordonnées sphériques.

Le choix de la méthode de reconstruction est aussi fonction de contraintes liées à l'application comme :

- la profondeur mesurable, c'est-à-dire la distance minimum en Z (en considérant la scène comme un espace X , Y et Z) mesurable entre deux points : ceci correspond à la précision du système de reconstruction ;
- la taille du système si des contraintes d'encombrement sont à prendre en compte ;
- la vitesse d'acquisition si la scène est statique ou non ;
- le champ de vue possible du système d'acquisition qui correspond à la dimension de la scène observable.

La figure 1.1.1 n'a pas l'ambition de proposer une liste exhaustive des différentes techniques de reconstruction 3D mais fournit néanmoins une vue d'ensemble des possibilités offertes. Le classement de ces techniques peut se faire de plusieurs façons différentes. Dans le cadre de ce document, la classifica-

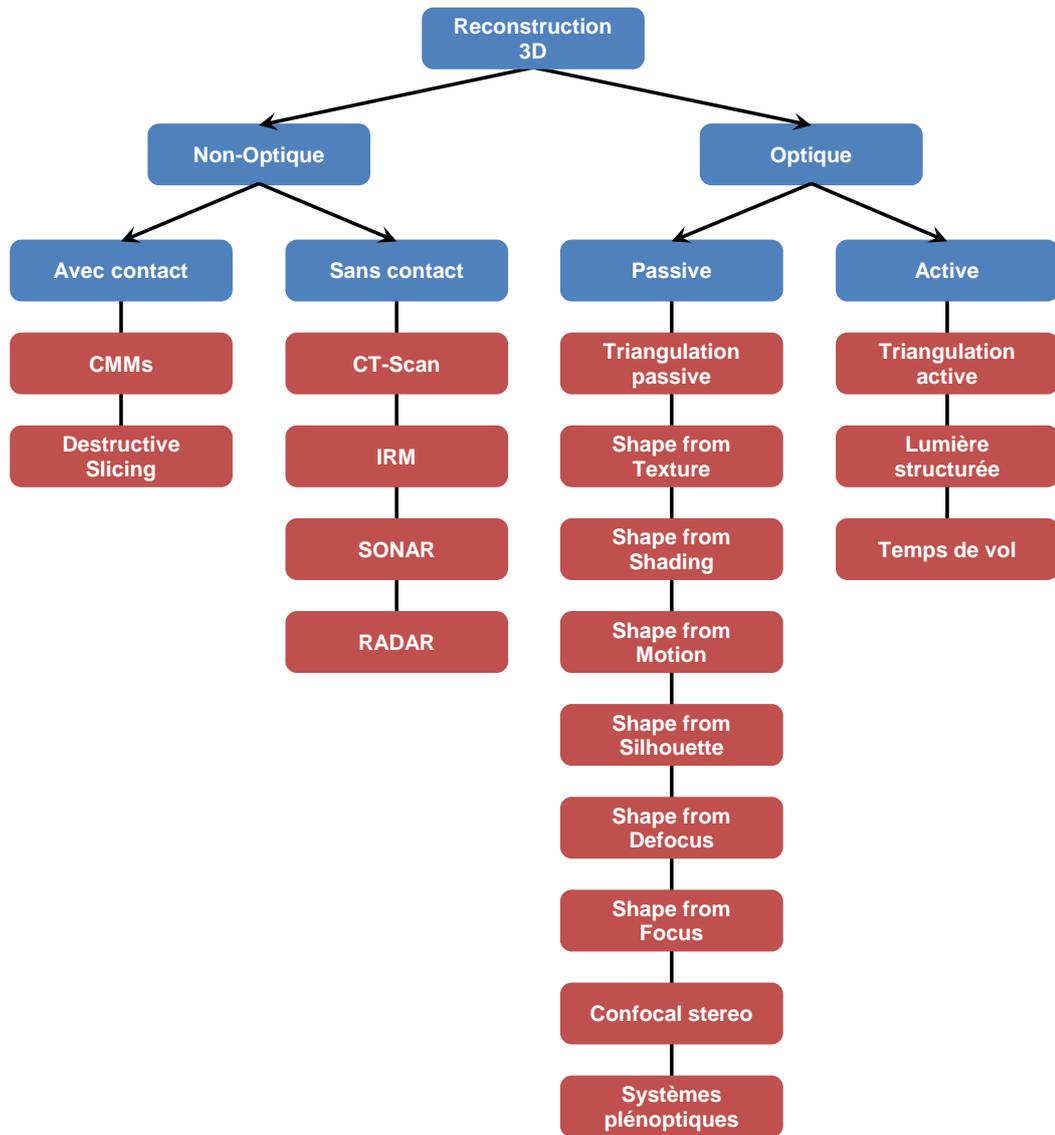


Figure 1.1.1: Classification des méthodes de reconstruction 3D.

tion proposée est basée sur la méthode d'acquisition. Ainsi on distingue les méthodes optiques des méthodes non optiques, avec ou sans contact, qui font l'objet de deux sous-parties de ce chapitre.

Les techniques optiques sont elles-mêmes divisées en deux catégories appelées vision active et vision passive. La différence principale entre ces deux catégories est l'adjonction ou non d'un illuminant particulier au capteur optique. Le principe de fonctionnement de chaque méthode est présenté ainsi que leurs avantages et inconvénients. Lorsqu'une utilisation pour une application agronomique d'une méthode existe, cela est mentionné. Les justifications concernant le choix de la méthode de reconstruction retenue sont présentées en fin de chapitre.

1.2 Techniques non optiques

1.2.1 Avec contact

1.2.1.1 CMMs

Acronyme de « Coordinate Measuring Machines », autrement dit machine de mesure de coordonnées [Seugling, 2000], une CMM est un appareil de mesure muni d'une sonde mobile de type palpeur (figure 1.2.1). La position de la sonde est suivie par une série de codeurs sur chaque axe. La numérisation est de type non destructive avec contact. Ces systèmes sont très précis (dixième de μm) avec une bonne répétabilité mais très lents. Ils sont plutôt utilisés dans le milieu industriel pour de la rétro-conception ou pour la vérification dimensionnelle de pièces rigides. Le système palpeur peut également être couplé avec des technologies de type scanner laser et intégré dans un système de bras polyarticulé principalement utilisé pour la numérisation précise d'objets immobiles.



Figure 1.2.1: Machine de mesure de coordonnées [AER, 2013].

1.2.1.2 Découpe

Méthode d'acquisition 3D destructive, cette technique consiste à injecter une résine dans l'objet à numériser puis à le découper en fines tranches. Ces tranches seront supprimées après chaque acquisition de l'image correspondante. La distance capteur/objet est contrôlée avec précision de telle sorte que l'échelle de chaque image soit connue. Les différents plans de coupe sont ensuite intégrés dans un logiciel de reconstitution afin d'obtenir le modèle 3D complet de l'objet. L'intérêt de cette méthode est l'obtention d'un modèle volumique intégral de l'objet et pas seulement surfacique. On retrouve l'utilisation de cette méthode dans le cadre de rétro-conception ou processus de vérification de conformité par corrélation entre le modèle acquis et le modèle CAO de la pièce. Bien entendu, les méthodes avec contact ne sont pas recommandées dans notre cas puisque nos scènes sont de type non rigide.

1.2.2 Sans contact

1.2.2.1 CT-Scan

Principalement utilisée dans le milieu médical, la tomodensitométrie [Doyon, 2000] (Computed Tomography), appelée couramment scanner, est de plus en plus utilisée dans d'autres domaines. Cette technique permet une numérisation sans contact et non destructive de l'objet. Le principe de base consiste à envoyer des rayons X à travers l'objet à scanner, puis à mesurer les rayons X transmis afin de mettre en évidence les différentes zones d'absorptions des rayons par l'objet. Une succession de coupes de l'objet est ainsi acquise permettant une reconstruction volumique de celui-ci. On retrouve plusieurs applications de cette méthode en agronomie comme par exemple la reconstruction 3D de troncs d'arbres, utilisée afin de détecter les nœuds présents dans ceux-ci [Krähenbühl et al., 2012].

1.2.2.2 IRM

L'Imagerie par Résonance Magnétique [Doyon, 2004] se base sur l'utilisation d'un champ magnétique (aimant) et des ondes radio. Contrairement au CT-Scan, aucune radiation ionisante n'est émise. Son principe consiste à réaliser des images en coupe de l'objet étudié ; principalement du corps humain de part sa composition importante en atomes d'hydrogène. Entourés d'un champ magnétique, tous les atomes d'hydrogène s'orientent dans la même direction : ils sont alors excités par des ondes radio durant une très courte période afin de les mettre en résonance. A l'arrêt de cette stimulation, les atomes restituent l'énergie accumulée en produisant un signal qui sera enregistré et traité par les outils associés. Une succession de coupes est ainsi acquise qui, en les superposant et en les recalant, donne le modèle 3D volumique de l'objet étudié. On trouve des utilisations agronomiques de cette méthode comme, par exemple, dans le cadre d'étude de racines de plantes ou de cultures [Haber-Pohlmeier et Pohlmeier, 2010].

Les deux méthodes présentées ici sont très utilisées pour l'étude de scènes organiques mais ont pour principal inconvénient la nécessité d'une installation

encombrante. C'est donc le sujet de l'étude qui doit être placé au sein des scanners, il est donc limité aux scènes pouvant être déplacées.

1.2.2.3 SONAR / RADAR

Acronyme de « SOund Navigation And Ranging » et « RAdio Detecting And Ranging », ces deux techniques utilisent respectivement une onde sonore et une onde radio comme moyen de déterminer l'information de profondeur. Le principe consiste à émettre une onde sur l'objet à numériser et capter l'onde réfléchie. La distance capteur/objet est déterminée par le temps (t_{total}) que met l'onde (de vitesse de propagation v_{onde}) à faire l'aller/retour :

$$d = \frac{t_{total} \times v_{onde}}{2}. \quad (1.2.1)$$

Plus de détails sur ces techniques peuvent être trouvés dans [Le Chevalier, 2002]. L'utilisation de technique RADAR est courante en télédétection, comme par exemple l'étude de la composition des sols et son impact sur la croissance des cultures de blé [Mattia et al., 2003].

1.3 Techniques optiques

1.3.1 Vision active

Le principe général des méthodes de reconstruction 3D utilisant la vision active repose sur l'utilisation conjointe d'une caméra (ou capteur) et d'une source d'énergie lumineuse. L'émission de lumière utilisée peut être de différents types tels que le laser ou la projection d'un motif particulier et va dépendre de l'application visée. Cette catégorie de méthodes de reconstruction est composée de plusieurs principes de base différents qui sont présentés dans les sous-sections suivantes.

1.3.1.1 Triangulation active

Cette catégorie de techniques regroupe tout un ensemble de méthodes basées sur l'utilisation d'une source d'émission et d'un capteur formant un triangle avec l'objet à numériser. Plusieurs types de projection sont possibles dont résultent des techniques de reconstruction différentes. On peut d'abord citer les projections de lumière type laser.

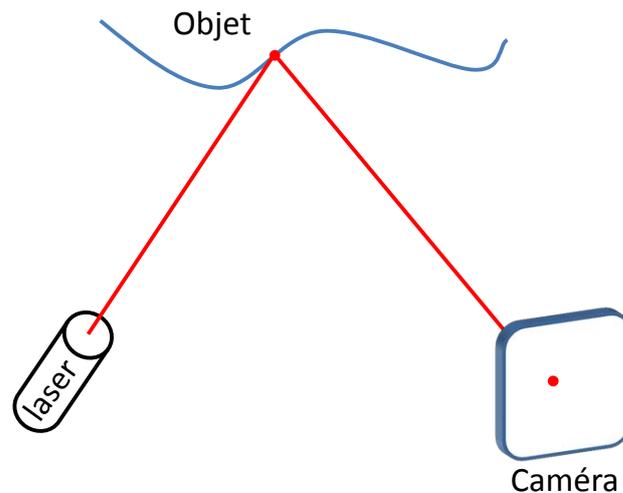


Figure 1.3.1: Triangulation active par projection de point laser.

Un point laser est projeté sur l'objet à numériser et acquis par la caméra (Figure 1.3.1). Le système étant préalablement calibré, l'orientation et l'inclinaison spatiale de la caméra et du laser sont connues. Ceci permet, par triangulation, de déterminer la distance entre la caméra et l'objet au niveau du point projeté. Le principal inconvénient de cette méthode est le nombre très élevé de points à acquérir pour obtenir une reconstruction acceptable de l'objet. Cette méthode est principalement utilisée en remplacement des techniques CMMs.

Afin d'accélérer la procédure d'acquisition, le point laser peut être orienté vers un système de lentilles divergentes (cylindriques) afin de transformer ce point en une ligne laser (Figure 1.3.2).

Contrairement au système point laser, les relations de triangulation ne sont plus utilisées pour déterminer la profondeur mais une étude de la déformation

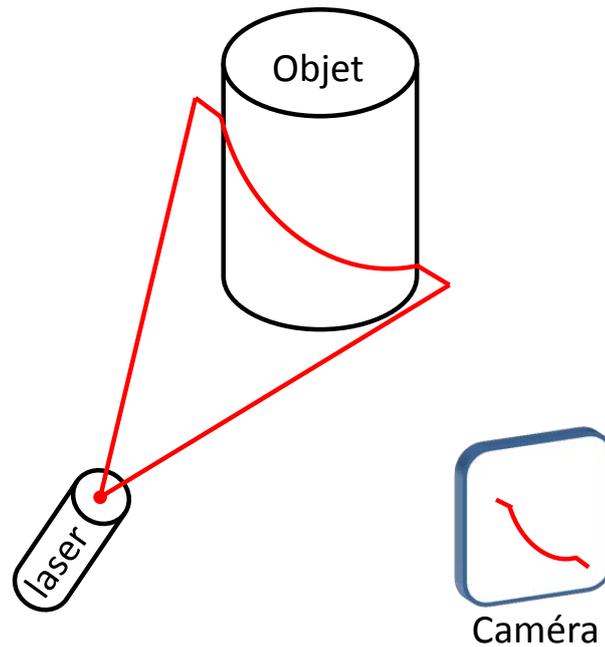


Figure 1.3.2: Triangulation active par projection de ligne laser.

de la ligne par l'objet permet de remonter à cette information. Un balayage de l'objet est nécessaire pour la numérisation complète, d'où le nom « scanner 3D » couramment utilisé pour dénommer cette méthode. On retrouve ce type de système sur des bras polyarticulés ou des appareils de scanning utilisant des miroirs rotatifs pour balayer la scène. L'utilisation principale concerne le domaine industriel pour de la rétro-conception ou du contrôle qualité et, de plus en plus, le domaine du patrimoine pour de la numérisation d'œuvres d'art [Chane et al., 2012]. L'inconvénient de l'utilisation de pointeurs laser est la nécessité du déplacement ainsi que le nombre d'acquisitions importantes à effectuer. Un autre problème survient lorsque la ligne n'est pas entièrement visible par la caméra (phénomène appelé occlusion), ce qui entraîne un problème de mise en correspondance.

Une autre solution, dénommée lumière structurée, est la projection d'un motif particulier 2D sur la scène considérée en lieu et place du laser. Ces méthodes correspondent au deuxième type de projection couramment utilisé en vision active. Un projecteur classique est utilisé pour projeter un motif 2D sur la scène complète. Il existe beaucoup de motifs différents utilisables mais

ceux-ci peuvent être regroupés en trois catégories : le multiplexage temporel, le codage spatial et le codage direct. Plusieurs travaux de recherche sur le regroupement et la comparaison des différents types de motifs ont été publiés depuis plus de dix ans. Parmi les plus populaires, on peut citer [Batlle et al., 1998] et [Salvi et al., 2004] dont une mise à jour récente a été réalisée dans [Salvi et al., 2010]. Une présentation succincte des différentes catégories est faite ici.

Multiplexage temporel

Le principe de cette catégorie repose sur une projection successive de plusieurs motifs sur la scène considérée et l'acquisition de la scène après chaque projection. Un exemple classique de motifs utilisés durant la phase d'acquisition est illustré figure 1.3.3 et a été présenté la première fois par [Posdamer et Altschuler, 1982].

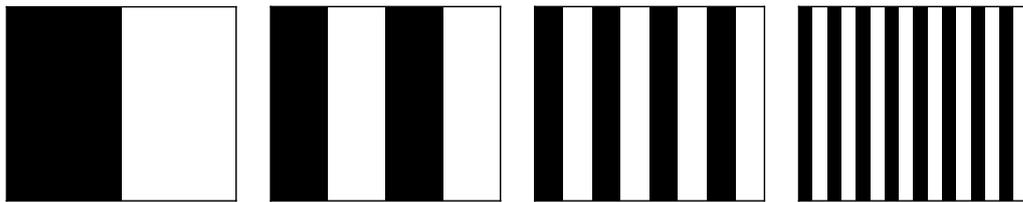


Figure 1.3.3: Motifs de projection basés sur le code binaire [Posdamer et Altschuler, 1982].

Chaque pixel de la scène acquise va donc prendre une valeur dépendante du motif projeté au cours du temps qui constituera un mot binaire. La mise en correspondance sera ensuite effectuée en comparant le code binaire des pixels des images acquises aux codes binaires des images projetées.

Plusieurs améliorations de la projection de code binaire ont été proposées, notamment l'utilisation d'un codage de Gray par [Inokuchi et al., 1984] qui permet une meilleure robustesse au bruit. [Rocchini et al., 2001] proposent quant à eux d'utiliser non pas une mire noire et blanche mais rouge et bleue ainsi qu'un léger espace de couleur verte d'un pixel entre les bandes projetées. Ceci permet de faciliter la localisation de la transition des bandes. Le principal inconvénient induit par l'utilisation de ce type de motif est le grand

nombre d'acquisitions nécessaires à la reconstruction et, par conséquent, seules des scènes statiques sont reconstruites. Pour diminuer ce nombre d'images à projeter, [Caspi et al., 1998] utilisent des motifs couleurs basés sur les codes n-aires pour lesquels les franges sont codées dans l'espace RGB. Ceci permet également un meilleur fonctionnement avec des scènes non neutres colorimétriquement parlant car le code couleur peut être adapté à l'environnement. On peut également citer des méthodes hybrides projetant conjointement un codage temporel et spatial et permettant une acquisition rapide [Ishii et al., 2007].

Codage spatial

Contrairement au multiplexage temporel où les mots de code sont obtenus suite à une projection d'images successives, la projection de motifs codés spatialement permet une obtention des mots à partir de l'information des pixels voisins. Ceci a pour effet premier de permettre la reconstruction 3D de scènes en mouvement du fait de l'acquisition unique du motif nécessaire. Les informations des pixels voisins utilisées peuvent être les variations d'intensité et de couleur pour former le mot codé de chaque pixel. On retrouve pour ces codages plusieurs types de motifs projetés tels que ceux basés sur une séquence de Bruijn [Pajes et Forest, 2004, Salvi et al., 2004], des codes couleurs [Pribanić et al., 2009], des codages non formels [Maruyama et Abe, 1993] ou encore le codage M-arrays [Vuylsteke et Oosterlinck, 1990].

Des exemples de motifs sont illustrés figure 1.3.4. Les principales limitations de ces motifs sont la sensibilité aux occlusions et, pour les motifs couleurs, la difficulté de mise en correspondance sur une scène colorée.

Un exemple d'utilisation de ce type de motif en agronomie est présenté par [Chéné et al., 2012]. L'appareil Kinect (Microsoft, Redmond, Washington), développé par la société Primasense, est utilisé et permet une reconstruction 3D de la scène. Le capteur comprend un système de projection de lumière structurée infrarouge constituée de points répartis semi-aléatoirement en neuf blocs ainsi qu'un capteur infrarouge. La reconstruction 3D des plantes considérées permet une segmentation simple des feuilles en vue de l'identification de pathogènes.

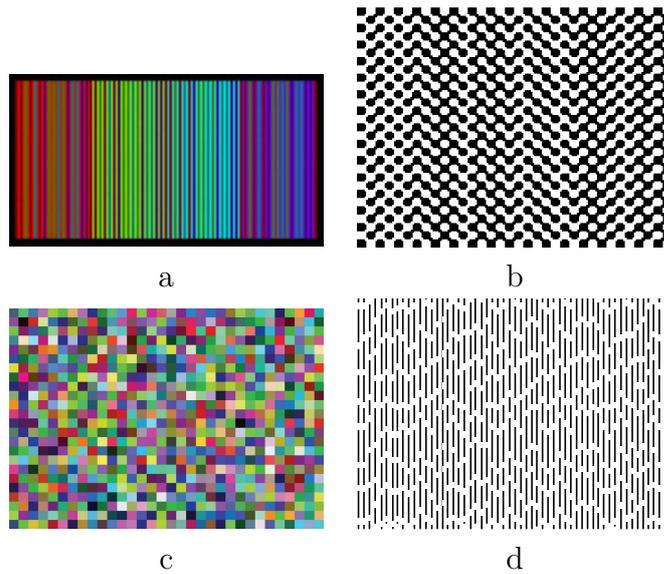


Figure 1.3.4: Motifs codés spatialement. a. De Bruijn [Pajes et Forest, 2004]. b. Codage M-arrays [Vuylsteke et Oosterlinck, 1990]. c. Codage couleur [Pribanić et al., 2009]. d. Codage non formel [Maruyama et Abe, 1993].

Codage direct

Pour ce type de codage, une seule mire est utilisée comme pour un codage spatial mais le codage est réalisé directement au niveau du pixel projeté. Deux types de codage direct sont à différencier : ceux en niveaux de gris et ceux en couleur. [Carrhill et Hummel, 1985] proposent l'utilisation d'une mire représentant un dégradé de niveaux de gris horizontal (figure 1.3.5a). La méthode est améliorée par [Miyasaka et al., 2000] qui obtiennent une reconstruction plus précise en utilisant une caméra tri-CCD en lieu et place d'une caméra couleur. Une mire couleur, semblable à un arc en ciel (figure 1.3.5b) projetée sur la scène, est proposée par [Tajima et Iwakawa, 1990]. Ce type de mire permet un codage d'une plus grande dynamique que celles en niveaux de gris mais est, de fait, plus sensible aux propriétés colorimétriques de la scène considérée. Un inconvénient majeur de ce type de codage est aussi lié à la haute sensibilité au bruit additif due aux différences très faibles entre deux niveaux successifs. Ainsi, un niveau de bruit modifiant la valeur d'un pixel entraîne une estimation erronée de sa profondeur.

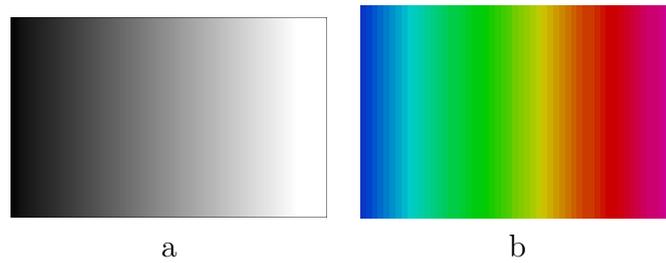


Figure 1.3.5: Motifs codage direct. a. Codage direct niveau de gris [Carrihill et Hummel, 1985]. b. Codage direct couleur [Tajima et Iwakawa, 1990].

Pour conclure, ces techniques, basées sur la projection de lumière structurée, sont considérées comme très efficaces pour une reconstruction 3D de la scène tant que cette dernière est constituée de couleurs plutôt neutres avec un relief peu accentué.

1.3.1.2 Télémétrie

La télémétrie consiste à mesurer la distance séparant un objet et un émetteur/récepteur d'après l'étude de l'onde émise par ce dernier. L'onde peut être non optique, telle que RADAR ou SONAR présentées précédemment, mais également optique, comme c'est le cas dans les systèmes LIDAR (Ligh Detection And Ranging) aussi appelés télémètre laser. La mesure de la distance peut être effectuée de plusieurs façons.

La première, à l'aide d'une caméra dite « temps de vol », consiste à émettre une impulsion de lumière laser qui est réfléchiée par l'objet à scanner. Le résultat de cette réflexion est détecté par un capteur et le temps écoulé entre l'émission et la détection donne la distance objet/capteur (eq. 2.2.1 page 13).

La deuxième technique consiste à moduler en amplitude l'onde laser puis à mesurer le déphasage entre l'onde émise et l'onde reçue. La mesure de distance est déduite de cette différence de phase.

La télémétrie est particulièrement bien adaptée aux reconstructions de grandes structures comme des bâtiments du fait de leur portée élevée (> 300m) et de leur précision de l'ordre du millimètre. Plus de détails sur ces

différentes méthodes peuvent être trouvés dans [Adams, 2002].

Dans le domaine agronomique, on retrouve une utilisation du LIDAR de plus en plus importante, principalement en télédétection et à des fins de caractérisation de la structure de la végétation [Omasa et al., 2007].

1.3.2 Vision passive

Contrairement à la vision active, les méthodes dites « passives » ne nécessitent pas d'illuminant additionnel pour fonctionner. Dans cette catégorie de méthodes de reconstruction 3D, on retrouve également des méthodes par triangulation mais utilisant plusieurs caméras, des méthodes se basant sur l'analyse des images acquises et des méthodes nécessitant une optique particulière.

1.3.2.1 Triangulation passive

Cette catégorie de méthodes de reconstruction 3D comprend une des techniques les plus répandues : la stéréovision aussi appelée « Shape from Stereo ». Son origine est probablement la plus ancienne puisque se basant sur le procédé de perception humaine : l'obtention du relief à partir de deux points de vue légèrement différents. Un des premiers dispositifs d'affichage d'images 3D appelé stéréoscope est d'ailleurs basé sur la stéréovision et a été inventé par le physicien anglais Charles Wheatstone en 1838.

La mesure de la distance entre les caméras et les points de la scène revient à une simple triangulation entre ces trois éléments. Cependant, cette triangulation ne devient possible que lorsque chaque point de la scène est projeté sur chaque image avec une position connue. Les deux images (ou plus) de la scène sont acquises avec des points de vue différents, ce qui a pour effet de créer une disparité géométrique entre elles. La mesure de cette disparité est réalisée en déterminant les correspondances entre chaque image qui représentent la visibilité d'un même point de la scène dans les images acquises. Les différentes étapes composant cette méthode, introduites par [Barnard et Fischler, 1982] et généralisées dans le cadre de la géométrie multivues par [Hartley, 2008],

sont :

1. Calibrage des caméras permettant l'obtention des paramètres intrinsèques et extrinsèques du système multi-caméras.
2. Rectification des images acquises suite au calibrage. Ceci permet de réduire la complexité d'appariement puisque un point dans une image ne pourra se trouver que sur une ligne dans la deuxième image (droite épipolaire).
3. Mise en correspondance des points par une méthode d'appariement.
4. Calcul de la carte de disparité permettant la reconstruction 3D de la scène.

L'étape d'appariement représente la principale difficulté pour le bon fonctionnement de cette technique et dépend du choix de la base entre les caméras ainsi que l'angle d'inclinaison de celles-ci. En effet, plus la base est grande et plus la mesure de distance est précise, plus on est exposé au phénomène d'occlusion (un point de la scène vu par une caméra ne l'est pas nécessairement par l'autre). Des informations détaillées sur le concept de stéréovision peuvent être trouvées dans [Faugeras, 1993].

On trouvera dans [Rovira-Más et al., 2008] un exemple d'utilisation d'une caméra stéréoscopique afin d'obtenir une cartographie de terrain 3D utile en agriculture de précision.

1.3.2.2 Shape from Texture

De la même façon que le fait la vision humaine, cette technique permet de reconstruire une forme à partir de sa surface en utilisant les informations contenues dans sa texture [Kanatani et Chou, 1989]. Les variations de cette texture sur l'image permettent d'estimer la forme de la surface observée à partir d'une seule image. Le champ de normales à la surface de l'objet est retrouvé à l'aide d'algorithmes complexes valables en théorie mais peu utilisés en pratique [Forsyth, 2002]. En effet, la reconstruction tridimensionnelle de la forme n'est possible que si les normales sont assez denses et la surface assez lisse avec une texture régulière (figure 1.3.6). Une texture régulière étant caractérisée

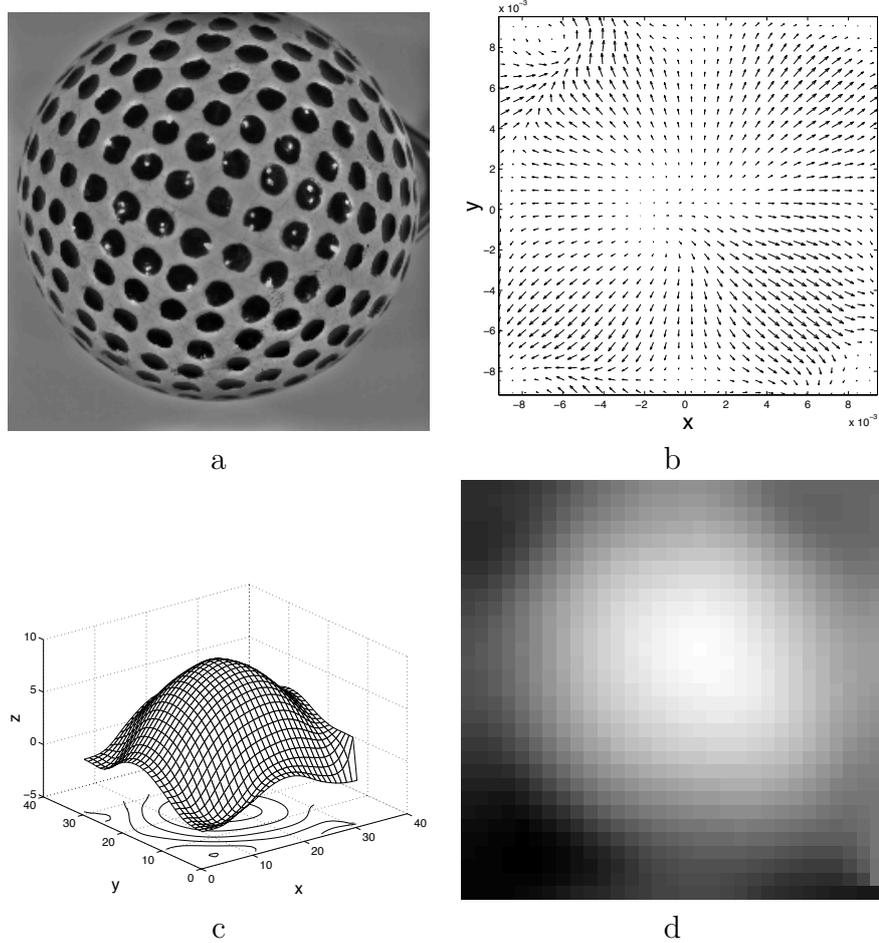


Figure 1.3.6: Shape from Texture. a. Image originale. b. Champ de normales. c. Surface reconstruite à partir du champ de normales. d. Carte de profondeur. [Clerc et Mallat, 2002].

par un motif de base se répétant périodiquement. Cette technique est donc valable seulement si la surface de l'objet ne présente pas de discontinuités.

1.3.2.3 Shape from Shading

Cette technique de reconstruction 3D monoculaire est basée sur l'analyse de la distribution d'intensité lumineuse réfléchie par l'objet ainsi que les ombres propres à la surface de l'objet. Elle fait l'objet de recherches depuis le début des années 70 [Horn, 1970] et un état de l'art des différentes approches de cette méthode peut être trouvé dans [Zhang et al., 1999]. L'analyse des niveaux de

gris de l'image permettant la reconstruction 3D dépend donc de plusieurs paramètres tels que les propriétés de réflectance de la surface, sa forme et les conditions d'illumination. Un certain nombre d'hypothèses doivent donc être formulées afin de pouvoir décrire la forme de l'objet. Cette technique implique beaucoup de contraintes et ne fonctionne que sous certaines conditions particulières comme, par exemple, la numérisation et la mise à plat de documents scannés [Courteille, 2006].

1.3.2.4 Shape from Motion

Autrement appelée « Structure From Motion », cette méthode permet de retrouver l'information de surface à partir du mouvement relatif estimé entre l'objet et la caméra. L'acquisition est réalisée soit par une caméra en mouvement réalisant une prise d'images successives ou une vidéo (figure 1.3.7), soit par un objet en mouvement par rapport au système d'acquisition. Globalement, la reconstruction est réalisée en calculant le déplacement entre les caméras par la recherche de points caractéristiques [Pollefeys et al., 2004].

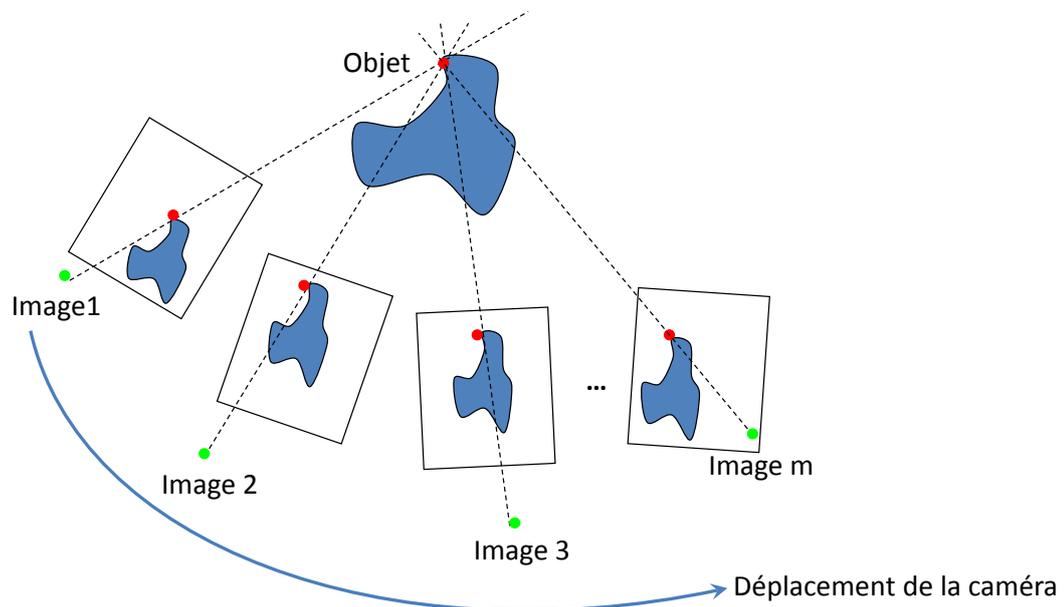


Figure 1.3.7: Principe d'acquisition « Shape from Motion ».

L'information de profondeur est retrouvée en analysant le mouvement apparent des points caractéristiques en fonction du mouvement de la caméra entre deux acquisitions (plus un point est éloigné de la caméra, plus celui-ci paraîtra avoir un faible déplacement entre deux acquisitions). Une application couramment utilisée est le SLAM (Simultaneous Localisation And Mapping) permettant de reconstruire un environnement en temps réel en fonction du déplacement d'une caméra présente sur un robot [Bailey, 2002]. Un exemple grand public est *Google Street View* qui propose une visualisation 3D des villes réalisée par « Structure from Motion » à l'aide d'un ensemble de caméras montées sur un véhicule.

1.3.2.5 Shape from Silhouette

Également connue sous le nom de « Visual Hull » [Laurentini, 1994], cette technique s'appuie sur l'acquisition de plusieurs images de l'objet à reconstruire depuis différents points de vue. Une fois la segmentation permettant de séparer l'objet de l'arrière plan effectuée, on obtient la silhouette de l'objet pour chaque point de vue. Afin d'obtenir les images, on peut soit utiliser plusieurs caméras (figure 1.3.8), soit positionner l'objet sur un plateau rotatif et réaliser les acquisitions avec une seule caméra. La reconstruction 3D est ensuite obtenue par assemblage des différentes silhouettes. Une application de cette méthode conjointement à la stéréovision peut être trouvée dans [Cheung et al., 2003]. Une limite importante de ce procédé est son non-fonctionnement dans le cas d'objets présentant des zones concaves. Contrairement aux méthodes précédemment présentées, les suivantes reposent sur la connaissance et la maîtrise des propriétés optiques du système utilisé.

1.3.2.6 Shape from Defocus

Dans le cas du « Shape from Defocus », on utilise une série limitée d'images (au minimum deux) acquises avec des paramètres optiques différents afin de ne pas avoir un flou optique identique dans les images. Pour obtenir l'information de profondeur, cette technique exploite le flou optique perceptible sur les images. Ce flou correspond à un étalement de l'intensité lumineuse cor-

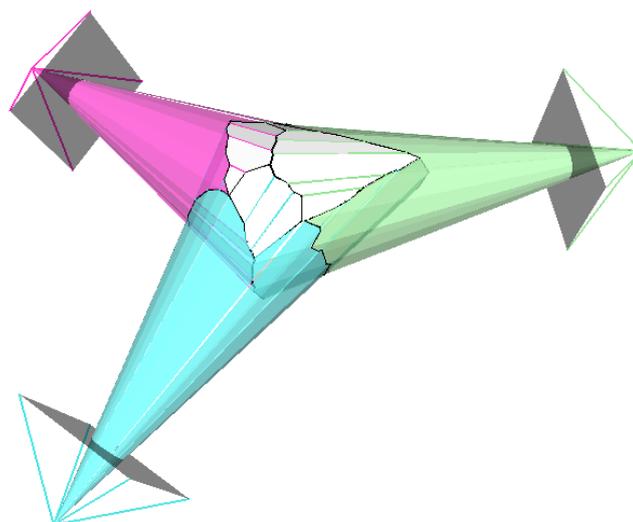


Figure 1.3.8: Schéma représentatif du « Visual Hull ».

respondant à un point de la scène sur le capteur. Cet étalement est lié à une mauvaise focalisation de l'image sur le plan capteur (cf chapitre 3 [Généralités optiques](#)). C'est l'estimation de cette fonction d'étalement du point (ou PSF : Point Spread Function) qui va permettre de remonter à l'information de profondeur car plus l'étalement est important, plus la distance objet/capteur l'est également. Plusieurs modélisations de cette fonction de flou sont proposées dans la littérature. On peut citer [[Pentland, 1987](#)] qui considère la fonction de flou comme une fonction Gaussienne et [[Subbarao, 1988](#)] qui propose de calculer la fonction de transfert optique (OTF) correspondant à la transformée de Fourier de la PSF. Quelle que soit la modélisation de la PSF choisie, la fonction de flou se comporte comme un filtre passe-bas dont la fréquence de coupure dépend de la défocalisation.

On notera que cette méthode est aussi quelquefois utilisée conjointement à un système de projection, faisant d'elle une technique active [[Pentland et al., 1994](#)]. Le principe est de projeter un motif sur la scène et de comparer ensuite l'image du motif projeté sur la scène à l'image du modèle projeté connue pour déterminer la profondeur de la scène. [[Favaro et al., 2003](#)] mettent en évidence le fait que cette méthode permet de reconstruire efficacement des surfaces lisses notamment lorsque la radiance de la scène est contrôlée (figure [1.3.9](#)). Cette méthode est difficilement applicable à des scènes complexes.

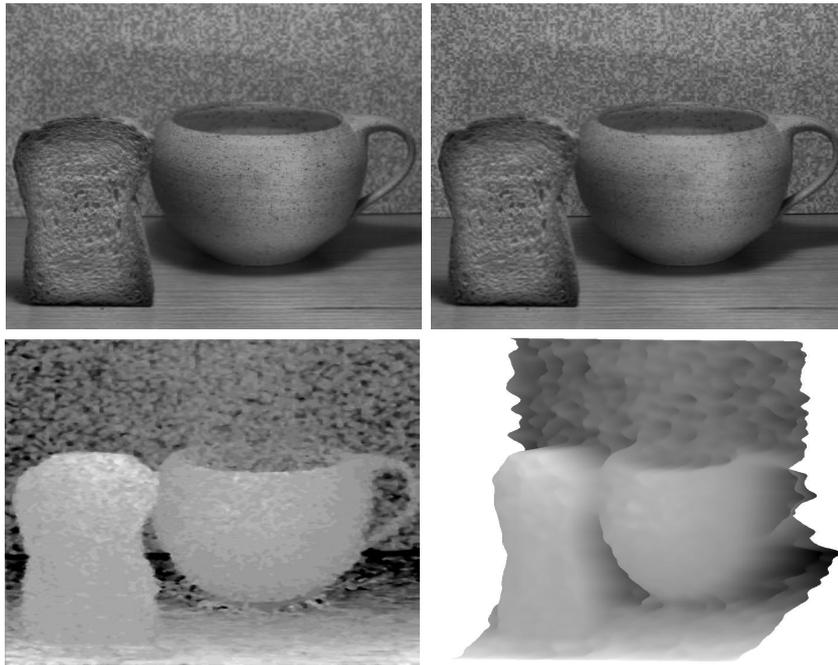


Figure 1.3.9: Exemple de reconstruction 3D par « Shape from Defocus ». [Favaro et Soatto, 2002]

1.3.2.7 Shape from Focus

Alors que la technique « Shape from Defocus » utilise l'information de flou pour déterminer la profondeur, la méthode « Shape from Focus » se base sur l'information de netteté. Également appelée « Depth from Focus », cette technique de reconstruction 3D a été introduite par [Nayar, 1989]. Elle se décompose en plusieurs étapes représentées sur la figure 1.3.10.

La première étape est l'acquisition de la séquence d'images de la scène dont on souhaite effectuer la reconstruction 3D. La procédure détaillée d'acquisition est expliquée au prochain chapitre mais il est bon de noter qu'à chaque pixel correspond une image dans laquelle celui-ci est net. L'idée générale d'acquisition étant un balayage complet de la scène par le plan focal.

Une fois la séquence d'images obtenue, une évaluation de la netteté de chaque pixel de la séquence est faite. Le même pixel de chaque image de la séquence doit représenter le même point spatial dans la scène. Une mesure de netteté locale est appliquée à chaque pixel et son voisinage permettant de

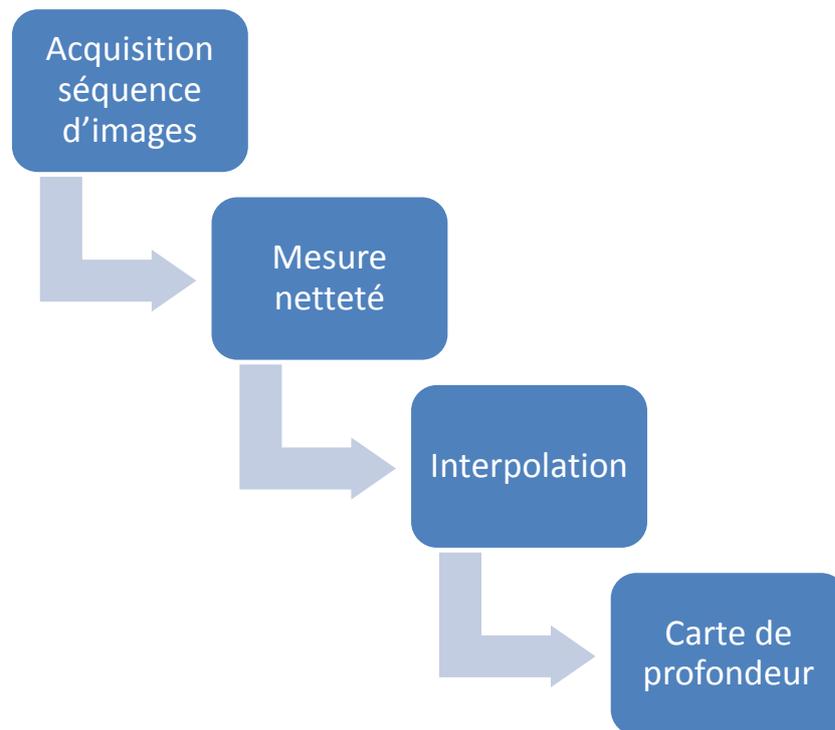


Figure 1.3.10: Principe du « Shape from Focus ».

déterminer l'image dans laquelle chaque point de la scène est le plus net. On retrouve ici le principe de base de l'autofocus en photographie numérique qui consiste à maximiser une mesure de netteté afin de focaliser correctement sur l'objet visé.

Une valeur de netteté étant déterminée pour chaque pixel de chaque image, il peut être intéressant d'interpoler la courbe de netteté obtenue pour chaque point de la scène. En effet, en fonction de la précision du système d'acquisition utilisé, la position de la valeur maximum de netteté n'est pas nécessairement la position réelle du point de la scène. Comme le but est d'obtenir une estimation de la profondeur en tout point de la scène la plus précise possible, il est bon d'approximer la position du point maximum de la courbe de netteté. Quelques méthodes d'interpolation utilisées en « Shape from Focus » sont présentées chapitre 3 section 3.3.

Finalement, la reconstruction 3D de la scène se déroule en plusieurs étapes en fonction des situations. L'étape commune à toutes les situations est la

création de la carte de profondeur. Cette dernière est constituée des différentes positions spatiales des points de la scène correspondant aux positions approximées lors de l'étape précédente. Ces positions spatiales peuvent être représentées par des niveaux de gris pour une meilleure clarté. Par exemple, plus un pixel est blanc et plus le point correspondant est éloigné du système d'acquisition. Inversement, plus il est sombre et plus il est proche du système. Avec un calibrage du système, ces niveaux de gris renvoient directement à une position métrique de chaque point. Une représentation en fausses couleurs est également possible pour permettre une meilleure visualisation.

Si la situation n'exige que l'obtention de l'information de profondeur en tout point de la scène considérée, alors seule la carte de profondeur est nécessaire. Si une représentation 3D de la scène est souhaitée, la carte de profondeur est projetée sur trois axes et la texture de la scène est plaquée sur la surface projetée. Connaissant les images où les pixels sont nets, une image de la scène entièrement nette peut être construite et utilisée pour le plaquage de texture.

Cette technique de reconstruction 3D est habituellement utilisée dans le domaine de l'infiniment petit, comme la microscopie confocale qui peut être assimilée au « Shape from Focus ». La différence est l'utilisation d'une source lumineuse collimatée ne permettant pas de rayons divergeants, et donc seules les zones nettes sont conservées. L'avantage principal de cette méthode est son fonctionnement sur des scènes où le phénomène d'occlusion est très présent. De plus, en fonction des propriétés optiques du système, elle permet une reconstruction très précise et dense de la scène considérée.

1.3.2.8 Confocal stereo

Méthode pouvant être considérée comme une technique dérivée du « Shape from Focus » et « Defocus », celle-ci a été récemment proposée par [[Hasinoff et Kutulakos, 2009](#)]. L'idée est de ne pas évaluer la netteté directement sur les images acquises localement mais sur une image particulière nommée « AFI » pour « Aperture-Focus Image ». Cette image est le résultat de la mosaïque construite par un même point de la scène projeté sur le capteur au travers de différents paramètres optiques. La séquence d'images nécessaire est acquise

par un balayage de la scène correspondant à une variation de la focalisation (similaire au « Shape from Focus ») mais pour chaque focalisation, une variation de l'ouverture est également effectuée. Une image différente est donc acquise pour chaque couple focalisation/ouverture. L'analyse de chaque image AFI, correspondant à chaque point de la scène, permet de déterminer quel doublet focalisation/ouverture renvoie au point le plus net. Ceci permet l'obtention de l'information de profondeur et la reconstruction 3D de la scène. L'inconvénient majeur de cette méthode est le nombre bien plus important d'images nécessaires pour une bonne reconstruction en comparaison du « Shape from Focus ».

1.3.2.9 Systèmes plénoptiques

Les prémices de ce procédé ont été décrites par le prix Nobel de Physique Gabriel Lippmann dans [Lippmann, 1908] sous le terme de « Photographie intégrale ». Pourtant, c'est plus récemment que le procédé a vraiment vu le jour sous une forme opérationnelle. Le principe repose sur la superposition d'un réseau de micro-lentilles sur le capteur image. La maîtrise de ce réseau de lentilles permet de capturer plusieurs images de la même scène issues de plusieurs distances de focalisation différentes. Toutes ces images sont imbriquées sur le capteur en fonction du réseau de lentilles. La profondeur est calculée en étudiant la netteté des micro-images associées aux micro-lentilles puisque la distance entre un point de la scène et ces dernières est connue. Deux entreprises commercialisent actuellement ce type de système : Raytrix pour des caméras industrielles et Lytro pour des appareils plutôt grands publics (figure 1.3.11).

L'entreprise Lytro est le fruit des recherches réalisées sur les systèmes plénoptiques par [Ng, 2006]. Le grand intérêt de ce type d'appareil en photographie est la possibilité de faire la mise au point *a posteriori* de la prise d'image. La limitation principale de ces systèmes est induite par son mode de fonctionnement : un seul capteur pour plusieurs distances de focalisation. Cela entraîne une réduction de la surface utile du capteur et donc une définition moindre des images. A l'heure actuelle, seules les caméras Raytrix proposent

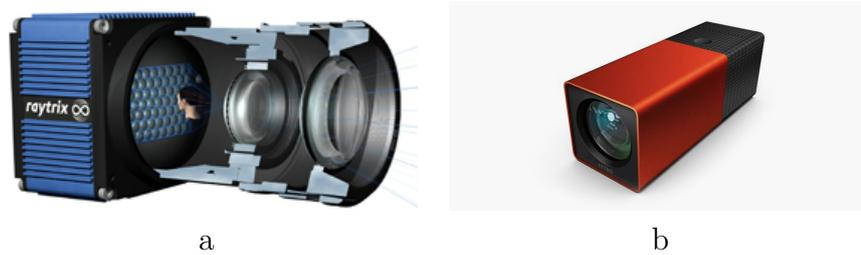


Figure 1.3.11: Exemple de systèmes plénoptiques. a. Caméra Raytrix. b. Appareil Lytro.

une reconstruction 3D avec leurs appareils.

1.4 Conclusion

De nombreuses méthodes, permettant une reconstruction 3D d'une scène ou d'un objet, existent à l'heure actuelle. La méthode à utiliser dépend principalement des contraintes liées à l'application et à la scène considérée. Dans notre cas, les scènes issues du domaine agronomique sont bien souvent jugées complexes. En effet, les scènes naturelles auxquelles nous nous intéressons présentent beaucoup de discontinuités ainsi qu'une répartition aléatoire des objets les composant. Un premier constat est de ne pas retenir les méthodes nécessitant des objets ou scènes présentant une surface continue. Un second constat est la forte présence d'occlusions, dans le cas d'utilisation de techniques multi-vues, biaisant l'appariement nécessaire à la triangulation. Les méthodes à base de triangulation sont donc à écarter. De plus, le système de reconstruction 3D devra être transportable pour une acquisition en plein champ, en plus des acquisitions faites en laboratoire. Enfin, l'objectif n'est pas une reconstruction 3D ayant pour but une visualisation, mais la détermination de l'information de profondeur en tout point de la scène.

Du fait de la volonté de création d'un système de proxidétection de terrain sur des scènes naturelles, les méthodes avec contact sont donc à proscrire. De même pour les méthodes non-optiques difficilement utilisables sur le terrain. Certaines techniques à base de lumière structurée seraient envisageables mais le caractère transportable du système rend ces méthodes difficilement

adaptables.

En résumé et, par rapport à nos contraintes, les méthodes potentiellement utilisables sont les suivantes :

- Systèmes plénoptiques ;
- Confocal stereo ;
- Shape from Focus.

Parmi les systèmes plénoptiques actuellement disponibles sur le marché, les caméras Raytrix sont des outils intéressants. En effet, une précision en profondeur de l'ordre de 2 mm est obtenue pour un champ de vue de 210 mm par 140 mm, le tout en une seule acquisition. Cependant, de tels systèmes d'acquisitions sont des outils très onéreux (de 25 k€ à 45 k€ en fonction de la caméra, de l'objectif et du logiciel associé choisi).

Dans le cas de la récente technique baptisée « Confocal stereo », l'utilisation d'appareil photographique numérique professionnel permettant un contrôle automatique des différents paramètres optiques et des prises de vues est nécessaire. Plus abordable que des appareils plénoptiques, cette technique nécessite cependant un nombre important d'images de la scène pour une bonne reconstruction.

Enfin, la méthode « Shape from Focus » permet l'utilisation de caméras et objectifs standards. Le nombre d'images nécessaire à cette méthode de reconstruction 3D est bien moins élevé que pour la méthode « Confocal stereo ». Nous avons donc choisi de nous orienter sur ce type de méthode qui semble être un compromis entre coût raisonnable du système global et précision. De plus, cette méthode étant habituellement utilisée dans le domaine microscopique, il est intéressant de tester sa faisabilité pour des utilisations macroscopiques.

Chapitre 2

Systemes d'acquisition

L'objectif de ce chapitre est de détailler les travaux de réflexion menant à la conception du système de proxidétection. Ce dernier permet l'acquisition de la séquence d'images alimentant la méthode de reconstruction 3D « Shape from Focus ». Cette méthode repose principalement sur la maîtrise des propriétés optiques du système. Une première sous-section est consacrée aux rappels des principes optiques liés à ces travaux ainsi que des détails des paramètres optiques dont nous avons le contrôle. Un premier système expérimental, ayant pour but la validation de la faisabilité de la méthode de reconstruction 3D retenue, est décrit par la suite. Enfin, la description du système final, son calibrage et la procédure d'acquisition associée sont présentés.

2.1 Généralités optiques

Cette section présente les notions de base permettant de mieux comprendre les phénomènes intervenant lors du processus de formation des images. La présentation de ce dernier est l'occasion de définir la notion de profondeur de champ, directement liée à la précision de la reconstruction 3D. Cette profondeur de champ est en lien avec plusieurs paramètres optiques détaillés dans une deuxième sous-section.

2.1.1 Formation géométrique des images

La façon la plus simple d'expliquer le processus de formation d'une image passe par la présentation de l'appareil sténopé. L'appellation anglophone de cet appareil est « pinhole camera » et permet aisément d'imager le processus de formation d'une image. En effet, cet appareil est une boîte hermétique à la lumière dans laquelle est percé un trou dont le diamètre est supposé nul. Ainsi, tous les rayons lumineux passant par le trou projettent la scène observée renversée sur le fond de la boîte (figure 2.1.1). En réalité, le diamètre du trou ne pouvant être nul, il revient à un compromis entre une petite taille engendrant trop peu de lumière et un diamètre trop grand rendant l'image projetée floue.

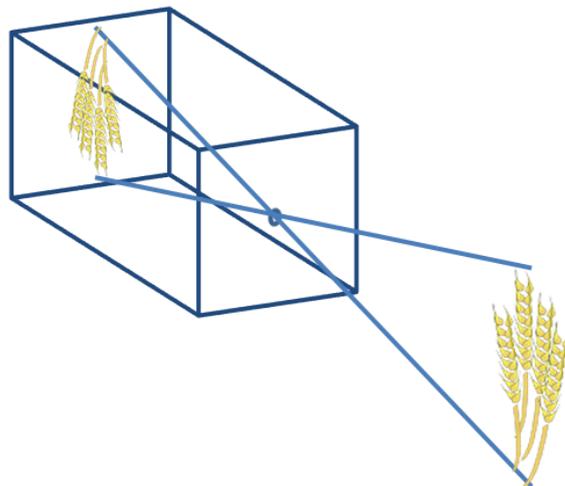


Figure 2.1.1: Schéma d'un appareil sténopé.

Le remplacement de ce trou par une lentille permet de laisser passer plus de lumière tout en obtenant des images nettes. Cependant, l'utilisation de lentilles ajoute une complexité au processus de formation des images. En effet, les rayons lumineux provenant de la scène et passant par une lentille subissent un changement de milieu puisqu'ils traversent l'air puis un matériau spécifique. Le comportement des rayons lumineux lors de ce changement de milieu suit la loi de Snell-Descartes qui, d'après le schéma figure 2.1.2, conduit à l'équation :

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2). \quad (2.1.1)$$

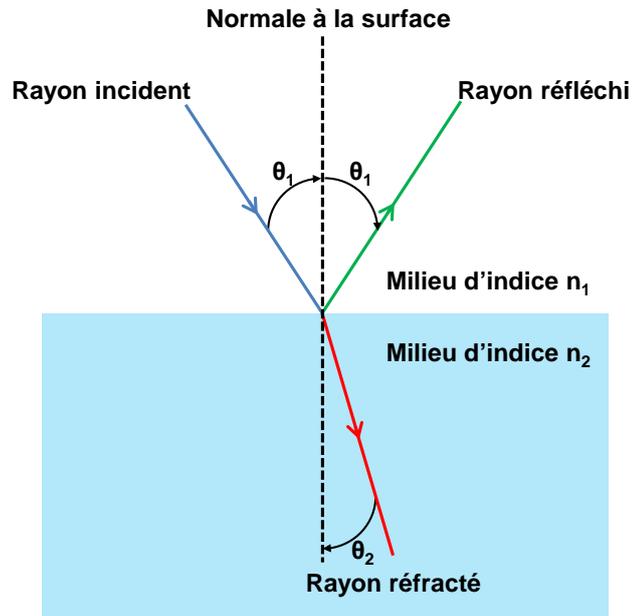


Figure 2.1.2: Loi de Snell-Descartes.

A cette loi est associé un des principes de base en optique qui est que la lumière, constituée de rayons lumineux parallèles, se propage en ligne droite dans un milieu transparent, homogène et isotrope. Ainsi, chaque rayon incident est réfléchi et réfracté lors du passage par le point d'incidence. Ce point d'incidence correspond au changement de milieu entre l'air et le matériau constituant la lentille. L'angle d'incidence θ_1 du rayon incident conduit, en fonction de l'indice du milieu n_2 lié au matériau de la lentille, à un angle de réfraction θ_2 . L'angle du rayon réfléchi par la surface de la lentille est quant à lui d'un angle identique à l'angle d'incidence.

La constitution des lentilles permet donc une focalisation des rayons lumineux sur le capteur. Par définition, une lentille est un milieu transparent, homogène et isotrope et est constituée de deux dioptries. Un dioptrie est la surface séparant deux milieux d'indices différents. On distingue les lentilles convergentes permettant le stigmatisme, c'est-à-dire une focalisation de tous les rayons issus d'un même point de la scène, des lentilles divergentes entraînant un angle $\theta_2 > \theta_1$.

On retrouve deux modèles de lentille : la lentille mince et la lentille épaisse.

Le second modèle est utilisé pour les calculs optiques nécessaires à la conception des objectifs puisqu'étant au plus proche de la réalité. Cependant, afin de simplifier les calculs dans les cas autres que la conception optique, l'épaisseur des lentilles est considérée comme négligeable et donc le modèle de lentille mince est utilisé (figure 2.1.3).

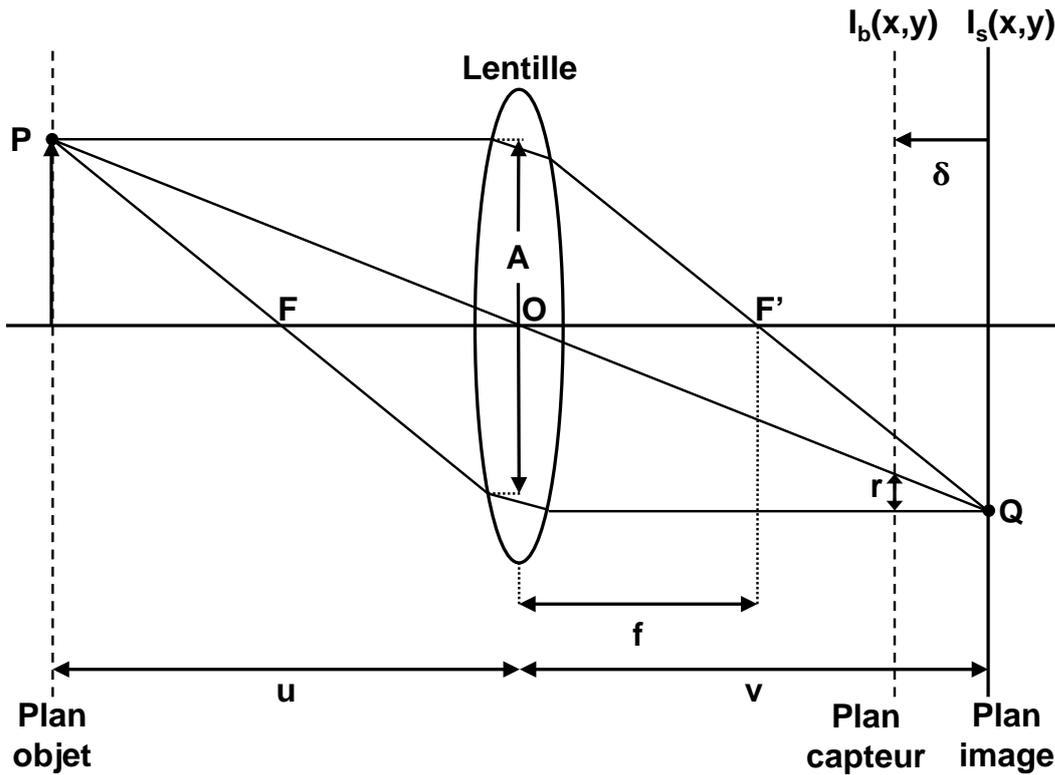


Figure 2.1.3: Schéma de formation d'une image par une lentille mince convergente.

Où $f = OF = OF'$ est la focale de la lentille ;

u la distance entre un point P d'un objet et le centre optique de la lentille O ;

v la distance entre le centre optique O et le point de focalisation Q correspondant au point P sur le plan image ;

δ est la distance entre le plan image et le plan capteur ;

O est le centre optique de la lentille ;

F est le foyer objet ;

F' est le foyer image ;

P est un point de la scène ;

Q est le point projeté de P par la lentille ;

A est l'ouverture relative du diaphragme ;

r est le rayon d'étalement de l'énergie d'un point défocalisé.

Les propriétés spécifiques à une lentille mince sont :

- tout rayon passant par le centre optique de la lentille O n'est pas dévié ;
- tout rayon arrivant parallèlement à l'axe optique est dévié puis suit une droite passant par un foyer image F' ;
- tout rayon passant par le foyer objet F ressortira parallèle à l'axe optique.

L'équation liée au modèle de lentille mince présenté figure 2.1.3 est la suivante :

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}. \quad (2.1.2)$$

Chaque point de l'objet ou de la scène considéré est projeté sur le plan image formant ainsi l'image $I_s(x, y)$. Dans le cas où le plan image ne coïncide pas avec le plan capteur d'une distance δ , alors l'énergie reçue du point P par

la lentille n'est pas concentrée en un point mais distribuée sur le plan capteur formant une image floue $I_b(x, y)$. La forme de cette distribution est fonction de la forme de l'ouverture de l'objectif qui est considérée comme circulaire. L'énergie est donc distribuée suivant une forme circulaire. Le rayon de cette forme est approximé par l'équation :

$$r = \frac{\delta \cdot A}{v}. \quad (2.1.3)$$

L'image floue $I_b(x, y)$ ainsi formée sur le plan capteur est considérée comme le résultat d'une convolution entre une image nette $I_s(x, y)$ et une fonction d'étalement du point $h(x, y)$ (PSF en anglais pour « Point Spread Function »). [Pentland, 1987] conçoit la fonction d'étalement comme approximée par une gaussienne :

$$h(x, y) = \frac{1}{2\pi\sigma_h^2} e^{-\frac{x^2+y^2}{2\sigma_h^2}}. \quad (2.1.4)$$

Le paramètre d'étalement de l'énergie σ_h est proportionnel au rayon r donc plus la distance δ est grande, plus les hautes fréquences sont filtrées entraînant une image de plus en plus floue.

Beaucoup de modélisations de la fonction d'étalement existent dans la littérature mais l'approche de Pentland a l'intérêt d'être appliquée pour une recherche de l'information de profondeur comme dans notre cas. Globalement, la fonction d'étalement correspond souvent à un filtre passe-bas.

2.1.2 Profondeur de champ

C'est la notion essentielle sur laquelle jouer pour obtenir une reconstruction 3D la plus précise possible de la scène par « Shape from Focus ». Contrairement au système sténopé, l'utilisation d'un système optique à base de lentille n'entraînera pas une projection nette d'un seul plan objet mais d'un ensemble de plans (figure 2.1.4).

En effet, chaque point de la scène est considéré comme net dès lors que le diamètre de l'énergie distribuée sur le plan capteur est inférieur à C . Il s'agit donc d'une zone de la scène d'une profondeur PdC dont la projection sur le

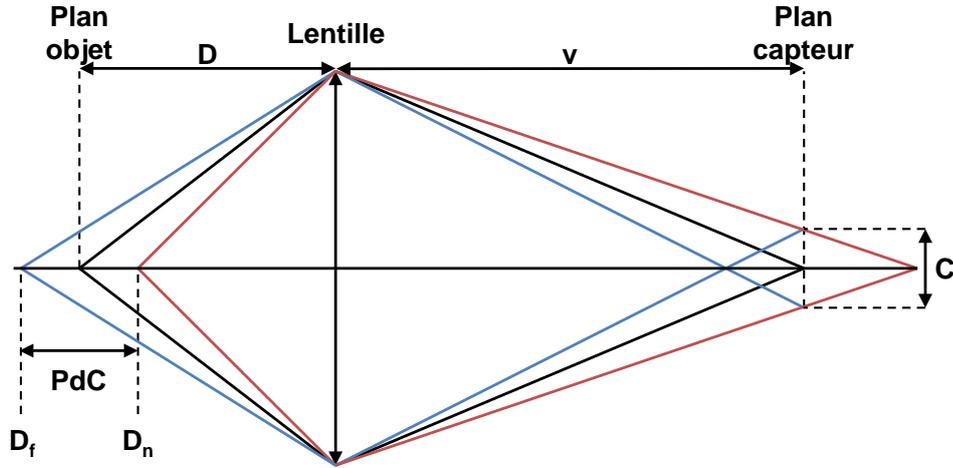


Figure 2.1.4: Schéma illustrant la profondeur de champ.

plan capteur est nette.

La profondeur de champ PdC est calculée par les équations suivantes :

$$\begin{cases} D_n = \frac{D}{1 + C.A. \frac{D-f}{f^2}} \\ D_f = \frac{D}{1 - C.A. \frac{D-f}{f^2}} \\ PdC = D_f - D_n \end{cases} \quad (2.1.5)$$

- où
- D est la distance entre la scène et le système optique (mm) ;
 - D_n est la distance entre le premier plan net et l'objectif (mm) ;
 - D_f est la distance entre le dernier plan net et l'objectif (mm) ;
 - C est le diamètre du cercle de confusion (mm) ;
 - A est l'ouverture relative du diaphragme ;
 - f est la focale de l'objectif (mm).

L'équation simplifiée est la suivante :

$$PdC = \frac{2.A.C.f^2.D.(D-f)}{f^4 - A^2.C^2.(D-f)^2}. \quad (2.1.6)$$

La profondeur de champ est donc dépendante de quatre paramètres : la

focale, l'ouverture du diaphragme, la distance d'acquisition et le diamètre du cercle de confusion. Ces paramètres, ainsi que leurs effets sur l'image acquise et la profondeur de champ, sont détaillés dans les prochains paragraphes.

Focale

La focale f de l'objectif utilisé est exprimée en millimètre et correspond à la distance focale. Le choix de ce paramètre est souvent lié à la dimension souhaitée de la scène à visualiser. En effet, à chaque distance focale correspond un angle de champ particulier qui sera grand pour une focale courte et petit pour une focale longue d'après l'équation suivante :

$$\alpha = 2 \cdot \arctan\left(\frac{d}{2 \cdot f}\right), \quad (2.1.7)$$

où f est la distance focale (mm) et d la diagonale du capteur (mm). Une scène de dimension élevée est acquise par un objectif de focale courte alors qu'une scène de dimension réduite l'est par un objectif de focale élevée. On retrouve ici le phénomène communément appelé « zoom » qui n'est rien d'autre qu'une augmentation de la distance focale.

Plus la focale est courte et plus la quantité de lumière captée est importante du fait de l'élargissement du champ de vue. Cet effet est pris en compte lors de la conception des objectifs en diminuant le diamètre réel de l'ouverture proportionnellement à la focale. Ainsi, un éclairage constant est obtenu pour une même valeur d'ouverture mais une focale différente.

Pour en revenir à la profondeur de champ, celle-ci est affectée par une variation de la distance focale qui est le paramètre l'impactant le plus celle-ci du fait de l'ordre supérieur dans l'équation 2.1.6 par rapport aux autres paramètres. Ainsi, une petite augmentation de la focale entraîne une forte diminution de la profondeur de champ.

Ouverture

Par convention, l'ouverture relative A est exprimée comme étant $A = f/D$ avec f la focale et D le diamètre de l'objectif (nommé « $F - Number$ » en

anglais). L'ouverture relative varie selon une plage dépendante de l'objectif selon une suite géométrique de raison $\sqrt{2}$. Par exemple, $A = 1.4, 2, 2.8 \dots 16$ où $A = 1.4$ correspond à une grande ouverture du diaphragme alors qu'une valeur plus grande ($A = 16$) indique une ouverture plus faible où, pour chaque incrémentation de A , deux fois moins de lumière traverse l'objectif (figure 2.1.5). La valeur de cet indice est la valeur à utiliser dans la formule de calcul de la profondeur de champ (équation 2.1.6). Ainsi, plus l'ouverture relative A est faible et plus la profondeur de champ est réduite.

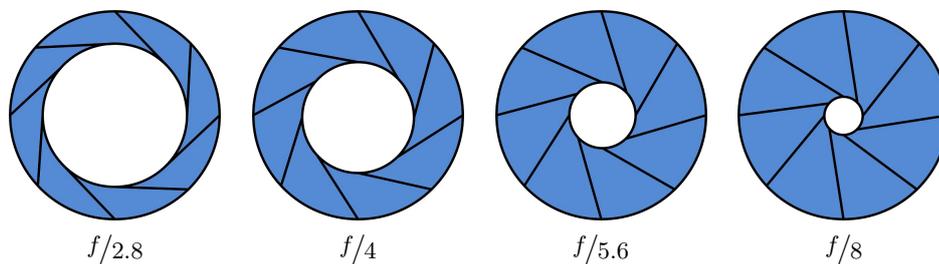


Figure 2.1.5: Exemples d'ouvertures relatives.

Distance

La distance D est la distance de travail entre l'objectif et le point de la scène considérée. Elle est définie comme la distance entre l'objectif et le plan objet le plus net. Cette distance est corrélée à la valeur de réglage de la bague de mise au point de l'objectif également appelée « focus ». En règle générale, le réglage de cette distance de mise au point permet de faire coïncider le plan objet avec le point de l'espace où la netteté est souhaitée. En photographie, la mise au point peut être automatique (« Autofocus » en anglais) et de comportement similaire au phénomène d'accommodation de l'œil humain, c'est-à-dire que le plan objet net coïncide automatiquement sur le point de l'espace considéré. Mécaniquement, une variation du focus est réalisée par une translation du réseau de lentilles constituant l'objectif afin de le rapprocher ou l'éloigner du capteur. Ainsi, plus la distance de mise au point est proche et plus la profondeur de champ est réduite.

Cercle de confusion

A l'origine en photographie, le diamètre du cercle de confusion correspondait au diamètre des plus petits points juxtaposés identifiables à l'œil nu d'une distance d'environ un mètre. On retrouvait également l'utilisation de la formule dite de Zeiss considérant le diamètre de ce cercle comme égal à $d/1730$, avec d le diamètre de la surface photosensible. Depuis l'avènement des capteurs numériques (CCD ou CMOS), les fabricants conçoivent le diamètre du cercle de confusion comme étant de la même taille qu'un pixel composant le capteur [Falco, 2010].

Ainsi, d'après la formule 2.1.6 et la figure 2.1.3, plus le diamètre du cercle de confusion est petit et plus la profondeur de champ est faible.

En résumé, la profondeur de champ diminue lorsque la longueur focale augmente, la distance de mise au point diminue, le diamètre d'ouverture du diaphragme augmente et celui du cercle de confusion diminue.

Les changements de focale et de distance de travail ont également pour effet une modification du champ de vue (« Field of View » en anglais), c'est-à-dire des dimensions de la scène visualisée. Celui-ci est également dépendant du type de capteur utilisé comme l'indiquent les équations suivantes :

$$\frac{w}{W} = \frac{h}{H} = \frac{f}{D}, \quad (2.1.8)$$

$$W = \frac{w \cdot D}{f}, \quad (2.1.9)$$

$$H = \frac{h \cdot D}{f}, \quad (2.1.10)$$

où D est la distance de mise au point (mm);
 w est la largeur du capteur (mm);
 h est la hauteur du capteur (mm);
 W est la largeur de la scène (mm);
 H est la hauteur de la scène (mm);
 f est la focale de l'objectif (mm).

Ainsi, un exemple de la profondeur de champ et du champ de vue accessible en fonction du type d'objectif est illustré tableau 2.1. Pour les calculs, on considère la caméra utilisée dans le système expérimental (section 2.2) constituée d'un capteur 1/2 pouce ($w = 6.4\text{ mm}$ et $h = 4.8\text{ mm}$) dont les pixels mesurent $4.65\ \mu\text{m}$. On remarque qu'une faible profondeur de champ va de pair avec une diminution du champ de vue.

D	25mm f1.4			35mm f1.6			50mm f2		
	L	H	PdC	L	H	PdC	L	H	PdC
500	128	96	4.9	91.4	68.5	2.8	64	48	1.67
600	153.6	115.2	7.2	109.7	82.3	4.1	76.8	57.6	2.4
700	179.2	134.4	9.8	128	96	5.6	89.6	67.2	3.4
800	204.8	153.6	12.9	146.3	109.7	7.4	102.4	76.8	4.4
900	230.4	172.8	16.4	164.5	123.4	9.4	115.2	86.4	5.7
1000	256	192	20.3	182.8	137.1	11.7	128	96	7

Tableau 2.1: Champ de vue (Largeur et Hauteur) et profondeur de champ (PdC) en fonction du type d'objectif (valeurs en mm).

2.1.3 Acquisition d'une séquence d'images

Notre objectif est le développement du système d'acquisition de la séquence d'images nécessaire à une reconstruction 3D par « Shape from Focus ». Le principe d'acquisition de la séquence d'images repose sur le balayage de la scène par le plan objet. Ce plan objet est en réalité l'association d'un nombre fini de plans constituant la profondeur de champ.

Ainsi, la précision de l'information de profondeur (Z) va dépendre directement de cette profondeur de champ induite par les paramètres optiques. En effet, plus la profondeur de champ sera petite et plus la précision en Z

sera importante. La contrainte est un nombre d'images nécessaires bien plus important. Cela revient donc à un compromis entre contraintes d'acquisition, précision et champ de vue souhaité.

Ce balayage de la scène peut être obtenu de différentes façons.

1. Déplacer l'objet à reconstruite par rapport au système d'acquisition fixe ;
2. Déplacer le système d'acquisition par rapport à la scène ;
3. Faire varier le plan objet par la variation de la distance de mise au point ;
4. Faire la varier la distance focale par l'utilisation d'un objectif à focale variable (zoom).

On peut d'ores et déjà éliminer la quatrième solution du fait de la modification importante de la profondeur de champ entraînée par un changement de focale. Une variation de la distance de mise au point engendre une variation de la profondeur de champ minime tout en permettant un balayage de la scène par le plan objet avec un système d'acquisition fixe. De plus, ces solutions entraînent une variation du grossissement entre chaque image de la séquence. Ceci implique la nécessité d'un recalage de toutes les images de la séquence. Les solutions de correction de ce grossissement sont étudiées sous-section [2.4.2](#).

Bien entendu, un déplacement de la scène n'est, dans notre cas, pas envisageable contrairement à un déplacement du système d'acquisition. C'est cette dernière solution qui, même si elle implique également des effets de grossissement, a été retenue dans un premier temps. La raison principale de ce choix est la volonté de création d'un premier système expérimental afin de tester la faisabilité de cette méthode de reconstruction 3D.

2.2 Système expérimental

2.2.1 Vue d'ensemble

La solution de déplacement du système optique a été retenue afin d'obtenir la séquence d'images représentant un balayage de la scène par le plan objet. Ce procédé d'acquisition est schématisé figure [2.2.1](#).

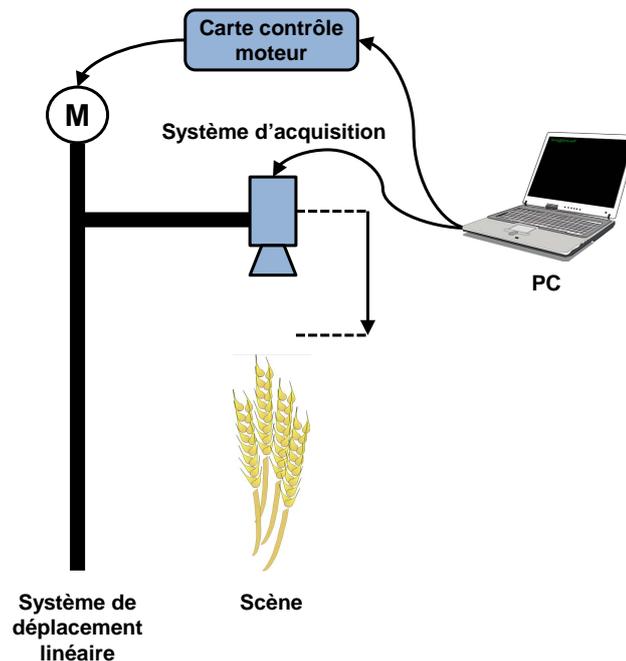


Figure 2.2.1: Schématisation du procédé d'acquisition.

Ainsi, on peut scinder l'ensemble du procédé d'acquisition en deux grandes parties : la partie optique permettant les acquisitions en tant que telles et la partie électronique permettant le déplacement de l'ensemble optique. L'utilisateur contrôle l'acquisition et le déplacement par le biais d'un ordinateur. Les deux parties sont détaillées ci-dessous.

2.2.2 Partie optique

L'ensemble des éléments permettant l'acquisition des images est illustré figure 2.2.2. Il est constitué de plusieurs éléments listés ci-dessous avec leurs caractéristiques.

Caméra

La caméra utilisée dans ce système est une caméra couleur à filtre de Bayer superposant un capteur CCD (Imaging Source DBK 41BU02.H). Le capteur est de type $1/2''$ (6.4×4.8 mm) avec une taille de photosite de $4.65 \mu\text{m}^2$. La



Figure 2.2.2: Ensemble constituant la partie optique.

définition finale des images est de 1280 par 960 pixels (1.2 mégapixels). Le transfert et la commande de la caméra sont réalisés par le protocole USB 2.0.

Objectif

L'objectif associé à la caméra est un objectif de focale 50 mm et d'ouverture relative $f1.4$. La distance de mise au point minimum de cet objectif est de 82 cm.

Illuminant

Afin de ne pas dépendre des conditions d'illumination extérieures mais surtout de permettre un éclairage constant tout au long du processus d'acquisition, deux diodes électroluminescentes sont disposées de part et d'autre de la caméra. Les diodes, de température de couleur de 4000 kelvins (blanc neutre), ont une puissance de 3 watts et sont montées sur un dissipateur thermique en

aluminium. Cette couleur permet de couvrir l'ensemble du spectre du visible en ne privilégiant pas de longueurs d'onde particulières, ce qui permet une utilisation relativement générale sur différents types de scène.

En considérant une distance de mise au point de 1 mètre, la scène couverte par le système d'acquisition est de 12.8 cm par 9.6 cm avec une profondeur de champ de 5 mm. Le diaphragme de l'objectif est ouvert au maximum, c'est-à-dire à l'ouverture relative $f1.4$.

2.2.3 Partie électronique

De la même façon que la partie optique, la partie électronique est constituée de plusieurs éléments permettant le déplacement du système d'acquisition. Ce déplacement se doit d'être commandé précisément et facilement par un ordinateur.

Système de déplacement

Le déplacement linéaire est assuré par un module de guidage linéaire (Drylin SHT-12-AWM-1000). Long d'un mètre, un système de vis sans fin trapézoïdale permet l'entraînement de la partie mobile sur laquelle est fixée la partie optique. La charge axiale maximale entraînée par le module de guidage dépend du couple du moteur utilisé et suit l'abaque constructeur illustré figure [2.2.3](#).

Moteur

Le couplage entre l'axe moteur et l'axe du système de guidage est direct par montage axial. Ainsi, les recommandations de la figure [2.2.3](#) sont respectées puisque les forces appliquées à la vis sans fin sont essentiellement de type axial lors d'un couplage direct. Le moteur choisi doit avoir un couple permettant l'entraînement d'une charge maximale de 2 kg (partie optique). D'après l'abaque, une charge de 2 kg (environ 20 N) est entraînée dès lors que le couple du moteur est de 0.3 Nm. Le type de moteur utilisé dans notre cas est un moteur pas à pas. Contrairement à d'autres types de moteurs (courant continu, servomoteur ...), le contrôle de la position est plus précis pour

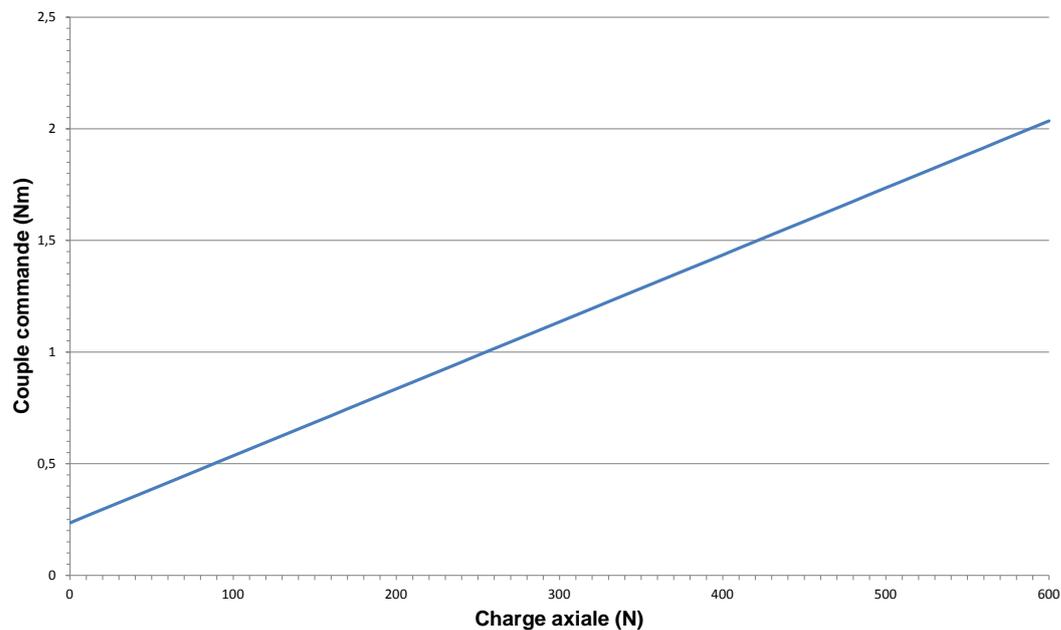


Figure 2.2.3: Couple moteur nécessaire au module de guidage en fonction de la charge axiale entraînée.

un moteur pas à pas du fait que la commande d'une impulsion entraîne un léger déplacement appelé pas. Ainsi la position est contrôlée précisément et est connue à tout instant alors qu'un moteur à courant continu, par exemple, nécessite un codeur pour obtenir un retour d'information sur celle-ci. C'est cette importante précision qui rend ces moteurs très utilisés dans des systèmes de positionnement. On retrouve plusieurs types de moteurs pas à pas : les moteurs à aimants permanents, à réluctance variable et les hybrides. Le premier est bon marché mais d'un nombre de pas par tour peu important. Le couple d'utilisation est plus élevé qu'un moteur à réluctance variable mais, contrairement à ceux-ci, présente un couple résiduel lorsque le moteur est hors tension. La résolution des moteurs à réluctance variable est meilleure mais de couple plus faible que les moteurs à aimants permanents.

Les moteurs hybrides combinent les deux technologies. Ainsi, ce sont des moteurs avec un meilleur couple, une vitesse plus élevée et une résolution importante. Notre moteur est un moteur 5V, 1A développant un couple de 500 mNm. La résolution est de 200 pas par tour en mode pas entier et 400 pas par tour en mode demi-pas. A raison d'un déplacement du module de guidage

de 2 mm pour une rotation de l'axe de un tour, on obtient une précision de 5 μm en mode demi-pas et 10 μm en mode pas entier. La précision est donc extrêmement importante avec ce type de moteur. A noter que le mode demi-pas implique un couple deux fois plus important qu'en mode pas entier pour une vitesse deux fois plus faible. En effet, le demi-pas est obtenu par une alimentation intermédiaire des bobines permettant une rotation du rotor de 0.9 degré au lieu de 1.8 degrés en mode pas entier.

Cartes de commande et de puissance

La partie électronique nécessaire à la gestion du déplacement par l'utilisateur est constituée de deux cartes. La première est la carte de commande dont l'architecture tourne autour d'un microcontrôleur (Microchip PIC 18F4550). L'objectif de cette carte est l'interfaçage entre la carte de puissance et le logiciel de commande. Une liaison USB entre l'ordinateur et la carte de commande permet le transfert des informations utiles à la commande du moteur. Ces informations sont la marche ou l'arrêt du moteur, son sens et sa vitesse de rotation, le choix du mode pas entier ou demi-pas ainsi que le nombre de pas (ou de tours) désiré. Ces informations sont ensuite transmises à la carte de puissance chargée de délivrer les bons signaux de commandes au moteur. Cette séparation de la partie commande et puissance a une double justification. La première est que le moteur, pour un couple maximum, nécessite un courant de 1A qui ne peut pas être fourni directement par les sorties du microcontrôleur limitées à 20mA. Une partie puissance assure également une tension stabilisée d'alimentation du moteur. La seconde raison est la sécurisation de la partie commande en cas de dysfonctionnement de la partie puissance.

Alimentation

L'alimentation des diodes électroluminescentes ainsi que des cartes électroniques est assurée par une batterie 12V de capacité 12Ah. À raison de 700mA de courant pour les LEDs, 1A pour le moteur et quelques milliampères pour la carte de puissance (la carte de commande étant alimentée par la liaison



Figure 2.2.4: Vue d'ensemble du système.

USB), soit environ 2A de consommation, la durée de fonctionnement continu est de 6 heures. Cette alimentation permet donc au moins une journée d'utilisation sachant que le moteur n'est pas alimenté continuellement. La conversion 12V/5V nécessaire pour l'alimentation de la carte de puissance et du moteur est assurée par un convertisseur 9.2-18V/5V.

L'ensemble constitué des deux parties présentées précédemment est intégré dans un châssis en aluminium muni de panneaux opaques (figure 2.2.4). Ces panneaux permettent une isolation de la scène considérée par rapport à l'illumination extérieure et contre les mouvements des éléments constituant la scène dus au vent lors de la procédure d'acquisition. La partie supérieure du châssis est intégrable à la partie inférieure, munie de poignées, afin de faciliter le transport et diminuer l'encombrement du système.

2.2.4 Contraintes du système expérimental

Lors du mouvement, les vibrations créées par le déplacement de la partie optique entraînent un léger déplacement du centre optique entre chaque image de la séquence. Ceci est problématique puisque, dans le cadre de la reconstruction 3D par « Shape from Focus », chaque pixel doit correspondre au même point de la scène dans chacune des images de la séquence. La séquence d'images nécessite donc un recalage afin d'obtenir la correspondance entre les pixels. Le recalage d'images est un problème récurrent en imagerie. Il consiste essentiellement à établir une relation géométrique entre les représentations de deux images permettant une mise en correspondance de celles-ci. Cette relation géométrique peut être de type rigide, dans le cas d'objets non-déformables, ou non-rigide dans le cas contraire. Globalement, les méthodes de recalage sont réparties en deux familles : les méthodes géométriques et iconiques.

Les méthodes géométriques se basent sur l'identification des caractéristiques géométriques communes aux images à recaler. Ces caractéristiques sont souvent des primitives de l'image comme les arrêtes ou des coins dont les coordonnées spatiales sont appariées afin de déterminer la transformation entre les deux images.

Les méthodes iconiques consistent à optimiser un critère de ressemblance (mesure de similarité) fondé sur l'intensité des pixels des images, sans tenir compte de la géométrie de celles-ci. La transformation recherchée est celle pour laquelle la mesure de similarité est maximum. Ces méthodes sont souvent divisées en trois parties : le choix de la mesure de similarité à utiliser (somme des différences au carré, coefficient de corrélation, entropie mutuelle ...), le type de transformation recherchée (rigide, affine, projective ...) et une méthode d'optimisation assurant le lien entre les deux premières étapes.

Le domaine du recalage d'images est un domaine de recherche très vaste et couvrant plusieurs domaines d'application. L'état de l'art est complexe du fait du nombre de méthodes existantes et ne présente pas d'intérêt dans le cadre de ces travaux. Cependant, afin de choisir une méthode adaptée à notre problématique, il convient tout d'abord de définir quelle famille de techniques

de recalage convient. Dans notre cas, chacune des images de la séquence met en évidence une zone nette différente du fait du balayage par la zone de profondeur de champ. Ainsi, les méthodes de recalage géométriques basées sur une mise en correspondance de points d'intérêt ne sont pas adaptées. Les méthodes iconiques basées sur l'information mutuelle entre les images semblent beaucoup plus adaptées. Notre type de séquence d'images présente plusieurs similitudes avec des séquences d'images de type IRM. Ainsi, on retrouve la méthode basée sur l'information mutuelle de [Thévenaz et al., 1998] souvent utilisée en imagerie médicale. L'intérêt principal de cette technique est sa rapidité et sa convergence robuste vers l'estimation de la transformation liant chaque image comme le précisent les études de [Zitova et Flusser, 2003, Ghollipour et al., 2007].

2.3 Système final

Ce premier système ayant permis la validation de la faisabilité de l'utilisation de la méthode de reconstruction 3D « Shape from Focus » (des exemples de scènes reconstruites se trouvent chapitre 4.1.1), la conception d'un second système entraînant l'achat de matériel spécifique a été effectuée.

Afin de s'affranchir du déplacement de la partie optique entraînant une légère translation du centre optique, on s'est orientés vers le balayage de la scène par la variation de la distance de mise au point. Comme expliqué à la sous-section 2.1.2, une variation de cette distance induit une modification de la profondeur de champ. Cependant, tant que la profondeur de champ est plus faible que la précision de l'estimation de la profondeur recherchée, ce type de déplacement est parfaitement utilisable. Le détail de l'influence des différentes variations de mise au point est expliqué sous-section 2.4.3.

De façon à réaliser cette variation de la distance de mise au point précisément et surtout de manière répétable, l'utilisation d'un objectif motorisé est nécessaire. Cependant, un objectif avec une bague de réglage de la distance de mise au point motorisée n'est pas fréquent contrairement à ceux avec ouverture et focale motorisées. Le fabricant Qioptiq propose des objectifs conçus pour

des capteurs à hautes définitions ainsi qu'une version motorisée de ces objectifs. Ainsi, nous avons opté pour un objectif de focale 35 mm et d'ouvertures relatives $f1.6 - 16$ (Qioptiq MeVis-Cm figure 2.3.1). Cet objectif de monture C assure une correction linéaire de 450 nm à 900 nm et est compatible pour des capteurs ayant une diagonale maximum de 1". Les variations de l'ouverture et de la mise au point sont motorisées et la commande passe par l'utilisation d'un module de contrôle. Ce module de contrôle permet la commande de l'objectif par une liaison RS-232 avec l'ordinateur. Un jeu d'instructions simples permet la variation de la position de la bague de réglage de l'ouverture et de mise au point. Dans notre cas, la profondeur de champ souhaitée étant la plus faible possible, on veillera à toujours avoir le diaphragme ouvert au maximum (soit à l'ouverture relative $f1.6$).

Toujours dans l'optique d'une profondeur de champ réduite, le choix de la caméra (IDS UI-2280SE figure 2.3.1) s'est porté sur un capteur CCD couleur à filtre de Bayer 2/3 pouces (8.446×7.066 mm) de définition maximale 2448 par 2048 (5 mégapixels) dont les photo-éléments mesurent $3.45 \mu\text{m}^2$. Cette petite taille de pixel occasionne un cercle de confusion très faible ayant pour effet une diminution de la profondeur de champ. La taille de capteur supérieure comparée au premier système induit également une augmentation du champ de vue disponible de la scène considérée. La liaison de la caméra avec l'ordinateur est assurée par protocole USB.



Figure 2.3.1: Objectif et caméra utilisés.

A une distance de mise au point de 1 mètre, cet ensemble objectif/caméra génère une profondeur de champ de 8.7 mm et un champ de vue de 25 par 19

cm. Un autre avantage de ce second système est le contrôle direct de la caméra et de l'objectif par l'ordinateur sans avoir de partie électronique conséquente. Ceci permet par ailleurs une réduction de l'encombrement général du système d'acquisition.

2.4 Calibrage

Permettre une variation de plusieurs paramètres étant la fonction principale des deux systèmes d'acquisitions conçus, ceci entraîne invariablement un besoin de correction. De plus, aucun objectif n'étant parfait, un calibrage de celui-ci permet une correction des éventuelles distorsions. Les deux premières sous-sections suivantes présentent les corrections communes aux deux systèmes. La troisième sous-section est dédiée au système final puisque présentant l'étalonnage de l'objectif motorisé.

2.4.1 Correction des aberrations

Les aberrations occasionnées par l'objectif utilisé peuvent être géométriques ou chromatiques. Parmi les aberrations géométriques on retrouve :

- les aberrations sphériques : ce type d'aberration survient lorsque les rayons passant par les bords de l'objectif ne se focalisent pas à la même distance que ceux traversant l'optique au centre. Il en résulte une perte de netteté. Ce type d'aberration est corrigé par l'utilisation de lentilles asphériques dans la conception de l'objectif.
- la coma : ce défaut apparaît pour des points de la scène non alignés avec l'axe optique qui seront focalisés sous forme de comète. La conception d'objectif avec plusieurs groupes de lentilles permet de compenser ce défaut.
- l'astigmatisme : proche de la coma, ce défaut a pour effet une focalisation différente des rayons dans le plan sagittal de l'axe optique (focale sagittale) et ceux dans le plan tangentiel (focale tangentielle). Les objectifs actuels sont exempts de ce défaut et sont appelés anastigmats.
- la courbure de champ : ceci correspond à une projection d'un plan de la

scène sous forme parabolique sur le capteur. Comme l'astigmatisme, ce défaut est corrigé par les objectifs actuels.

- les distorsions : on retrouve les distorsions tangentielles et les distorsions radiales. Les distorsions tangentielles apparaissent lors d'un mauvais alignement des différentes lentilles composant l'objectif. Ces distorsions ne sont donc pas présentes pour des objectifs de précision. Les distorsions radiales peuvent être de deux types : en barillet ou en coussinet. La première a pour effet une transformation des lignes en courbes incurvées vers l'extérieur de l'image. La seconde est l'inverse, avec des courbes incurvées vers l'intérieur de l'image. La distorsion en barillet est de plus en plus présente à mesure que la focale est courte. Ainsi, un objectif de type « fisheye » aura de très fortes distorsions. Inversement, les téléobjectifs de grande focale font état de distorsions en coussinet. Même si ces phénomènes sont en partie corrigés par l'utilisation de lentilles asphériques, un traitement *a posteriori* est nécessaire pour les corriger.

Les aberrations chromatiques correspondent quant à elles à la différence de focalisation des rayons lumineux de longueurs d'onde différentes. En effet, comme expliqué section 2.1.1, un rayon lumineux changeant de milieu est réfracté suivant un angle dépendant de l'indice de réfraction du milieu. Cependant, cet indice de réfraction dépend également de la longueur d'onde à laquelle correspond ce rayon incident puisqu'il augmente lorsque la longueur d'onde diminue. Ainsi, un rayon de longueur d'onde 400 nm (bleu) aura un indice de réfraction plus élevé qu'un rayon de longueur d'onde 650 nm (rouge). C'est d'ailleurs sur ce principe que se base l'utilisation d'un prisme permettant la dispersion des longueurs d'onde.

Notre objectif étant un objectif de précision spécialement conçu pour des applications de mesure, il permet une élimination d'un certain nombre d'aberrations ou tout du moins une atténuation de leur impact. Ainsi, les aberrations chromatiques sont corrigées par la correction linéaire de 450 à 900 nm qui permet une utilisation de l'objectif dans le visible et le très proche infrarouge sans crainte d'aberrations. Concernant les aberrations géométriques, elles sont corrigées par la conception particulière de l'objectif mais, concernant les distorsions éventuelles, il convient de vérifier si un traitement *a posteriori*

est nécessaire. Précisons que, outre les distorsions, les autres aberrations sont atténuées à mesure que le diamètre du diaphragme diminue. Cependant, souhaitant obtenir une profondeur de champ la plus faible possible, une ouverture maximum du diaphragme est nécessaire.

Pour cela, l'outil de calibrage développé par [Bouguet, 2010] pour le logiciel *Matlab* est utilisé. Cet outil implémente la méthode de calibrage proposée par [Zhang, 2000]. La procédure de calibrage commence par l'acquisition de plusieurs images d'une mire (figure 2.4.1) en faisant varier la position de celle-ci.

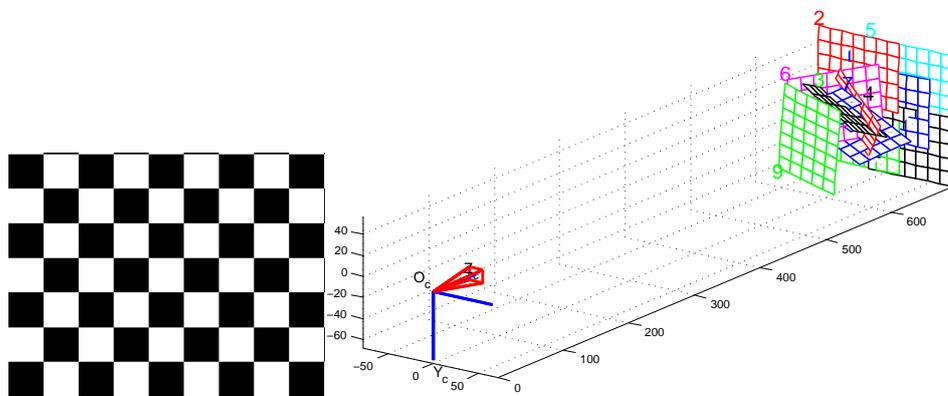


Figure 2.4.1: Mire de calibrage et projection des positions utilisées.

Les points d'intérêt des images sont détectés puis les paramètres intrinsèques (focale, centre optique) et extrinsèques (position repère monde de la caméra par rapport à la scène) sont déterminés. Les coefficients de distorsions radiales et tangentiels sont ensuite estimés avant une optimisation de l'estimation des différents paramètres par minimisation de la fonction de maximum de vraisemblance.

Le vecteur estimé contenant les coefficients de distorsions se présente sous la forme $k_c = [k_1 \ k_2 \ p_1 \ p_2 \ k_3]$ avec k_1 , k_2 et k_3 les coefficients de distorsions radiales et p_1 , p_2 les coefficients de distorsions tangentiels. On ne considèrera pas les coefficients de distorsions tangentiels dans notre calibration.

Le modèle de distorsions radiales du second ordre utilise seulement le coefficient k_1 pour estimer les nouvelles coordonnées de l'image sans distorsions [Tsai, 1987]. Zhang préconise l'utilisation du modèle de distorsions radiales du

4^{ème} ordre utilisant les deux premiers coefficients. En pratique, le modèle du 6^{ème} ordre est rarement utilisé donc nous ne considérerons pas le coefficient k_3 pour le calibrage.

Une fois le calibrage réalisé pour nos deux objectifs (50 mm et 35 mm motorisé), les modélisations des distorsions radiales sont obtenues et illustrées figure 2.4.2.

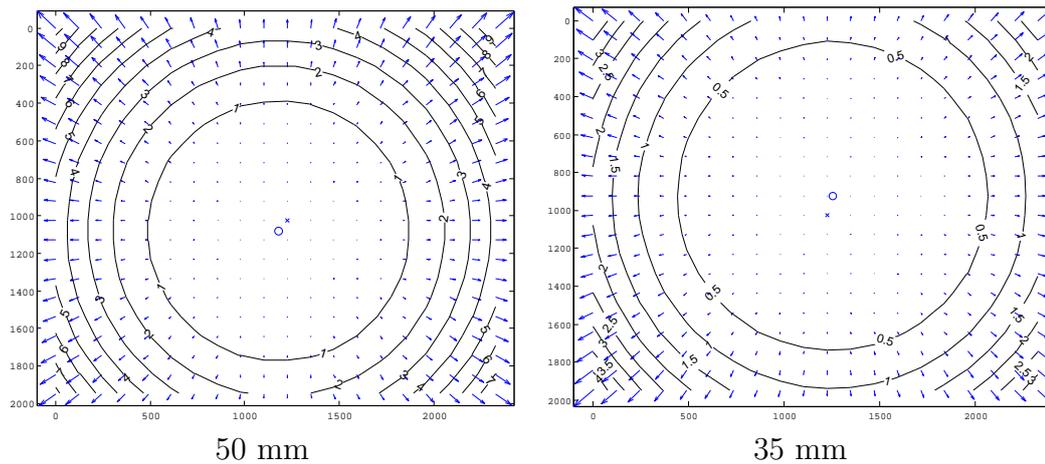


Figure 2.4.2: Distorsions radiales des objectifs utilisés.

On constate qu'une légère distorsion est présente sur les deux objectifs. Ainsi, on a 9 pixels de décalage maximum en bordure d'image pour l'objectif 50 mm et 4 pixels pour le 35 mm. Contrairement à ce qu'on aurait pu envisager, l'objectif 35 mm fait état de moins de distorsions que le 50 mm alors qu'il est de focale plus faible. L'objectif du second système est donc de meilleure facture que l'objectif précédemment utilisé.

2.4.2 Correction du grossissement

Comme expliqué précédemment, le fait de déplacer le plan focal, que ce soit par une translation de l'ensemble de la partie optique ou par la variation de la distance de mise au point, entraîne un effet de grossissement au cours de l'acquisition. Ce grossissement induit plusieurs effets indésirables :

- les parties de la scène situées près des bords sortent progressivement du champ de vue au fur-et-à-mesure du balayage. Une diminution du champ

de vue apparaît ;

- la zone locale utilisée dans chaque image pour la mesure de netteté d'un pixel ne représente pas les mêmes informations spatiales.

En pratique, ce grossissement peut être négligé pour des déplacements très courts comme c'est le cas en microscopie. Cependant, l'aspect macroscopique de notre application nécessite une correction de ce grossissement. Si aucune correction n'est apportée à la séquence d'images, les mesures de netteté biaisées entraînent une reconstruction 3D erronée de la scène.

On retrouve dans la littérature plusieurs solutions pour corriger ou s'affranchir de ce grossissement.

Redimensionnement des images

Proposé par [Darrell et Wohn, 1988], le principe est de réaliser la mise à l'échelle des images *a posteriori* de l'acquisition. Bien que cette méthode soit simple à mettre en œuvre et adaptée aux objectifs à focale fixe, elle a des inconvénients. Du fait du changement d'échelle appliqué aux images nécessitant une interpolation et un rééchantillonnage, une perte d'information apparaît. Cependant, une même zone autour d'un pixel représente bien la même information spatiale pour la mesure de netteté.

Compensation du grossissement

D'abord proposée par Darell mais non implémentée par celui-ci, cette méthode est reprise et utilisée par [Willson et Shafer, 1991]. Leur solution est de compenser le grossissement, résultant du déplacement, par un changement de la distance focale dans le but de toujours visualiser la même zone spatiale. Pour cela, un modèle calibré du comportement du changement de la focale (zoom) en fonction du déplacement (du système ou dans leur cas de la distance de mise au point) est réalisé (figure 2.4.3).

L'inconvénient majeur de cette solution est le besoin d'un objectif à focale variable motorisée qui entraîne par ailleurs une modification de la profondeur de champ due au changement de focale. La profondeur de champ est également modifiée indirectement lors du changement de focale car l'ouverture n'est pas

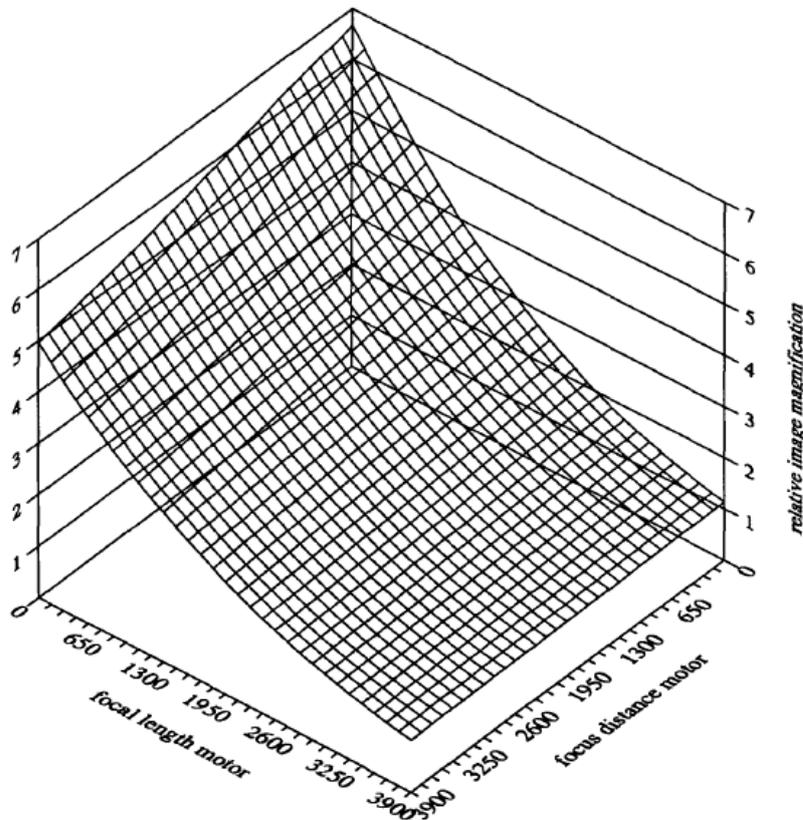


Figure 2.4.3: Grossissement relatif en fonction de la focale et de la distance de mise au point [Willson et Shafer, 1991].

identique pour toutes les variations de focale de ce type d'objectif. Ceci est dû au fait que la quantité de lumière passant par l'objectif n'est pas la même en fonction de la focale utilisée. Une adaptation de l'ouverture est donc nécessaire afin d'obtenir une quantité de lumière constante passant par l'objectif.

Utilisation d'un objectif télécentrique

[Watanabe et Nayar, 1995] proposent de modifier mécaniquement un objectif en lui associant un second système de diaphragme positionné au niveau du plan focal. L'adjonction de cette ouverture permet l'obtention du procédé télécentrique. En effet, si le diaphragme d'ouverture se situe dans le plan focal du système optique qui le précède, les rayons lumineux passant par l'objectif et focalisés sur le capteur peuvent être considérés comme provenant de

l'infini et donc parallèles les uns aux autres. Ce type d'objectif permet donc une taille d'image identique à mesure du déplacement sous réserve de rester dans une certaine profondeur. L'effet de grossissement n'est plus présent, de même que les distorsions radiales qui sont quasi-nulles avec ce type d'objectif. Cependant, la scène visualisée par ces objectifs est souvent petite (quelques centimètres carrés) et ces objectifs induisent une profondeur de champ un peu plus élevée que les objectifs standards.

Adaptation de la fenêtre de mesure

Puisque qu'une variation du grossissement de la scène visualisée apparaît lors de l'acquisition, [Nair et Stewart, 1992] proposent d'adapter la taille de la fenêtre de mesure de netteté locale en fonction du changement de résolution spatiale. Ainsi, la mesure de netteté est toujours appliquée à la même zone spatiale pour chaque pixel. L'inconvénient est le nombre d'échantillons non constant pour la mesure de netteté qui ne permet pas de comparer celles-ci de façon rigoureuse. De plus, une augmentation de la taille de la fenêtre d'analyse a pour effet une diminution de la résolution spatiale et de la profondeur bien qu'étant plus robuste quant à l'évaluation de la zone la plus nette. Donc une modification de la taille de la fenêtre d'analyse ne permet pas d'avoir une carte de profondeur homogène au niveau de sa résolution.

Dans notre cas, la correction du grossissement passe par une méthode à l'interface du redimensionnement des images et l'adaptation de la fenêtre de mesure. Tout d'abord l'estimation du grossissement est aisée du fait que le déplacement entre deux images est connu précisément (Δd dans le schéma illustré figure 2.4.4).

Le coefficient de grossissement h_n/h_1 peut donc être déterminé par le théorème de Thalès :

$$\frac{h_n}{h_1} = \frac{d}{d + \Delta d}. \quad (2.4.1)$$

Ce coefficient est calculé pour chaque image et est utilisé pour la redimensionner afin de toujours visualiser la même zone spatiale. Une mise à l'échelle de toutes les images est ensuite réalisée en appliquant un redimen-

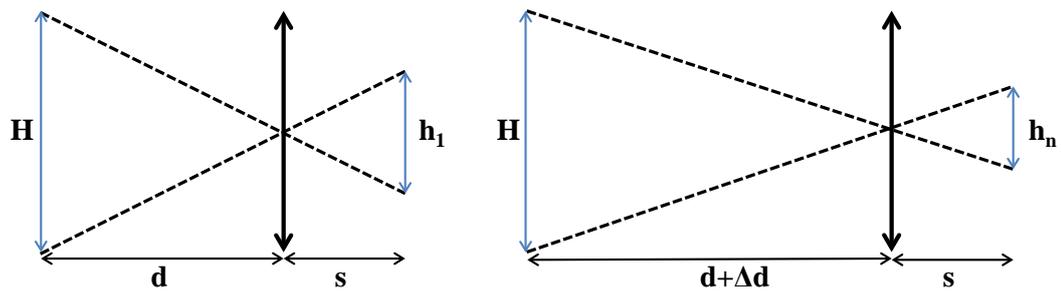


Figure 2.4.4: Schématisation du grossissement entre la 1^{ère} et la $n^{\text{ième}}$ image.

sionnement à chaque image. Ce redimensionnement est calqué en fonction de la taille de l'image la plus éloignée du capteur. Ainsi, une sous-interpolation des autres images est appliquée. L'information est certes dégradée mais appliquer un redimensionnement dans ce sens ne génère pas d'informations erronées contrairement à une sur-interpolation. La définition du capteur étant de base élevée, une réduction de la taille des images implique toujours une bonne définition. On obtient de ce fait une séquence d'images représentant la même scène avec un déplacement de la zone nette et permettant une mesure de netteté sur les mêmes zones spatiales.

2.4.3 Étalonnage de l'objectif motorisé

Contrairement au premier système d'acquisition qui permet un pas constant entre chaque image comme, par exemple, tous les centimètres, l'objectif motorisé du second système nécessite un étalonnage afin de maîtriser la commande. En effet, une rotation à intervalle constant de la bague de variation de distance de mise au point n'entraîne pas une variation constante de cette distance dans l'espace. Cette variation doit être cependant parfaitement maîtrisée afin de garantir une reconstruction 3D fidèle à la réalité. Effectivement, l'étape de mesure de netteté permet de déterminer la position de l'image où chaque pixel est le plus net et donc le pas séparant chaque image de la séquence doit être connu.

L'objectif motorisé autorise une variation de la distance sur 94 positions de la distance de mise au point minimum (45 mm selon constructeur) à l'infini.

Pour notre application, nous souhaitons balayer la scène considérée de la distance de mise au point la plus proche jusqu'à une distance de 1 mètre. Ainsi, pour chaque position du moteur, la distance de mise au point correspondante est déterminée et représentée figure 2.4.5.

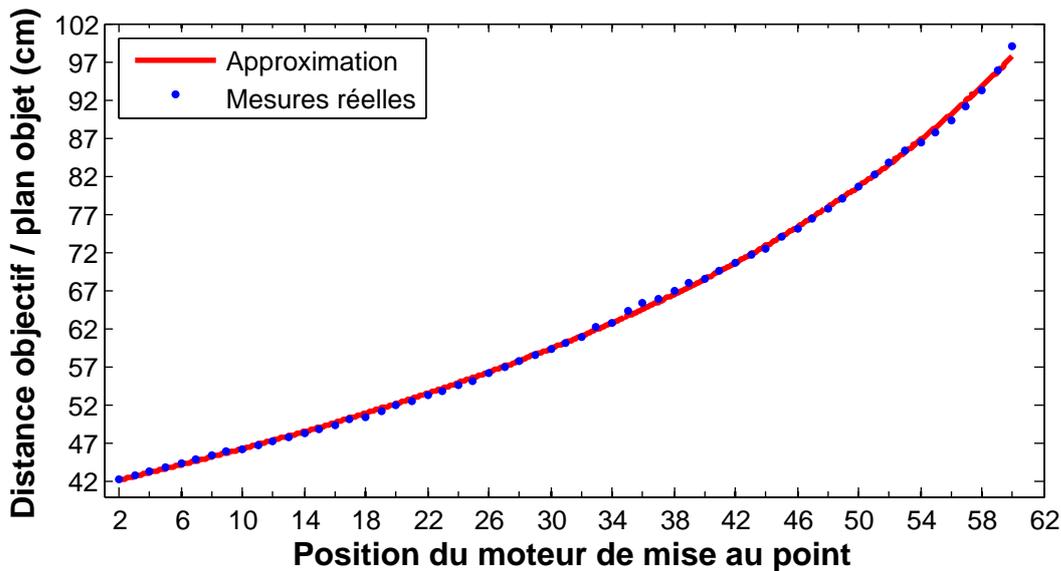


Figure 2.4.5: Distance de mise au point en fonction de la position du moteur.

Pour déterminer avec précision à quelle distance exacte correspond une position moteur, un opérateur de mesure de netteté (chapitre 3) est utilisé afin de mesurer précisément cette distance. Ainsi, pour chaque position moteur on obtient une courbe analogue à celle illustrée figure 2.4.6.

On remarque effectivement une non-linéarité entre l'évolution de la distance de mise au point en fonction de la position du moteur. De plus, après mesure, la distance minimum de mise au point est de 42.1 cm et non 45 cm.

Suite à l'étalonnage de l'objectif, il est intéressant de calculer la profondeur de champ (eq. 2.1.6) associée à chaque distance de mise au point. L'évolution de celle-ci est représentée figure 2.4.7.

Avec un balayage de la scène inférieur à 1 mètre, la profondeur de champ n'excède pas 7.5 mm. Ceci permet une précision de l'estimation de la profondeur par la méthode « Shape from Focus » car cela garantit un non recouvrement des mêmes zones nettes pour deux images successives. En effet,

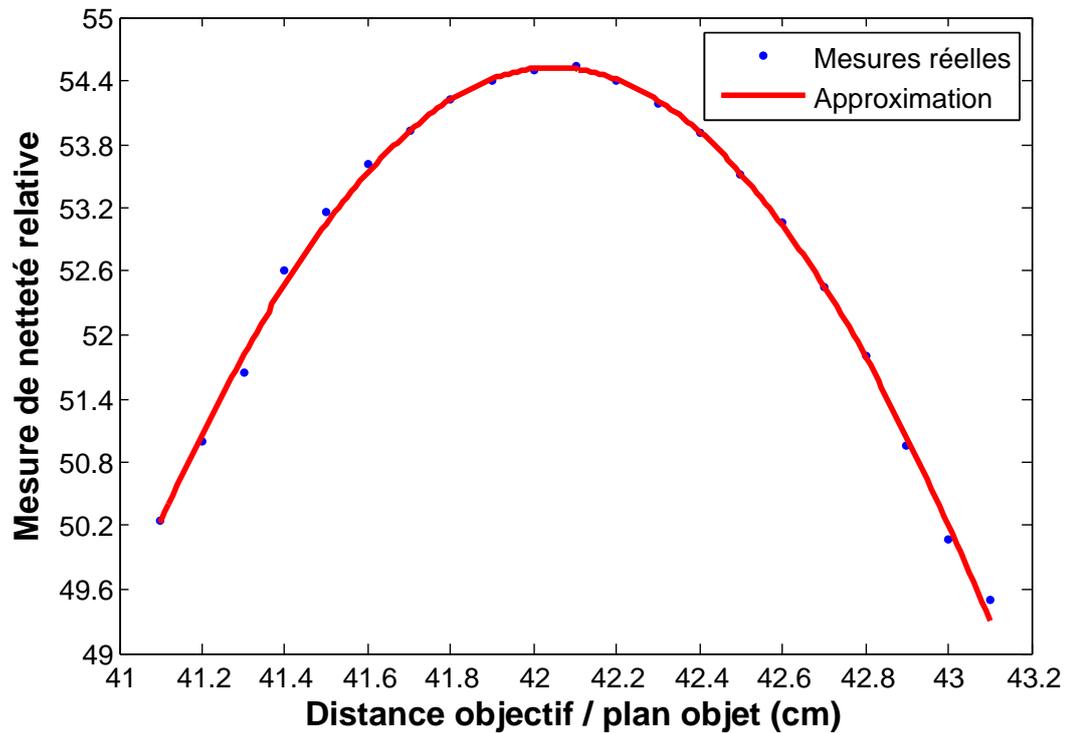


Figure 2.4.6: Mesure de netteté pour la 1^{ère} position moteur.

l'espacement entre deux images est contrôlé de façon à être toujours supérieur à la profondeur de champ.

Quant au champ de vue résultant d'un balayage de 42 cm jusqu'à 100 cm de la scène, il varie respectivement de 9.5 par 8 cm (42 cm) jusqu'à 22.3 par 18.6 cm (100 cm).

2.5 Procédure d'acquisition

La procédure d'acquisition va dépendre principalement du type de scène à visualiser ainsi que de la précision de la reconstruction souhaitée. Ainsi le logiciel développé pour le contrôle du premier et du second système permet de choisir le pas d'acquisition voulu, la distance de départ du balayage ainsi que le nombre d'images à acquérir. Le choix du pas entre chaque acquisition va être choisi en fonction de la profondeur de champ engendrée par le système.

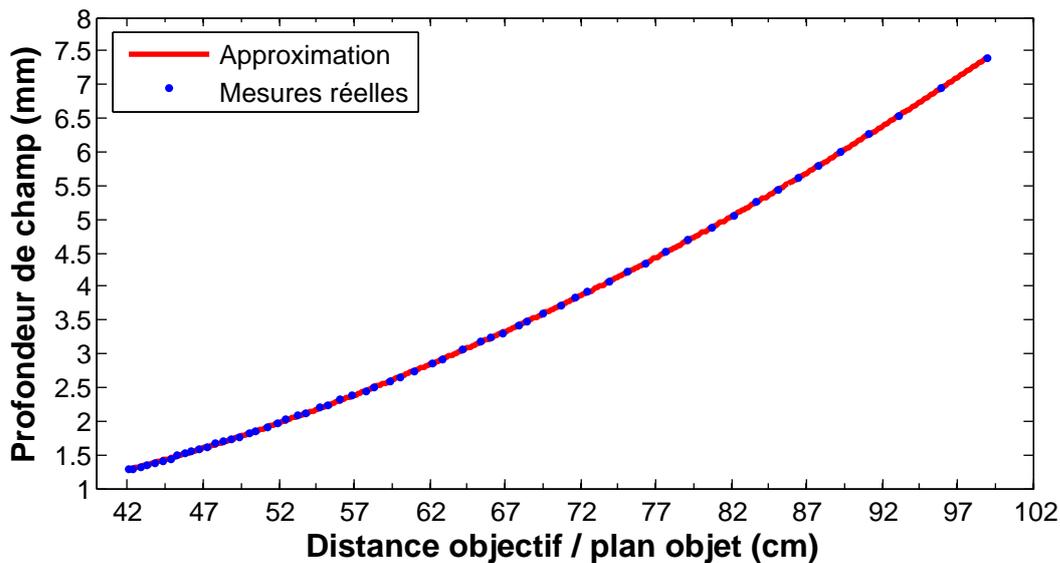


Figure 2.4.7: Profondeur de champ en fonction de la distance de mise au point.

En effet, pour ne pas avoir d'effet de chevauchement des zones dans le plan de netteté, le pas doit être choisi au minimum égal à la profondeur de champ. Par exemple, pour une profondeur de champ de 1 cm, le déplacement minimal entre deux images doit être de 1 cm. La distance de départ du balayage permet de placer le premier plan de balayage au début de la scène à reconstruire. Enfin, le nombre d'images à acquérir va évidemment dépendre de la profondeur de reconstruction souhaitée de la scène considérée. Une fois l'ensemble des paramètres choisis par l'utilisateur, l'acquisition est entièrement automatisée.

2.6 Conclusion

Dans ce chapitre, les deux systèmes permettant l'acquisition de la séquence d'images nécessaire à une reconstruction 3D par la méthode « Shape from Focus » sont présentés. La compréhension des phénomènes optiques induits par les caméras et objectifs utilisés est le point central de nos choix quant au dimensionnement des deux systèmes d'acquisition. L'automatisation de l'acquisition permet une utilisation aisée des deux systèmes par des non spécialistes. De plus, les deux systèmes sont complémentaires en fonction de l'application

visée. En effet, le premier système, bien qu'étant sujet à des erreurs d'alignement des images acquises, permet un changement d'objectif en fonction de l'application visée. Le second système quant à lui, s'affranchit des erreurs d'alignement et, du fait de son encombrement réduit, permet son intégration dans des appareils mobiles ou espaces confinés.

Le chapitre suivant est consacré à l'évaluation de la netteté des pixels de chaque image de la séquence. Après l'acquisition, la mesure de netteté est la seconde étape très importante pour garantir une reconstruction 3D fidèle de la scène analysée.

Chapitre 3

Evaluation de la netteté

La notion de netteté correspond à une focalisation complète de l'énergie lumineuse réfléchiée par un point de la scène sur le plan capteur. Il existe en réalité une zone considérée comme nette appelée profondeur de champ introduite section 2.1.2. Cette zone nette est donc à mettre en correspondance avec l'objet ou la partie de la scène à acquérir. Pour faire un parallèle avec la vision humaine, cette focalisation sur la zone visualisée, et donc l'évaluation de la netteté permettant de juger cette zone comme la plus nette, est réalisée automatiquement par le processus physiologique d'accommodation. Cependant, en vision par ordinateur, ce processus est à recréer afin d'évaluer cette netteté de façon automatique sans dépendre de l'expertise subjective de l'utilisateur. Un des exemples les plus courants de la mise au point automatique d'un système d'acquisition est l'autofocus, terme anglo-saxon, présent sur les appareils photos, qui ajuste automatiquement la bague de mise au point jusqu'à maximiser la netteté sur le point de la scène visé. Une valeur de netteté est attribuée à chaque changement du paramètre optique et son maximum correspond la distance de mise au point assurant la netteté.

Ce processus de mise au point automatique, qui a évolué au cours du temps, se divise en plusieurs procédés répartis en deux catégories :

- Procédés actifs : utilisation d'un système télémétrique associé permettant de mesurer la distance séparant le point de la scène visé de l'appareil. On retrouve l'utilisation de systèmes télémétriques à base d'ultrasons

ou de faisceaux infrarouges (principe similaire à celui des méthodes de reconstruction 3D, basées sur les mêmes signaux, décrites chapitre 1). Ce sont les premiers systèmes d'autofocus créés mais très peu utilisés de nos jours.

- Procédés passifs : ces systèmes d'autofocus se basent directement sur l'information contenue dans la scène visualisée pour évaluer la netteté. On retrouve les autofocus à détection de phase qui utilisent plusieurs microcapteurs CCD qui vont recevoir l'énergie lumineuse de la scène par un système de condenseur et lentilles de focalisation. En fonction de la réponse de ces microcapteurs, le sens de rotation de la bague de mise au point ainsi que l'amplitude du déplacement sont simultanément déterminés pour permettre une mise au point correcte [Kusada, 1996]. Ce système est celui couramment utilisé sur les appareils reflex pour leur rapidité et leur qualité d'estimation. Un autre système d'autofocus passif se base sur la mesure de contraste des pixels adjacents à différents emplacements de la scène visualisée. Ainsi, plus la zone est contrastée, plus elle est nette. La mise au point est faite en continu jusqu'à maximiser la mesure de contraste. Cet autofocus se retrouve sur les appareils bas de gamme.

Dans notre cas, l'objectif n'est pas de converger vers la mise au point constituant l'image la plus nette mais de déterminer l'image pour laquelle chaque pixel est le plus net parmi les images de la séquence (figure 3.0.1). La position de l'image où le pixel considéré est le plus net renvoie directement à l'information de profondeur de celui-ci, puisque notre système est calibré (section 2.4.3). Une évaluation locale de la netteté est donc nécessaire. Au même titre que la mesure de contraste, de nombreuses méthodes permettant l'estimation de la netteté locale sont présentes dans la littérature. Dans ce chapitre, un état de l'art de ces méthodes est proposé ainsi que les méthodes d'approximation affinant l'estimation de la profondeur pour chaque pixel. Puis, la comparaison des différents opérateurs de mesure de netteté est effectuée par une méthode adaptée à notre problématique. Enfin, une méthode non supervisée et une méthode supervisée permettant le choix de l'opérateur à utiliser en fonction de la scène visualisée sont présentées.

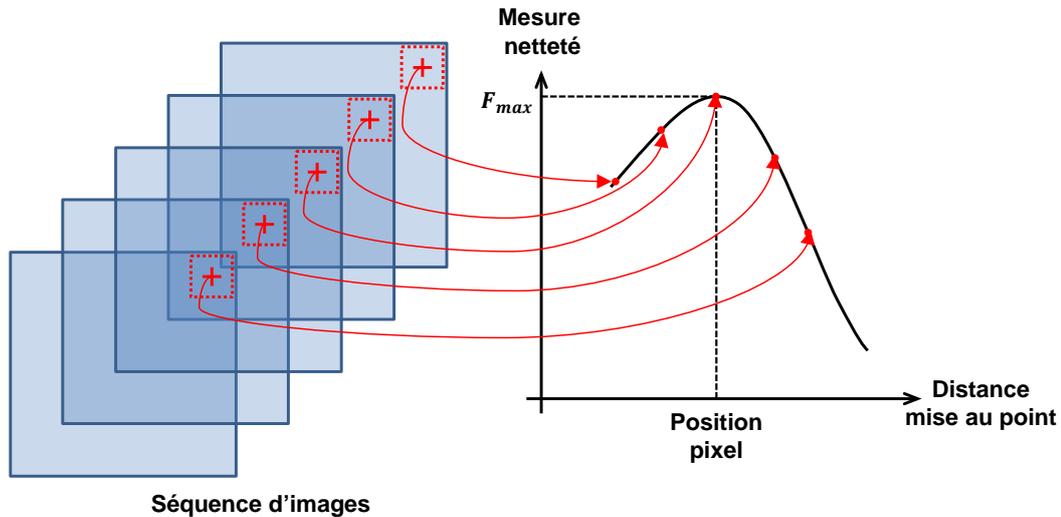


Figure 3.0.1: Schéma du principe de mesure de netteté d'un pixel de la séquence d'images.

3.1 État de l'art des opérateurs de mesure de netteté

On retrouve dans la littérature beaucoup d'opérateurs permettant l'évaluation de la netteté. A l'origine, ils étaient plutôt utilisés afin de déterminer la meilleure distance de mise au point en microscopie avec une évaluation de la netteté sur l'image totale. Ils sont depuis également au cœur de l'algorithme de reconstruction 3D « Shape from Focus » mais utilisés localement au voisinage des pixels. Globalement, les mesures de netteté se basent sur la maximisation de l'information haute fréquence puisqu'une image nette contient plus de changements d'intensité et de primitives à fort gradient (coins, arêtes ...) qu'une image floue. Bien évidemment, le principal inconvénient de cette méthode de reconstruction 3D est que son application est limitée aux scènes présentant des composantes hautes fréquences, donc aux scènes présentant des zones texturées. Dans le cas contraire, l'adjonction d'une lumière structurante est nécessaire [Noguchi et Nayar, 1994].

On retrouve dans l'état de l'art une répartition de ces opérateurs en quatre grandes familles : mesures différentielles, statistiques, fréquentielles et autres

types de mesures. Dans ces catégories, seul les opérateurs de mesure de netteté conservés pour la suite de l'étude sont détaillés. Ces opérateurs constituent un exemple représentatif des différentes familles d'opérateurs ainsi que ceux couramment utilisés et ayant fait leurs preuves. De plus, deux nouveaux opérateurs basés sur les informations fréquentielles sont décrits dans la section associée.

3.1.1 Mesures différentielles

Plus une image devient floue, plus le diamètre du cercle de la tache floue censée représenter un point de la scène est important. Ainsi, cette tache floue a pour effet de distribuer l'énergie d'un pixel sur un ensemble de pixels adjacents. Des fonctions mesurant la différenciation entre un pixel et ses voisins présentent donc un extrémum dans l'image pour laquelle le pixel est le mieux focalisé. Les fonctions mesurant cette différenciation sont nombreuses comme, par exemple, le gradient de Brenner [Brenner et al., 1976], l'énergie du Laplacien ou du gradient, mesures très utilisées et proposées par [Subbarao et al., 1993] :

$$FM_{nrjgrad}(i, j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} f_x^2 + f_y^2, \quad (3.1.1)$$

où $f_x = I(x+1, y) - I(x, y)$, $f_y = I(x, y+1) - I(x, y)$ (avec $I(x, y)$ la valeur du pixel à la position (x, y)). N représente la taille du voisinage considéré pour la mesure de netteté (il en est de même pour toutes les autres équations comprenant N) et (i, j) sont les coordonnées du pixel en cours de mesure.

Des variantes ont été proposées par [Ahmad et Choi, 2007] utilisant le gradient et [An et al., 2008] le Laplacien, mais en considérant non pas un voisinage 2D mais un voisinage 3D de chaque pixel. Les images précédentes et suivantes, par rapport à l'image en cours d'analyse, sont donc mises à contribution.

[Tenenbaum, 1970] propose d'estimer l'amplitude du gradient (eq. 3.1.4) à chaque pixel en additionnant les amplitudes de tous les pixels du voisinage supérieures à une valeur seuil T . Les opérateurs de Sobel sont utilisés comme

masque de convolution :

$$S_h = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad S_v = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}, \quad (3.1.2)$$

$$S(x, y) = \sqrt{[G_h(x, y)]^2 + [G_v(x, y)]^2}, \quad (3.1.3)$$

où $G_h(x, y)$ et $G_v(x, y)$ sont, respectivement, la convolution entre l'image I et les masques S_h et S_v aux coordonnées (x, y) . La mesure de netteté proposée appelée « Tenengrad » est donnée par :

$$FM_{ten}(i, j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} [S(x, y)]^2 \quad \text{avec } S(x, y) > T. \quad (3.1.4)$$

Une variante de cette mesure est suggérée par [Krotkov, 1987] pour laquelle aucune valeur seuil n'est utilisée. D'autre part, [Pech-Pacheco et al., 2000] calculent la variance de l'amplitude du gradient :

$$FM_{ten_var}(i, j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} [S(x, y) - \bar{S}]^2, \quad (3.1.5)$$

où $\bar{S} = \frac{1}{N^2} \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} S(x, y)$.

Dans le cadre du « Shape from Focus », un opérateur très utilisé est le Laplacien modifié présenté dans la publication originelle de la méthode de reconstruction 3D [Nayar, 1989]. Le Laplacien classique devient nul lorsque les deux dérivées secondes sont de même amplitude mais de signes opposés :

$$L(x, y) = \frac{\partial^2 I(x, y)}{\partial x^2} + \frac{\partial^2 I(x, y)}{\partial y^2}. \quad (3.1.6)$$

Le Laplacien modifié considère donc l'amplitude des deux dérivées secondes :

$$ML(x, y) = \left| \frac{\partial^2 I(x, y)}{\partial x^2} \right| + \left| \frac{\partial^2 I(x, y)}{\partial y^2} \right|. \quad (3.1.7)$$

L'approximation discrète du Laplacien modifié est définie comme la convolution du filtre Laplacien carré et le voisinage considéré :

$$ML(x, y) = \left| \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} * I(x, y) \right|. \quad (3.1.8)$$

La mesure de netteté correspondante est donc définie par l'équation 3.1.9 :

$$FM_{sml}(i, j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} ML(x, y) \quad \text{avec } ML(x, y) > T. \quad (3.1.9)$$

3.1.2 Mesures statistiques

Les mesures de contraste précédemment évoquées se trouvent dans cette famille de mesures. Ainsi, une des mesures de netteté les plus utilisées est basée sur la variance des niveaux de gris du voisinage [Groen et al., 1985] :

$$FM_{var}(i, j) = \frac{1}{N^2 - 1} \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} [I(x, y) - \bar{g}]^2, \quad (3.1.10)$$

où \bar{g} est la moyenne des niveaux de gris au voisinage du pixel $I(x, y)$.

Le contraste peut également être défini par la différence entre les niveaux de gris maximum et minimum. [Firestone et al., 1991] suggèrent d'utiliser cette différence comme mesure de netteté :

$$FM_{range}(i, j) = \max_{x,y} [I(x, y)] - \min_{x,y} [I(x, y)], \quad (3.1.11)$$

où $x = i - N, \dots, i + N$ et $y = j - N, \dots, j + N$.

Partant du principe qu'une image nette détient plus d'informations qu'une image floue, [Krotkov, 1987] évalue la netteté par la mesure de l'entropie dans

le voisinage du pixel étudié :

$$FM_{entropy} = - \sum_k p_k \log_2 p_k, \quad (3.1.12)$$

où p_k représente la fréquence relative du niveau de gris de valeur k .

Notons qu'une mesure de netteté statistique basée sur une analyse en composantes principales du voisinage 3D du pixel considéré est proposée par [Mahmood et Choi, 2008].

3.1.3 Mesures fréquentielles

Une image nette présente plus de composantes hautes fréquences qu'une image floue. L'étude de l'information fréquentielle du voisinage de chaque pixel est le fondement de plusieurs mesures de netteté. De ce fait, nombre d'auteurs proposent des mesures de netteté basées sur différentes transformées telles que la transformée de Fourier [Malik et Choi, 2008], la transformée en cosinus discrète [Shen et Chen, 2006, Kristan et al., 2006] ou encore la transformée en ondelettes discrètes [Yang et Nelson, 2003].

Lorsqu'une image devient floue, l'énergie des composantes hautes fréquences diminue alors que celle des basses fréquences augmente. C'est sur ce constat que se basent [Xie et al., 2006] qui suggèrent l'utilisation, comme mesure de netteté, du ratio entre la somme des coefficients hautes fréquences des moments d'ordre deux (eq. 3.1.13) et des coefficients basses fréquences 3.1.14) de la transformée en ondelettes discrètes (eq. 3.1.15) :

$$M_H^2 = \sum_{I=1}^K \left[\sum_{(x,y) \in S_{LHI}} W_{LHI}^2(x,y) + \sum_{(x,y) \in S_{HLI}} W_{HLI}^2(x,y) + \sum_{(x,y) \in S_{HHI}} W_{HHI}^2(x,y) \right], \quad (3.1.13)$$

$$M_L^2 = \sum_{(x,y) \in S_{LL}} W_{LLK}^2(x,y), \quad (3.1.14)$$

$$FM_{wavratio} = \frac{M_H^2}{M_L^2}, \quad (3.1.15)$$

où W_{LHI} , W_{HLI} et W_{HHI} sont les coefficients hautes fréquences des sous-bandes S_{LHI} , S_{HLI} et S_{HHI} (respectivement filtrage horizontal, vertical et diagonal). W_{LLK} est le coefficient basses fréquences de la sous-bande S_{LL} pour le $K^{ième}$ niveau d'ondelette. Après comparatifs, les auteurs utilisent les ondelettes de Daubechies d'ordre 6 et 3 étapes de décomposition. Comparée aux autres mesures de netteté, la complexité de celle-ci étant plus élevée, le temps de calcul l'est également, mais d'après [Shen et Chen, 2006] cette mesure permet une meilleure efficacité sur les zones peu contrastées.

3.1.4 Autres types de mesures

On retrouve enfin des opérateurs non classables dans les autres catégories. C'est le cas des opérateurs basés sur une mesure de corrélation comme celle suggérée par [Vollath, 1987] :

$$FM_{corr}(i, j) = \sum_{x=i-N}^{i+(N-1)} \sum_{y=j-N}^{j+N} I(x, y) \cdot I(x+1, y) - \sum_{x=i-N}^{i+(N-2)} \sum_{y=j-N}^{j+N} I(x, y) \cdot I(x+2, y). \quad (3.1.16)$$

Cette mesure de netteté s'appuie sur le fait que la variabilité de l'information est plus importante pour une image nette que pour une image floue même pour un léger décalage du voisinage. [Helmlí et Scherer, 2001] suggèrent de mesurer le contraste local en calculant le ratio entre l'intensité du pixel mesuré et la moyenne des pixels avoisinants. Le ratio sera donc plus important pour un voisinage contrasté. Récemment, [Minhas et al., 2009] ont proposé une mesure de netteté basée sur des filtres orientables dans le cadre de la reconstruction 3D par « Shape from Focus ». Ils utilisent la fonction gaussienne $h(x, y)$ (eq. 2.1.4) dont les dérivées partielles constituent les filtres orientables de base G_x et G_y (eq. 3.1.17 et 3.1.18, respectivement 0 et 90 degrés) :

$$G_x = \frac{1}{\sigma^2}(-x \times h(x, y)), \quad (3.1.17)$$

$$G_y = \frac{1}{\sigma^2}(-y \times h(x, y)). \quad (3.1.18)$$

Le jeu de filtres orientables R est ensuite obtenu par les rotations θ de ces

filtres de base convolués au voisinage considéré :

$$R_x(x, y) = G_x * I(x, y), \quad (3.1.19)$$

$$R_y(x, y) = G_y * I(x, y), \quad (3.1.20)$$

$$R_\theta(x, y) = \cos(\theta).R_x + \sin(\theta).R_y, \quad (3.1.21)$$

où $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$. La mesure de netteté est obtenue en évaluant l'amplitude maximum de chaque filtre et le pixel dont l'amplitude est la plus élevée est considéré net :

$$FM_{orient} = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} \max_{\theta}(R_{\theta}(x, y)). \quad (3.1.22)$$

3.2 Mesure de netteté basée sur les descripteurs généralisés de Fourier

Comme le type de scène considéré pour nos applications agronomiques est souvent texturé, une mesure de netteté basée sur l'analyse de cette texture semble pertinente. Parmi les opérateurs d'analyse de texture couramment utilisés, on retrouve les 14 paramètres d'Haralick [Haralick et al., 1973] qui, pour certains, correspondent aux mesures de netteté présentées précédemment (entropie, variance, corrélation). L'analyse de netteté ne passe cependant pas par l'application de ces paramètres sur une matrice de cooccurrence mais sur le voisinage directement.

Récemment, [Journaux et al., 2011] ont utilisé les descripteurs généralisés de Fourier (DGF) dans le cadre d'analyses de rugosité de plantes pour la pulvérisation de précision en agronomie. Ces descripteurs ont été introduits par [Gauthier et al., 1991] et appliqués à des problématiques de reconnaissance de formes [Smach et al., 2008]. En considérant la fonction f , la transformée de Fourier \hat{f} de cette fonction est définie telle que :

$$\hat{f}(\lambda, \theta) = \int_{\mathbb{R}^2} f(x, y) e^{-j(\lambda x + \theta y)} dx dy, \quad (3.2.1)$$

où $\hat{f}(\lambda, \theta)$ représente la transformée de Fourier de f au point de coordonnées polaires (λ, θ) . Le descripteur généralisé de Fourier est défini tel que :

$$D_f(\lambda) = \int_0^{2\pi} |\hat{f}(\lambda, \theta)|^2 d\theta. \quad (3.2.2)$$

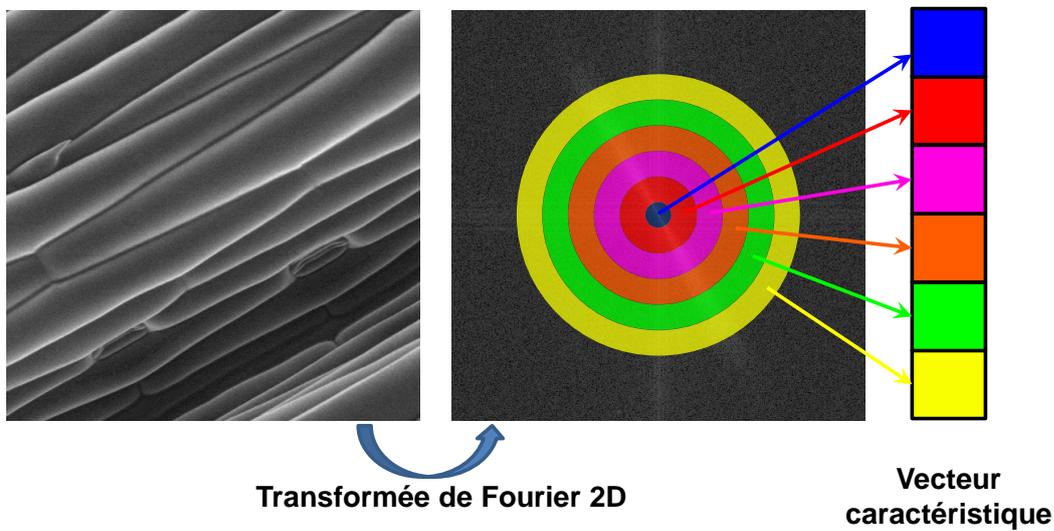


Figure 3.2.1: Schématisation de l'acquisition du descripteur généralisé de Fourier.

Un vecteur caractéristique du voisinage considéré pour la mesure de netteté est donc obtenu (figure 3.2.1). Une fois ce vecteur déterminé pour chaque pixel de la séquence d'images, nous proposons deux façons de transformer l'information de ces vecteurs en mesures de netteté.

La première est de simplement sommer l'ensemble des valeurs de chaque vecteur. En effet, plus élevée est la somme des énergies de chaque fréquence, plus net est le voisinage considéré. Cette mesure de netteté est définie par l'équation 3.2.3 :

$$FM_{DGFsom} = \sum_{\lambda=1}^M D_f(\lambda), \quad (3.2.3)$$

où M représente la taille du vecteur caractéristique égal à la moitié de la taille du voisinage.

La seconde méthode de mesure de netteté se base sur toute une séquence de voisinages d'un pixel de mêmes coordonnées dans chaque image. Ainsi, cette mesure s'apparente à une mesure de netteté 3D comme certaines citées précédemment. Le calcul de chaque vecteur caractéristique du même pixel de chaque image est effectué puis ces vecteurs sont concaténés en une matrice (figure 3.2.2). Une analyse en composantes principales [Jolliffe, 2005] est appliquée à cette matrice (M). Les scores factoriels issus de l'ACP permettent d'obtenir une projection optimisée des différents voisinages (ou pixels considérés) selon les vecteurs des descripteurs généralisés de Fourier correspondants. Cette projection est optimisée selon le premier axe de l'ACP qui contient le pourcentage d'inertie (% d'information) le plus important au regard des autres axes. La projection des scores factoriels de l'axe 1 met en évidence un gradient de netteté allant des valeurs les plus faibles pour les pixels les moins nets, aux scores factoriels les plus importants pour les pixels les plus nets. Autrement dit, obtenir le score factoriel le plus fort revient à déterminer la position de l'image dans laquelle le pixel est le plus net :

$$FM_{DGF_{acp}} = position(max(vecteur ACP)). \quad (3.2.4)$$

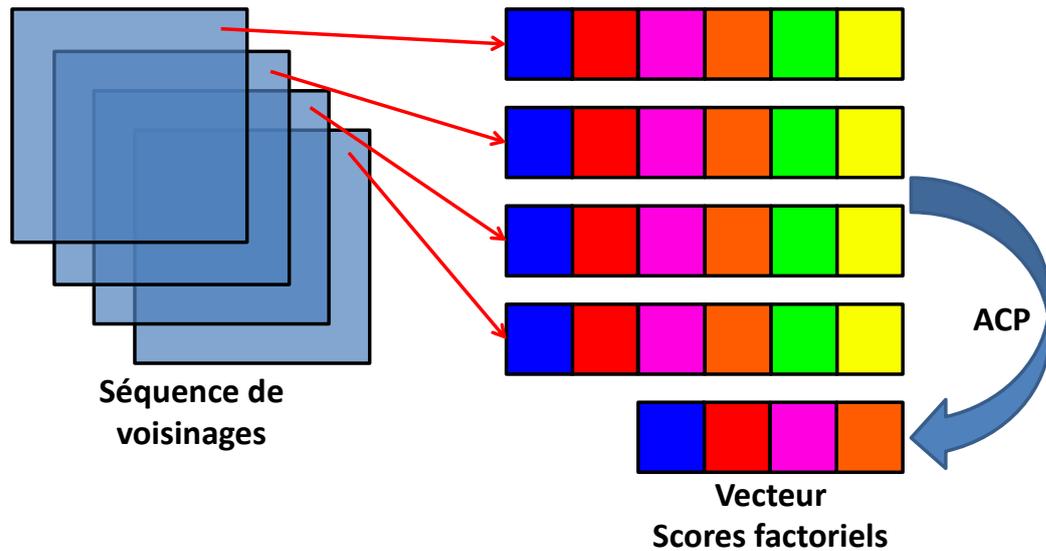


Figure 3.2.2: Obtention du vecteur des scores factoriels.

En résumé, les opérateurs de mesure de netteté conservés pour la suite de l'étude sont :

1. FM_{tenvar} : variance de l'amplitude du gradient ;
2. $FM_{nrjgrad}$: énergie du gradient ;
3. FM_{sml} : somme de laplaciens modifiés ;
4. FM_{var} : contraste par variance des niveaux de gris ;
5. $FM_{entropy}$: mesure de l'entropie locale ;
6. FM_{range} : contraste par différence entre maximum et minimum des niveaux de gris ;
7. FM_{corr} : mesure de corrélation ;
8. FM_{orient} : mesure par filtres orientables ;
9. $FM_{wavratio}$: mesure par ratio d'ondelettes ;
10. $FM_{DGF_{acp}}$: mesure par régression linéaire des descripteurs généralisés de Fourier ;
11. $FM_{DGF_{som}}$: mesure par somme des composantes des descripteurs généralisés de Fourier.

3.3 Approximation

Une fois que les positions des images pour lesquelles la netteté est estimée maximum pour chaque pixel sont connues, la création de la carte de profondeur pourrait être faite directement en affectant à chaque pixel la distance correspondante à la position de l'image associée. Cependant, la séquence d'images étant discrète (et le pas entre chaque image étant dépendant des paramètres choisis par l'utilisateur), cette carte de profondeur obtenue par la méthode directe souffre d'une topographie en « marche d'escalier ». Afin d'obtenir une carte de profondeur avec une meilleure continuité, une approximation de la courbe de mesure de netteté est utilisée. Ceci permet également une meilleure précision de l'estimation de profondeur puisque approximer la courbe permet de limiter l'impact des valeurs de netteté aberrantes. [Nayar, 1989] suggère d'interpoler les valeurs des mesures de netteté en admettant que celles-ci suivent une distribution gaussienne (figure 3.3.1).

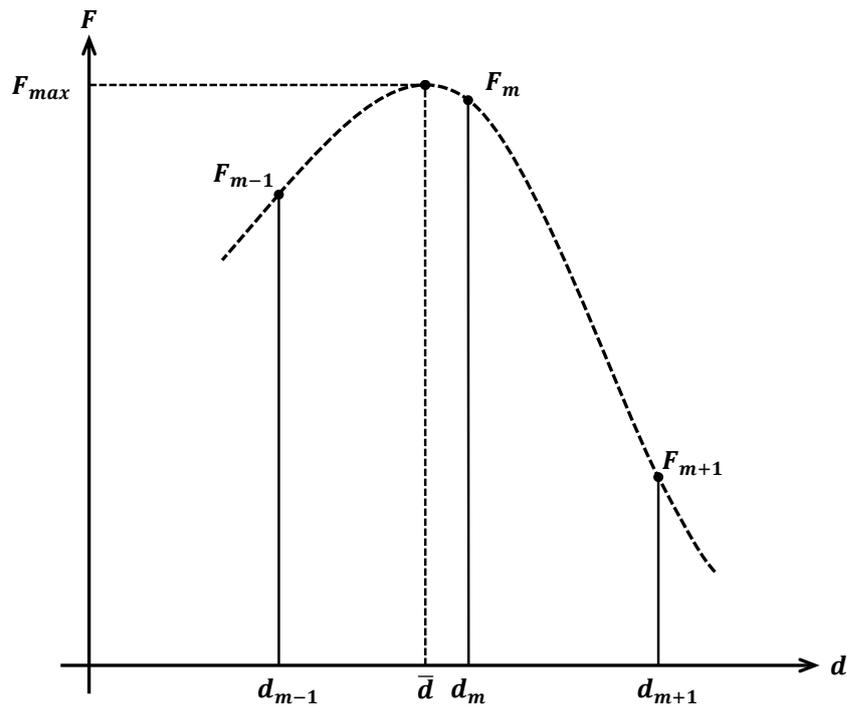


Figure 3.3.1: Approximation gaussienne.

La distance \bar{d} correspondant au maximum de la courbe approximée F_{max}

est obtenue à partir des 3 points F_{m-1} , F_{m+1} et F_m dont la distance associée est notée d_m :

$$\bar{d} = d_m + \frac{\log(F_{m-1}) - \log(F_{m+1})}{2\log(F_{m-1}) - 4\log(F_m) + 2\log(F_{m+1})}. \quad (3.3.1)$$

Dans le cas d'une courbe de netteté pour laquelle une interpolation trois points est délicate, un ajustement d'une courbe à distribution gaussienne est utilisé.

D'autres types d'approximations sont proposés comme, par exemple, un ajustement de polynômes de degrés 2 [Niederöst et al., 2002]. Une approche non basée sur l'ajustement de courbe est proposée par [Subbarao et Choi, 1995] et nommée « Focused Image Surface ». Cette méthode particulière de « Shape from Focus » se base sur une modélisation de la surface visualisée. L'approximation de la surface plane existante pour chaque voisinage 3D de pixel est calculée de manière à maximiser la mesure de netteté. [Choi et Yun, 2000] reprennent ce principe mais approximent la surface par des polynômes de Lagrange alors que [Asif et Choi, 2001] utilisent un réseau de neurones. L'inconvénient de ces méthodes est le temps de calcul beaucoup trop important afin d'obtenir la carte de profondeur. De plus, avec une mesure de netteté adaptée à la scène considérée, une approximation gaussienne se révèle efficace du fait du comportement de la mesure de netteté apparenté à une distribution gaussienne.

3.4 Comparaison

3.4.1 Méthodes de comparaison usuelles

Le choix de l'opérateur de mesure de netteté étant une étape fondamentale que ce soit dans le cadre de la méthode de « Shape from Focus » ou pour de la mise au point automatique, une comparaison de ces opérateurs est nécessaire. C'est dans le cadre de la recherche de la distance de mise au point optimale en microscopie que l'on retrouve le plus de comparaisons dans la littérature. Ainsi, [Firestone et al., 1991] comparent plusieurs algorithmes de mesure de netteté en évaluant quatre caractéristiques :

1. La précision : considérée comme la différence entre la position maximale de la courbe de netteté obtenue et la position exacte.
2. La plage : intervalle où la courbe de netteté est monotone. Cette caractéristique indique une réaction fidèle à une augmentation ou une diminution de la netteté relative dans la scène.
3. Le nombre de faux maximums.
4. La largeur de la courbe de netteté à mi-hauteur : plus cette donnée est faible et plus la mesure est considérée comme précise.

[[Pertuz et al., 2012](#)] ont proposé une comparaison très complète de la plupart des opérateurs appliqués à la méthode de « Shape from Focus ». Cette comparaison est basée sur la mesure de l'erreur quadratique moyenne entre une carte de profondeur obtenue pour chaque opérateur et une carte de profondeur de référence. Ce type de comparaison nécessite d'être en possession de la vérité terrain et c'est pourquoi elle est couramment basée sur des scènes de synthèse. De plus, la robustesse des opérateurs est évaluée en considérant cinq niveaux de jeux de paramètres affectant de plus en plus la qualité d'une mesure de netteté. Ainsi, les paramètres variant et affectant les mesures sont : une diminution de la taille du voisinage considérée, du bruit additif, une baisse du contraste et l'augmentation de la saturation dans les images.

Bien que donnant une idée des opérateurs à utiliser, ces méthodes de comparaison ne renseignent pas sur la robustesse des opérateurs aux variations de paramètres les affectant et ce, séparément. Avec nos systèmes d'acquisition pour lesquels l'éclairage est contrôlé, les mesures de netteté ne sont pas affectées par les facteurs d'influence que sont le changement de contraste et de saturation. En effet, chaque image de la séquence est acquise dans les mêmes conditions que les autres. Il reste donc deux paramètres influents sur les opérateurs de mesure de netteté : la taille du voisinage et la présence de bruit additif. La taille du voisinage considérée pour la mesure de netteté de chaque pixel est un élément délicat à fixer. En effet, en théorie, ce voisinage doit être le plus petit possible afin de tenir compte de toutes les discontinuités et d'obtenir la carte de profondeur la plus précise possible. Cependant, un petit voisinage entraîne nécessairement peu d'échantillons pour la mesure de netteté et donc une robustesse au bruit faible et une précision de l'estimation

de profondeur moins importante que pour un voisinage plus grand. Un compromis doit donc être fait et n'est pas implicite. Le bruit additif, quant à lui, est un élément affectant fortement les mesures de netteté puisque correspondant à des composantes hautes fréquences. Ainsi, un voisinage sans bruit et net peut être considéré moins net qu'un voisinage flou et bruité du fait de la présence de composantes hautes fréquences. On retrouvera, par exemple, du bruit thermique ou impulsionnel dans les images acquises. Le bruit thermique correspond principalement au courant d'obscurité dû à des électrons libres générés lorsque le capteur n'est pas assez refroidi. Ces électrons s'ajoutent à ceux d'origine lumineuse et faussent la valeur réelle du pixel affecté. Le bruit impulsionnel correspond à des erreurs de transmission ou à une défaillance de photo-éléments. Du bruit numérique, dû à une augmentation du gain ou une prise d'image en faibles conditions d'éclairage, peut également être présent mais résulte principalement d'une mauvaise gestion de l'acquisition.

3.4.2 Méthode de comparaison adaptée

Souhaitant évaluer la robustesse des opérateurs non seulement à la variation de la taille du voisinage et au bruit additif mais également à la variation simultanée de ces deux facteurs d'influence, une méthode de comparaison a été développée. Le processus de comparaison se décompose en quatre parties détaillées ci-après.

Séquences d'images de référence

La première étape est bien entendu la création des images de test simulant des séquences d'images. Compte tenu du fait que la méthode de reconstruction 3D a pour objectif d'être appliquée à des scènes naturelles texturées, l'utilisation de la base d'images VisTex (« Vision Texture ») [VisTex, 2002] semble pertinente. En effet, cette base est constituée d'images de scènes naturelles acquises en conditions non contrôlées et réparties en différentes classes : écorces d'arbres, murs, textiles, aliments, métaux, sols, fleurs, plantes, eau, bois et divers. Dans notre cas, nous utilisons un exemplaire de chaque texture différente soit 72 images de définition 512 par 512 pixels (figure 4.4.1 en annexe). De sorte à évaluer la robustesse au bruit des opérateurs de mesure

de netteté, les images de base sont bruitées selon 5 niveaux de bruit différents. [Malik et Choi, 2008] suggèrent d'évaluer la robustesse des images pour 3 types de bruits : thermique, impulsionnel et « speckle ». Comme précisé précédemment, le bruit impulsionnel survient lors d'une mauvaise transmission de l'information ou d'une défaillance de photo-élément. Le « speckle » est un bruit apparaissant lors de la réflexion d'une lumière cohérente, type laser, sur une surface texturée dont les motifs sont du même ordre de dimension que la longueur d'onde. Dans notre cas, le « speckle » n'est pas considéré du fait de notre système dénué de laser. De même pour le bruit impulsionnel dont la probabilité d'apparition est négligeable, ou alors apparaissant ponctuellement et tout aussi négligeable de par la définition des images. Le bruit thermique est donc le bruit étudié et modélisé par un bruit gaussien. [Malik et Choi, 2008] proposent de modéliser ce bruit par des niveaux de variance de 0.5, 0.05, 0.005, 0.0005 et 0.00005. Cependant, utiliser ces niveaux de bruit n'entraîne pas une augmentation croissante de la valeur de corrélation entre les images bruitées et l'image originale (figure 3.4.1).

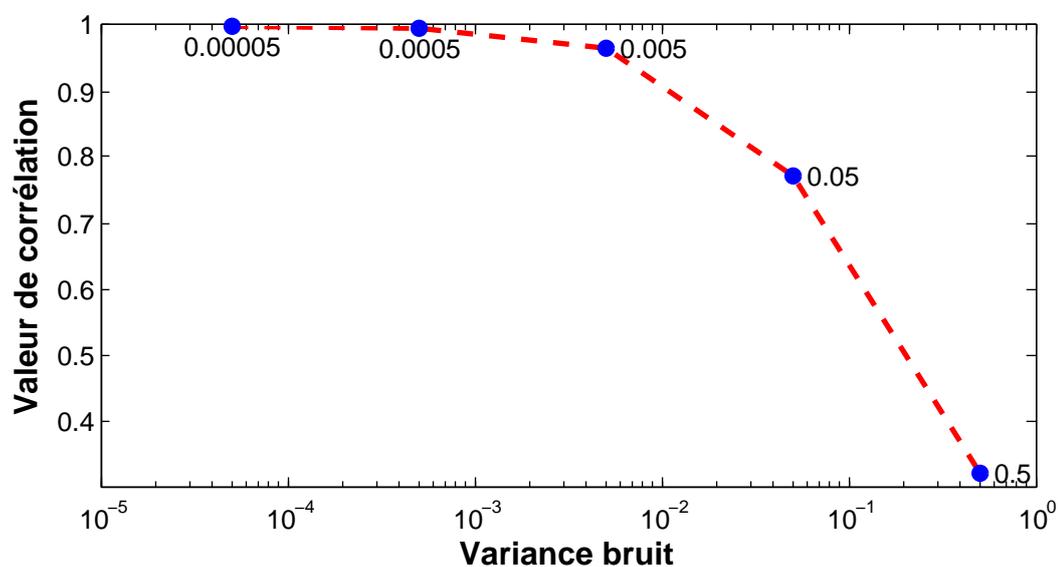


Figure 3.4.1: Effet des niveaux de bruit proposés par [Malik et Choi, 2008] sur la valeur de corrélation entre l'image originale et l'image bruitée.

On constate un impact mineur des bruits pour des variances de 0.00005, 0.0005 et 0.005 alors qu'une variance de 0.05 et 0.5 entraine respectivement une image bruitée à 23% et 68%. Afin d'avoir une évolution linéaire du bruit appliqué à chaque image, on détermine les variances à utiliser afin d'obtenir un bruit couvrant 5, 10, 15 et 20% des images. Ainsi, pour la première image de la base VisTex, on obtient les valeurs illustrées figure 3.4.2.

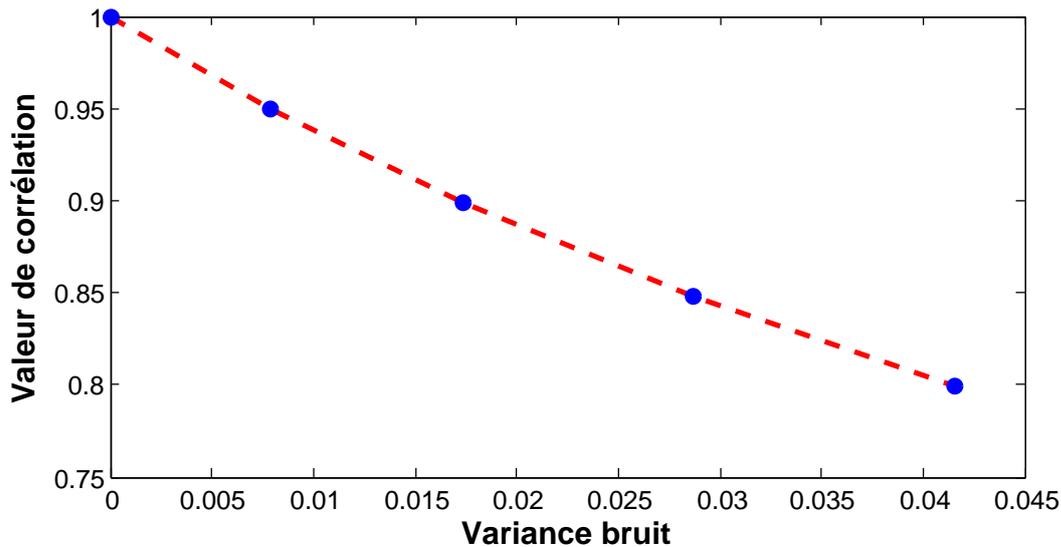


Figure 3.4.2: Effet de niveaux de bruit induisant une évolution constante de la valeur de corrélation entre l'image originale et l'image bruitée.

Une fois les images de la base bruitées, deux flous successifs sont appliqués à celles-ci afin de simuler une image de profondeur. Ainsi, pour chaque image de la base VisTex, 5 images bruitées à 0, 5, 10, 15, et 20% sont créées.

Calcul des cartes de profondeur

Les cartes de profondeur issues de ces séquences d'images sont ensuite calculées pour chaque opérateur à évaluer. Les tailles de voisinage considérées pour les mesures de netteté de chaque pixel sont carrées de largeurs 64, 32, 16, 8 et 4 pixels. Ainsi, pour un opérateur et une image de la base, on obtient 25 cartes de profondeur pour chaque combinaison de paramètres (bruit additif et taille du voisinage que l'on nommera par la suite doublet de paramètres). Au final, on obtient 25 doublets de paramètres \times 11 opérateurs de mesure de netteté \times 72 images de textures = 19800 cartes de profondeur. Ces cartes sont

constituées de 3 niveaux de gris pour chaque bande considérée nette (bande 1 - image originale, bande 2 - image avec léger flou et bande 3 - image avec flou plus important). Une carte de profondeur idéale (figure 3.4.3a) est entièrement constituée de niveaux de gris de valeur 85 alors qu'une erreur induit un niveau de gris de 170 ou 255 (figure 3.4.3b voisinage 64×64 et bruit de 10%) dans le cas où un opérateur considère comme nette une bande floutée. Ces cartes sont créées suite à une évaluation de la netteté bloc par bloc de la taille du voisinage choisie et pixel par pixel. Ceci afin d'acquérir l'ensemble des cartes de profondeur plus rapidement compte tenu de la quantité de combinaisons.

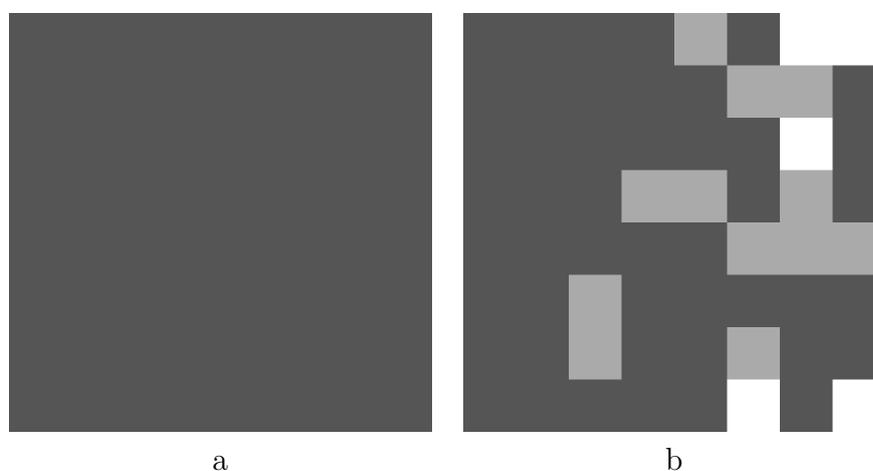


Figure 3.4.3: Exemple de cartes de profondeur obtenues : a. Voisinage 64×64 sans bruit b. Voisinage 64×64 et bruit de 10%.

Evaluation des erreurs

Un fois toutes les cartes de profondeur créées, une métrique doit être choisie afin de quantifier l'erreur induite par chacun des opérateurs de mesure de netteté pour chaque combinaison. Dans notre cas, l'utilisation de l'erreur moyenne quadratique n'est pas adaptée car un voisinage de 64 par 64 pixels présentant une erreur fait état d'une erreur quadratique plus importante que dans le cas d'un voisinage de 8 par 8 pixels, par exemple. L'erreur quadratique est donc différente alors que dans les deux cas l'erreur réelle constatée est identique. Dans le cas de notre application, les métriques classiques (erreur quadratique, corrélation) ne sont pas adaptées. Une métrique spécifique est

donc définie et repose sur le taux de voisinage présentant une erreur d'estimation de profondeur :

$$\text{Nombre voisinage faux} = \frac{\text{nbr pixel faux}}{\text{taille voisinage}^2}, \quad (3.4.1)$$

$$\text{Taux voisinage faux} = \frac{\text{Nombre voisinage faux}}{\text{Nombre voisinage présent dans image}}. \quad (3.4.2)$$

Un pourcentage d'erreur est ainsi calculé pour chaque carte de profondeur.

Plan d'expériences factoriel

Dans le but d'évaluer la robustesse des opérateurs de mesure de netteté face à la variation de la taille de voisinage et l'ajout de bruit, un plan d'expériences factoriel est mis en place. L'utilisation d'un plan d'expériences permet d'étudier le comportement d'un résultat en fonction de la variation des facteurs choisis. Dans notre cas, le résultat est l'erreur induite par chaque opérateur en fonction de la variation de 2 facteurs : la taille du voisinage et le bruit additif. Ainsi, un plan d'expériences à 2 niveaux est mis en place pour chaque opérateur. La première étape est la constitution de la matrice d'expériences qui est du même type quel que soit l'opérateur (tableau 3.1).

	Facteur 1	Facteur 2	Interaction	Réponse
1	-1	-1	+1	y_1
2	+1	-1	-1	y_2
3	-1	+1	-1	y_3
4	+1	+1	+1	y_4
Niveau bas	Valeur min facteur 1	Valeur min facteur 2		
Niveau haut	Valeur max facteur 1	Valeur max facteur 2		
Effets	E_1	E_2	E_{12}	

Tableau 3.1: Matrice d'expériences, effets et interaction pour 2 niveaux de paramètres.

Les valeurs minimales et maximales du facteur 1 sont les tailles de voisinages choisies pour l'étude alors que ce sont les niveaux de bruit minimum et maximum pour le facteur 2. Les réponses y_1 , y_2 , y_3 et y_4 correspondent aux erreurs calculées pour chaque opérateur pour les doublets de facteurs 1 et 2. On note que seulement 4 doublets de paramètres sont utilisés dans le cadre du plan d'expériences. Toutes les autres mesures d'erreurs pour les 21 autres doublets sont utilisées section 3.5.2 dans le but de déterminer les meilleurs opérateurs pour un doublet de paramètres donné.

Une fois la matrice d'expériences créée, les effets (E_1 , E_2 eq. 3.4.3 et 3.4.4) ainsi que l'interaction entre les 2 facteurs (E_{12} eq. 3.4.5) sont calculés pour chaque opérateur :

$$E_1 = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4), \quad (3.4.3)$$

$$E_2 = \frac{1}{4}(-y_1 - y_2 + y_3 + y_4), \quad (3.4.4)$$

$$E_{12} = \frac{1}{4}(+y_1 - y_2 - y_3 + y_4). \quad (3.4.5)$$

3.4.3 Analyse du comportement des opérateurs

Les tableaux 3.2 et 3.3 illustrent les effets et l'interaction de chaque opérateur alors que les figures 3.4.4 et 3.4.5 représentent le comportement des erreurs en fonction de doublets de paramètres différents. Les premiers tableau et figure illustrent le comportement de l'erreur en fonction d'un voisinage variant de 64×64 à 4×4 pixels et un bruit de 0% à 20%. Le second jeu de données est moins restrictif avec le facteur 1 variant de 64×64 à 8×8 et le bruit de 0% à 10%.

Opérateurs	E_1 (Voisinage)	E_2 (Bruit)	E_{12}	Moyenne des erreurs
FM_{tenvar}	0.6099	0.4574	0.4109	16.0909
$FM_{nrjgrad}$	0.4553	0.7349	-0.0512	35.9979
FM_{sml}	0.2131	0.8275	0.1410	27.0322
FM_{var}	0.6124	0.4667	0.3965	16.3457
$FM_{entropy}$	0.3098	0.7465	-0.3110	41.4041
FM_{range}	0.5517	0.5566	0.3685	17.9252
FM_{corr}	0.8439	0.1866	0.0551	30.4931
FM_{orient}	0.7080	0.4572	0.1992	24.4235
$FM_{wavratio}$	0.2705	0.8023	0.1819	25.1023
FM_{DGFacp}	0.5835	0.5132	0.3823	19.3078
FM_{DGFsom}	0.6119	0.4741	0.3883	19.1451

Tableau 3.2: Synthèse des résultats des plans d'expériences (voisinage 64×64 et 4×4 ; bruit 0% et 20%).

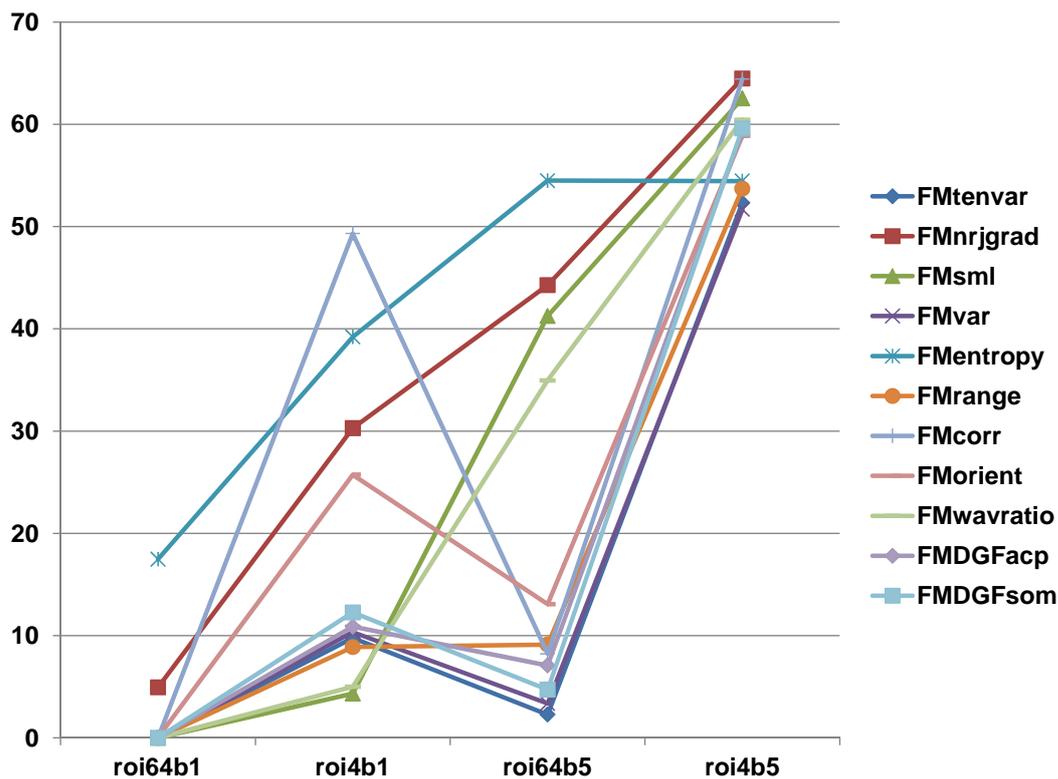


Figure 3.4.4: Evolution de l'erreur en fonction des doublets de paramètres (voisinage 64×64 et 4×4 ; bruit 0% et 20%).

Opérateurs	E_1 (Voisinage)	E_2 (Bruit)	E_{12}	Moyenne des erreurs
FM_{tenvar}	0.5498	0.4973	0.4477	6.96
$FM_{nrjgrad}$	0.5409	0.6491	0.1899	26.3819
FM_{sml}	0.2759	0.7774	0.2637	17.4966
FM_{var}	0.5686	0.4875	0.4348	7.2946
$FM_{entropy}$	0.5144	0.6947	0.0525	36.3676
FM_{range}	0.4314	0.6688	0.3415	11.1156
FM_{corr}	0.7775	0.3137	0.2170	19.1190
FM_{orient}	0.6793	0.4330	0.3179	16.1515
$FM_{wavratio}$	0.3144	0.7503	0.2971	15.2910
$FM_{DGF_{facp}}$	0.5255	0.5267	0.4431	10.3563
$FM_{DGF_{som}}$	0.5480	0.5002	0.4467	9.5095

Tableau 3.3: Synthèse des résultats des plans d'expériences (voisinage 64×64 et 8×8 ; bruit 0% et 10%).

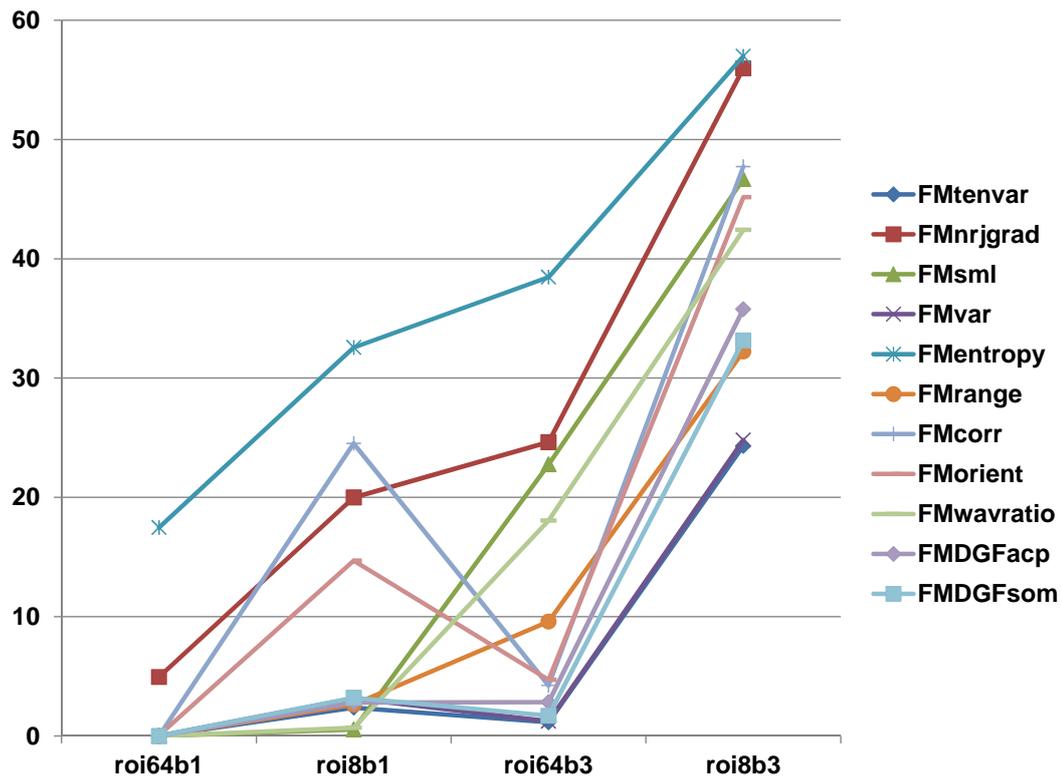


Figure 3.4.5: Evolution de l'erreur en fonction des doublets de paramètres (voisinage 64×64 et 8×8 ; bruit 0% et 10%).

En analysant les deux synthèses, on remarque que globalement le comportement de chaque opérateur suit la même tendance que ce soit pour une évolution de paramètres plus ou moins restrictifs. L'étude du comportement de chaque opérateur doit être réalisée non seulement au regard des effets et interactions mais également au regard de l'erreur induite par les doublets de paramètres choisis.

Ainsi, deux comportements principaux apparaissent : une robustesse à la variation de la taille du voisinage plus importante que pour une variation du bruit additif et, inversement, une robustesse plus élevée au bruit qu'à une réduction du voisinage considéré.

Le premier type de comportement concerne les opérateurs de mesure de netteté $FM_{nrjgrad}$, FM_{sml} , $FM_{entropy}$, FM_{range} et $FM_{wavratio}$. Pour ces opérateurs, l'ajout de bruit impacte d'avantage la qualité de la mesure qu'une variation de la taille du voisinage. Parmi ce groupe, les opérateurs $FM_{nrjgrad}$ et $FM_{entropy}$ se distinguent par une erreur d'estimation de netteté pour un voisinage important et aucun bruit additif. De plus, ces derniers présentent peu de robustesse non seulement à un bruit additif mais également à une variation de la taille de voisinage au vu de l'erreur comparée à tous les autres opérateurs. L'interaction calculée pour ces opérateurs est relativement faible mais cela s'explique par une erreur déjà importante pour une variation d'un seul paramètre. Pour les trois opérateurs restants, une distinction significative est faite entre FM_{range} légèrement moins robuste à une variation de voisinage mais moins sensible à un bruit additif comparé à FM_{sml} et $FM_{wavratio}$. Ces deux derniers opérateurs sont très sensibles au bruit mais font état des meilleurs résultats lors d'une variation de taille de voisinage.

Le second type de comportement est valable pour tous les autres opérateurs, à savoir FM_{tenvar} , FM_{var} , FM_{corr} , FM_{orient} , $FM_{DGF_{acp}}$ et $FM_{DGF_{som}}$. En effet, pour ces opérateurs, la variation de la taille de voisinage induit un effet plus important que celui calculé pour une variation de bruit additif. Pour l'ensemble de ces opérateurs, on constate une robustesse importante au bruit additif en comparaison aux opérateurs précédents. Un comportement global similaire des opérateurs FM_{var} , FM_{tenvar} , $FM_{DGF_{acp}}$ et $FM_{DGF_{som}}$ apparaît que ce soit au niveau des effets, interactions et erreurs. En revanche, FM_{corr} et

FM_{orient} font état d'une sensibilité aux variations de la taille du voisinage largement supérieure aux autres opérateurs. Dans le cas moins restrictif et avec une variation simultanée des deux paramètres étudiés, les opérateurs FM_{var} et FM_{tenvar} se distinguent par leur comportement vis-à-vis des erreurs des erreurs. Pour le cas plus restrictif de variations des paramètres, la différence est plus ténue entre les opérateurs de cette catégorie.

Pour résumer, on recommande les opérateurs selon 4 cas différents :

1. Taille de voisinage importante et pas de bruit additionnel : à l'exception de $FM_{entropy}$ et $FM_{nrjgrad}$, tous les autres opérateurs présentent un bon comportement dans ce cas ;
2. Taille de voisinage importante et bruit additif : les opérateurs FM_{tenvar} , FM_{var} , $FM_{DGF_{acp}}$ et $FM_{DGF_{som}}$ se comportent de façon correcte et similaire de par leur robustesse au bruit ;
3. Petite taille de voisinage et aucun bruit additionnel : à l'exception de FM_{corr} , $FM_{entropy}$, $FM_{nrjgrad}$ et FM_{orient} , les autres opérateurs sont à conseiller. Un avantage certain, dû à la robustesse à la variation de la taille du voisinage, est constaté pour FM_{sml} et $FM_{wavratio}$;
4. Petite taille de voisinage et bruit additif : cette configuration de paramètres influant sur la qualité de la mesure de netteté est la plus restrictive. La recherche de l'opérateur adapté dans ce cas passe par la recherche de ceux ne démultipliant pas l'erreur pour une variation simultanée des deux paramètres. C'est le cas de l'opérateur FM_{range} qui, même s'il n'est généralement pas en bonne position dans les tests comparatifs de la littérature, présente un comportement intéressant dans ce cas particulier. Les opérateurs dont la sensibilité est moindre face aux deux types de variations de paramètres apparaissent également comme un bon choix. On retrouve dans ce cas les opérateurs FM_{tenvar} , FM_{var} , $FM_{DGF_{acp}}$ et $FM_{DGF_{som}}$.

Pour conclure, cette méthode de comparaison permet d'étudier le comportement des différents opérateurs contrairement aux méthodes traditionnelles présentant un classement global. Ceci permet l'utilisation d'un opérateur adapté à un cas particulier potentiellement plus performant qu'un opé-

rateur considéré comme globalement efficace. De par le processus de calcul de l'erreur réalisé, un classement général est toutefois possible et est présenté section 3.5.2. On constate que nos opérateurs ont un comportement similaire aux opérateurs FM_{tenvar} et FM_{var} , considérés comme performants dans les études comparatives de la littérature.

3.5 Evaluation automatique

L'objectif étant la réalisation d'un système de reconstruction 3D dédié principalement à une application aux données agronomiques, le système se doit d'être accessible et convivial pour permettre son utilisation par des non spécialistes. Pour cela, une évaluation automatique de l'opérateur de mesure de netteté à utiliser en fonction de la scène considérée est nécessaire. Dans cette section, deux méthodes sont proposées dont l'une est non supervisée et la seconde supervisée. Une méthode non supervisée ne nécessite aucune intervention de l'utilisateur alors qu'une méthode supervisée nécessite une ou plusieurs informations en amont du processus d'automatisation.

3.5.1 Méthode non supervisée par fusion d'opérateurs

Traditionnellement, la méthode de reconstruction 3D « Shape from Focus » passe par l'utilisation d'un seul opérateur de mesure de netteté dans le cadre du processus de reconstruction. Cependant, l'étude comparative précédente met en évidence le fait que chaque opérateur fait état de forces et faiblesses quant à sa robustesse au bruit ou au changement de taille de voisinage. Ainsi, une fusion de l'information d'estimation de la netteté issue de plusieurs opérateurs est envisagée. Ceci dans l'optique de garantir une carte de profondeur relativement précise quel que soit le type de scène ou d'objet considérés en tirant partie des atouts de plusieurs opérateurs.

Pour ce faire, une analyse en composantes principales (également appelée transformée de Karhunen-Loeve en traitement d'image) est appliquée à nos

données. L'objectif de cette analyse est de déterminer un nouvel axe de projection des données maximisant la variabilité entre celles-ci.

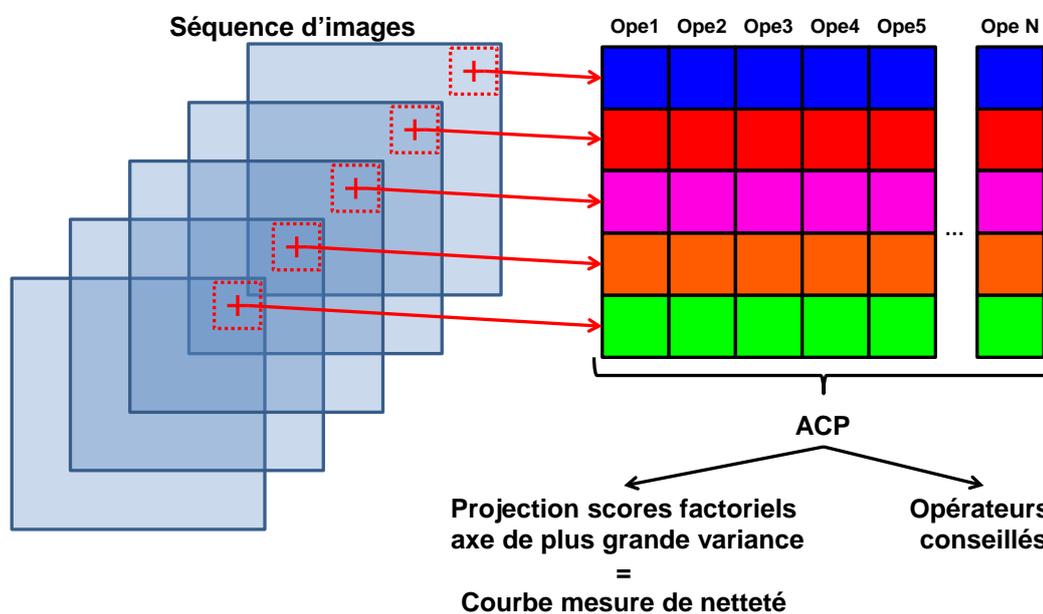


Figure 3.5.1: Processus de fusion d'opérateurs de mesure de netteté.

Nos données représentent la concaténation des différents vecteurs de mesure de netteté issus de nos opérateurs en une matrice dont les colonnes font référence aux variables (dans notre cas les opérateurs de mesure de netteté) et les lignes aux individus (les images de la séquence). Chaque vecteur de netteté est bien entendu normalisé entre 0 et 1 compte tenu du fait que la dynamique de mesure est différente pour chaque opérateur. Le processus de fusion est illustré figure 3.5.1 et un exemple de courbe de netteté obtenue est fournie figure 3.5.2. L'utilisation de cette courbe de netteté est la même que pour celles issues des opérateurs, c'est-à-dire une approximation permettant l'estimation de la profondeur du point de la scène considéré. Cette nouvelle courbe est le résultat de la projection des scores factoriels issus de l'analyse en composantes principales. Cela correspond à une projection optimisée des différentes valeurs de netteté des voisinages considérés selon les vecteurs de netteté de chaque opérateur utilisé. La projection concerne seulement les scores factoriels du premier axe de l'ACP en fonction de la position des voisinages correspondante dans la séquence d'images. En effet, le premier axe, dans notre cas, est celui

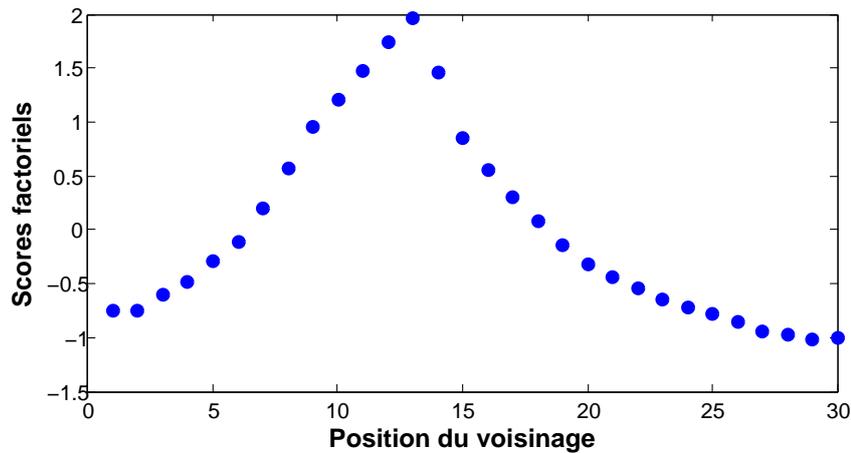


Figure 3.5.2: Courbe de mesure de netteté correspondant à la projection des scores factoriels.

faisant état du pourcentage d’inertie le plus important comparé aux autres axes (environ 90%).

Outre la courbe de mesure de netteté, l’information sur le choix de l’opérateur à utiliser pour la scène considérée est obtenue. En effet, puisque le maximum de cette nouvelle courbe correspond à l’estimation de la profondeur du pixel en cours, un poids peut être affecté aux opérateurs ayant estimé la même position. Ainsi, un pourcentage de bonne participation est obtenu pour l’ensemble de la séquence d’images permettant de conseiller les opérateurs dont l’estimation est la meilleure.

Ce nouveau processus d’estimation de la profondeur utilisant plusieurs opérateurs nécessite un temps de calcul évidemment plus important puisque la mesure de netteté est calculée pour chaque opérateur. Cependant, ceci permet à l’utilisateur n’ayant pas connaissance de l’opérateur à utiliser, d’obtenir une carte de profondeur plus précise que celle obtenue par un opérateur de mesure de netteté non adapté. Dans le cas où plusieurs reconstructions 3D d’un type de scène similaire sont envisagées, une première itération du processus permet de déterminer l’opérateur conseillé pour toutes les autres scènes similaires. Bien entendu, cette méthode ne fonctionne que si une majorité d’opérateurs efficaces sont utilisés. Outre l’analyse du comportement présentée précédemment, l’étude par la méthode supervisée suivante a pour but une mise en

évidence des opérateurs de mesure de netteté les plus efficaces.

3.5.2 Méthode supervisée par analyse de variance

L'étude comparative a permis de mettre en avant une correspondance entre chaque doublet de paramètres (taille du voisinage et bruit additif) et des opérateurs de mesure de netteté plus adaptés que d'autres pour chaque cas. Déterminer par avance quels opérateurs sont à utiliser en fonction du doublet de paramètres lié au type de scène considérée permet la création d'une méthode supervisée de choix de l'opérateur le plus adapté. Les données nécessaires à l'évaluation des meilleurs opérateurs pour chaque doublet sont celles obtenues lors du processus de comparaison détaillé section 3.4.2. Pour rappel, les doublets de paramètres correspondent à une variation de la taille de voisinage de 64×64 , 32×32 , 16×16 , 8×8 et 4×4 pixels et un bruit additif variant de 0%, 5%, 10%, 15% et 20% (respectivement dans les figures et tableaux suivant b1, b2, b3, b4 et b5 dans les dénominations de doublet).

On regroupe les erreurs calculées pour chaque doublet de paramètres. Ainsi, pour chaque doublet est retrouvé l'ensemble des erreurs calculées pour chaque opérateur et chaque image de la base VisTex. La comparaison d'un ensemble d'échantillons indépendants est effectuée par une analyse de variance (terme anglais ANOVA pour « ANalysis Of VAriance ») dès lors que l'on cherche à comparer plus de deux moyennes. Dans notre cas, ce sont onze facteurs, correspondant à nos onze opérateurs testés, qui nécessitent la comparaison de onze moyennes puisque les images de la base VisTex représentent des répétabilités de calcul des erreurs. De plus, nos échantillons (les erreurs calculées) sont indépendants du fait que chaque opérateur induit une erreur propre à celui-ci et n'infère en aucun cas sur les erreurs des autres opérateurs puisque calculées séparément. Nos données sont indépendantes mais ne suivent pas une distribution parfaitement gaussienne. Dans ce cas, les tests conseillés sont plutôt des tests non paramétriques comme le test de Friedman. Toutefois, les tests non paramétriques sont en général moins robustes qu'une analyse de variance qui reste robuste même en cas de données ne suivant pas une distribution normale [Kirk, 1995].

Si le test F (test de Fischer) issu de l'analyse de variance effectuée pour chaque doublet est significatif, c'est-à-dire que la valeur est inférieure au seuil arbitrairement fixé à 5%, alors au moins deux moyennes sont différentes. Pour identifier les moyennes différentes, un test de comparaison multiple de moyennes est nécessaire. Ce type de test permet de constituer des groupes de moyennes qui ne sont pas significativement différentes entre elles. Ces groupes permettent ensuite d'analyser l'influence du facteur étudié (dans notre cas les opérateurs) et ainsi préconiser l'utilisation, pour chaque doublet, des opérateurs appartenant au groupe ayant la plus petite moyenne.

De nombreux tests de comparaison multiple de moyennes existent. Toutefois, nos échantillons ont le même effectif puisque, pour chaque doublet de paramètres, chaque opérateur est associé à 72 valeurs d'erreurs calculées du fait des 72 images de la base VisTex. Le test préconisé en cas d'effectifs similaires (ou dispositifs équi-répétés) est le test de Newman-Keuls couramment abrégé SNK pour Student-Newman-Keuls.

Dans notre cas, l'analyse de variance fait état d'un test F très inférieur à 5% pour l'ensemble des 25 doublets de paramètres. On applique donc le test de Newman-Keuls pour chacune des 25 analyses de variance. La synthèse des groupes d'opérateurs mis en évidence par le test est illustrée tableau 3.4. A noter que le test, dans certains cas, propose de placer certains opérateurs dans plusieurs groupes lorsqu'une appartenance unique n'est pas franche.

En vue d'une évaluation automatique de l'opérateur (ou des opérateurs si fusion de leur estimation) en fonction du doublet de paramètres utilisé, c'est donc dans le 1^{er} groupe (en bleu dans le tableau) que sont choisis les opérateurs. Afin de simplifier la compréhension de cette synthèse, une illustration du 1^{er} groupe uniquement pour chaque doublet est représentée tableau 3.5.

Une première constatation est le fait que certains opérateurs ne sont jamais associés au groupe conseillé pour chaque doublet de paramètres. C'est le cas de l'opérateur basé sur l'énergie du gradient ($FM_{nrjgrad}$) et celui basé sur le calcul de l'entropie ($FM_{entropy}$). A l'inverse, les opérateurs basés sur la variance du Tenengrad (FM_{tenvar}) et la variance des niveaux de gris du voisinage (FM_{var}) sont ceux à conseiller dans la plupart des cas. On leur

	FM_{tenvar}	FM_{var}	FM_{DCFsom}	FM_{DCFacp}	$FM_{wavratio}$	FM_{sml}	FM_{range}	FM_{corr}	FM_{orient}	$FM_{nrigrad}$	$FM_{entropy}$
roi64b1	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Green	Red
roi64b2	Blue	Blue	Blue	Blue	Green	Green	Blue	Blue	Blue	Green	Red
roi64b3	Blue	Blue	Blue	Blue	Green	Green	Blue	Blue	Blue	Green	Red
roi64b4	Blue	Blue	Blue	Blue	Green	Green	Blue	Blue	Blue	Green	Red
roi64b5	Blue	Blue	Blue	Blue	Green	Green	Blue	Blue	Blue	Green	Red
roi32b1	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Green	Red
roi32b2	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi32b3	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi32b4	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi32b5	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi16b1	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi16b2	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi16b3	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi16b4	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi16b5	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi8b1	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi8b2	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi8b3	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi8b4	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi8b5	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi4b1	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi4b2	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi4b3	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi4b4	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black
roi4b5	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Black	Black

Tableau 3.4: Synthèse des groupes issus de l'analyse de variance.

Du meilleur groupe au moins bon : Bleu, Vert, Rouge, Orange, Bleu clair, Noir, Blanc.

	FM_{genvar}	FM_{var}	FM_{DGRsom}	FM_{DGRrap}	$FM_{varratio}$	FM_{smi}	FM_{range}	FM_{corr}	FM_{orient}	$FM_{artifrad}$	$FM_{entropy}$
roi64b1											
roi64b2											
roi64b3											
roi64b4											
roi64b5											
roi32b1											
roi32b2											
roi32b3											
roi32b4											
roi32b5											
roi16b1											
roi16b2											
roi16b3											
roi16b4											
roi16b5											
roi8b1											
roi8b2											
roi8b3											
roi8b4											
roi8b5											
roi4b1											
roi4b2											
roi4b3											
roi4b4											
roi4b5											

Tableau 3.5: Synthèse du groupe d'opérateurs les plus efficaces issus de l'analyse de variance.

préfèrera les opérateurs basés sur le ratio d'ondelettes ($FM_{wavratio}$) ou sur le Laplacien modifié (FM_{sml}) pour de faibles tailles de voisinage exempt de bruit additif. Nos deux opérateurs basés sur les descripteurs généralisés de Fourier ($FM_{DGF_{acp}}$ et $FM_{DGF_{som}}$) sont utilisés pour une taille de voisinage relativement importante avec présence de bruit additif. Globalement, on retrouve les comportements mis en évidence par le plan d'expérience mais l'analyse de variance permet d'affiner cette connaissance de chaque opérateur.

Au sein du 1^{er} groupe d'opérateurs pour chaque doublet, l'erreur moyenne minimum engendrée par chaque variation de paramètres permet de créer la courbe figure 3.5.3.

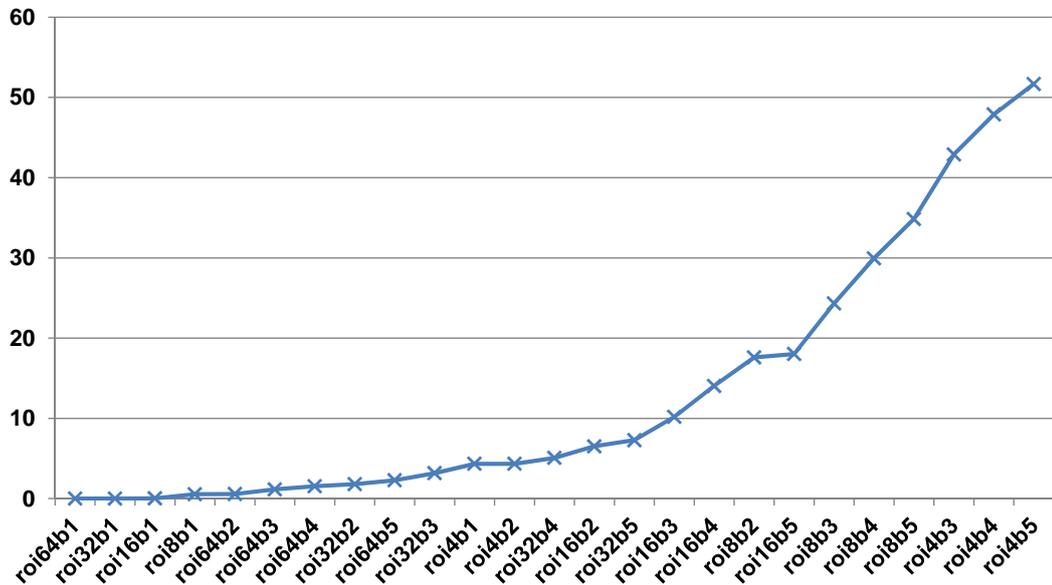


Figure 3.5.3: Courbe d'erreur moyenne minimale en fonction des doublets de paramètres.

Cette courbe permet de visualiser l'évolution de l'erreur moyenne minimale en fonction des variations de paramètres. Une forte augmentation de l'erreur à lieu lorsque la taille de voisinage se réduit conjointement à une augmentation du bruit additif. Ainsi, il est intéressant de mettre en évidence les doublets ayant un comportement similaire au regard de l'erreur moyenne engendrée. Pour cela, on effectue une analyse de variance avec pour facteur les dénominations des doublets et en considérant comme répétabilité l'ensemble des

autres facteurs que sont les opérateurs et images de la base VisTex. Le test de Newman-Keuls permet de dégager les groupes illustrés tableau 3.6.

Taille de voisinage	4					
	8					
	16					
	32					
	64					
		0%	5%	10%	15%	20%
		Bruit additif				

Tableau 3.6: Groupes de doublets de comportement similaire. Les couleurs identiques symbolisent les mêmes comportements. Le blanc est utilisé pour les doublets n'appartenant à aucun groupe.

Cette représentation permet un accompagnement de l'utilisateur quant au choix de la taille de voisinage à utiliser. En effet, la taille de voisinage est un paramètre devant être choisi par l'utilisateur en fonction de ses besoins. Une petite taille de voisinage engendre une carte de profondeur de plus grande précision tenant compte au maximum des discontinuités de profondeur. Cependant, le nombre d'échantillons limité induit par un petit voisinage entraîne une sensibilité accrue au bruit additif ainsi que des erreurs plus importantes lors de l'estimation de la profondeur. Un compromis doit donc être fait et le tableau 3.6 permet un accompagnement dans ce choix. Ainsi, on constate, par exemple, que lorsqu'une scène est exempte de bruit, la taille de voisinage peut être de 16×16 sans pour autant induire plus d'erreur que pour un voisinage de 64×64 ou 32×32 . On retrouve également des groupements relativement logiques pour une scène bruitée comme, par exemple, un voisinage de 16×16 de bruit 15% est similaire à un voisinage de 8×8 dont le bruit est de 5%.

Enfin, l'analyse de variance et le test de comparaison peuvent être utilisés afin d'obtenir un classement général des opérateurs même si ceci n'est pas représentatif de leurs forces et faiblesses par rapport aux variations de paramètres. Dans notre cas, le classement tableau 3.7 est obtenu.

Groupe	Opérateur	Moyenne erreurs (%)
A	FM_{tenvar}	14.87
A	FM_{var}	15.3
B	FM_{DGFsom}	19.65
B	FM_{range}	20.41
B	FM_{DGFacp}	21.22
C	FM_{orient}	27.73
C	FM_{corr}	28.25
C	$FM_{wavratio}$	29.55
D	FM_{sml}	32.95
E	$FM_{nrjgrad}$	40.61
F	$FM_{entropy}$	44.45

Tableau 3.7: Classement général des opérateurs de mesure de netteté (du meilleur groupe A au moins bon F).

Un groupe global considéré comme le meilleur est constitué de FM_{tenvar} et FM_{var} , ce qui suit la tendance des résultats expliqués précédemment. Nos opérateurs basés sur les descripteurs généralisés de Fourier se situent dans le 2^{ème} groupe avec une erreur moyenne totale 5% supérieure à celle du 1^{er} groupe. Sans pour autant égaler les opérateurs du 1^{er} groupe, ils se démarquent tout de même des autres opérateurs classés dans les groupes inférieurs.

3.6 Conclusion

Ce chapitre a été consacré à une étape fondamentale de l'estimation de profondeur en tout point d'une scène par « Shape from Focus » : l'évaluation de la netteté. Après un état de l'art des mesures usuelles existantes dans la littérature, deux nouveaux opérateurs basés sur les descripteurs généralisés de Fourier sont présentés. Ensuite, une présentation des méthodes d'approximation permettant un affinement de l'estimation de la netteté en vue d'une information de profondeur plus précise est effectuée.

Le choix de l'opérateur de mesure de netteté devant être parfaitement adapté à la scène considérée, il est crucial de proposer à l'utilisateur des outils permettant de faire ce choix. En règle générale, on retrouve dans la littérature des articles de comparaison des différents opérateurs de mesure de netteté. Cependant, ces comparaisons ne tiennent pas compte des forces et faiblesses de chaque opérateur par rapport aux différents paramètres affectant leur efficacité. Ainsi, une procédure d'étude du comportement des opérateurs en fonction de la taille de voisinage considérée pour la mesure et du bruit additif potentiellement présent dans les images est présentée. Elle se base sur la mise en place d'un plan d'expériences dont l'analyse des résultats permet une préconisation quant au choix de l'opérateur à utiliser en fonction du type de voisinage et la présence ou non de bruit. Préconisation qui est confirmée par l'analyse de variance des erreurs engendrées par plusieurs combinaisons opérateur / taille de voisinage / bruit / image de référence. A cette analyse de variance fait suite un test de comparaison multiple de moyennes (Newman-Keuls) permettant la mise en évidence de groupes d'opérateurs pour chaque doublet de paramètres (taille voisinage / bruit). Ces groupes permettent ensuite une automatisation du choix de l'opérateur à utiliser en fonction de la taille de voisinage et de la quantité de bruit estimée dans la séquence d'images.

Une méthode de choix automatique, non supervisée cette fois, est également présentée et repose sur la fusion des estimations de netteté de plusieurs opérateurs par le biais d'une analyse en composantes principales. Cette méthode permet de tendre vers une bonne estimation de la profondeur même en cas d'opérateurs peu performants des lors que plusieurs autres opérateurs performants sont également utilisés.

Bien entendu, l'ensemble de ces processus est à adapter en fonction des tailles maximales et minimales de voisinage étudiées ainsi que du pourcentage de bruit minimal et maximal. L'atout principal de cette méthode de comparaison est avant tout la caractérisation du comportement des opérateurs et non un simple classement comme ce peut être le cas traditionnellement.

Chapitre 4

Applications

Les deux chapitres précédents ont détaillé les deux étapes fondamentales de la méthode d'estimation de la profondeur « Shape from Focus », à savoir l'acquisition d'une séquence d'images et les outils permettant une évaluation locale de la netteté de la scène. Ce chapitre présente des résultats obtenus par l'ensemble de la méthode. Le processus global de reconstruction 3D est illustré figure [4.0.1](#).

Ainsi, ce processus est automatisé et amorcé suite à plusieurs choix concernant le nombre d'images à acquérir, le pas entre chaque acquisition (dépendant de la profondeur de champ induite par le système optique utilisé), l'automatisation (supervisée ou non), ou le choix manuel de l'opérateur de mesure de netteté. Ce schéma fonctionnel se décompose en trois parties :

1. **Génération de la séquence d'images.** L'acquisition peut être réalisée soit par le 1^{er} système soit par le 2nd en fonction du type de scène considéré et de la précision attendue. En effet, le second système a l'avantage de s'affranchir de la translation du centre optique durant le processus d'acquisition. En revanche, le 1er système permet d'utiliser tous types d'objectifs standards pour, par exemple, reconstruire de très petites scènes. Il est cependant exposé à une translation du centre optique due aux vibrations lors du processus d'acquisition. La séquence d'images acquise fait ensuite l'objet d'une correction des éventuelles distorsions ainsi que d'un recalage des images corrigeant le grossissement induit par

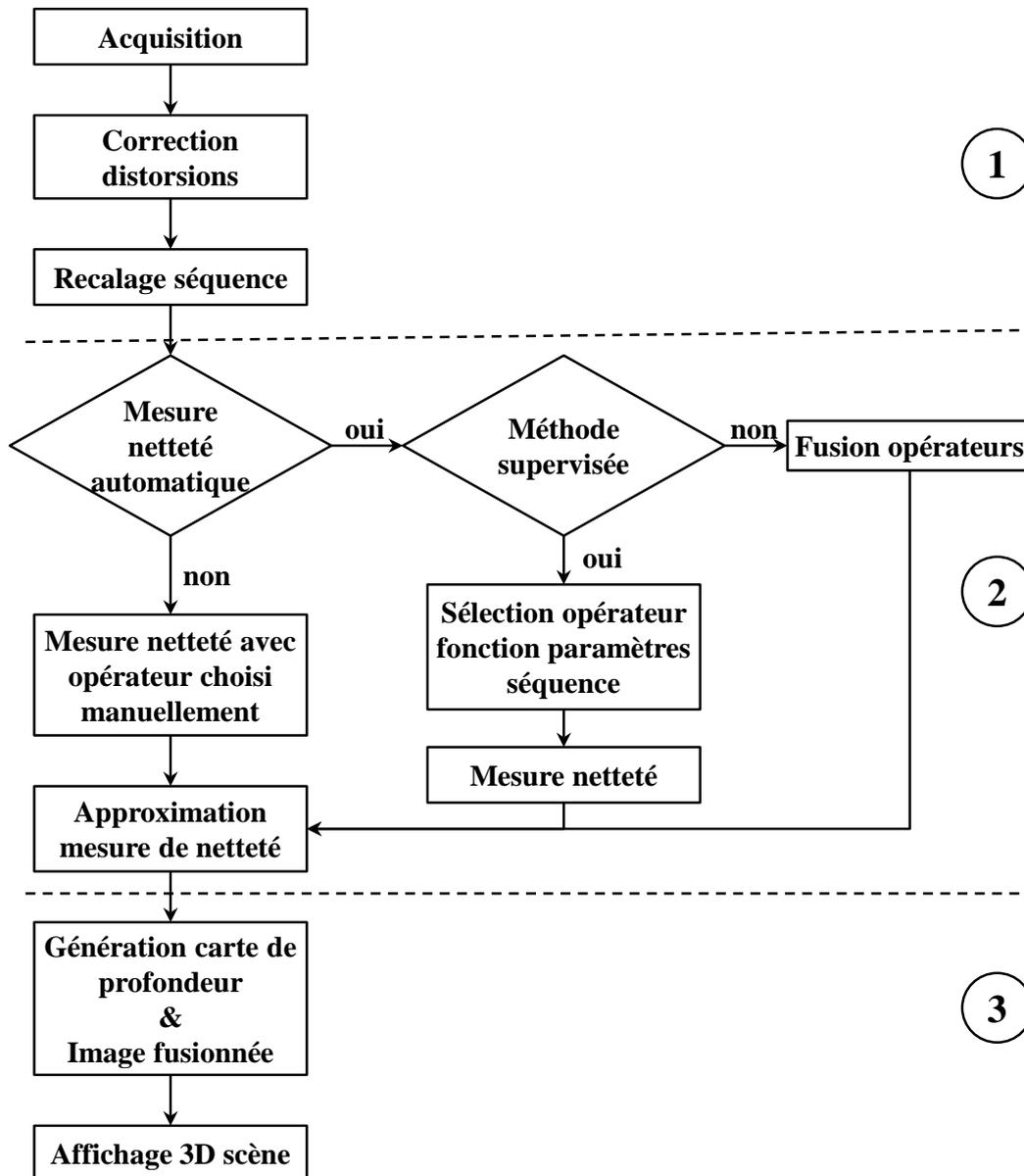


Figure 4.0.1: Schéma fonctionnel du processus complet d'estimation de la profondeur.

le déplacement du plan focal. La sortie de ce processus fournit une séquence d'images pour lesquelles chaque pixel est associé au même point de la scène considérée et dont la netteté est variable tout au long de la séquence.

2. **Evaluation de la netteté pour chaque pixel de la séquence d'images.** Etape cruciale d'obtention d'une information de profondeur juste pour chaque point de la scène, elle fait l'objet de la principale intervention de l'utilisateur. En effet, cette étape nécessite le choix de la taille de voisinage à utiliser lors de la mesure de netteté pour chaque pixel ainsi que le choix du procédé de mesure. Ce procédé peut être manuel avec un choix de l'opérateur de mesure de netteté à utiliser par l'utilisateur ou automatique suivant une approche supervisée (détaillée section 3.5.2) ou non supervisée (détaillée section 3.5.1).
3. **Génération de la carte de profondeur et de l'image fusionnée.** Une fois l'information de netteté obtenue pour chaque pixel, deux mises en forme de cette information sont réalisables. La première est la génération de la carte de profondeur illustrant la profondeur en tout point de la scène par un codage en niveaux de gris ou en fausses couleurs. La seconde est la création d'une image entièrement nette de la scène. Avec ces deux nouvelles images, une projection tridimensionnelle de l'information de profondeur suivie du plaquage de l'image nette peuvent être réalisés. A noter que cette représentation est purement explicite pour une meilleure compréhension de la scène par l'utilisateur mais l'information utile reste bien entendue la profondeur numérique pour chaque point.

Cette thèse de doctorat n'étant pas spécifiquement liée à une application précise, le type de scène n'est donc pas imposé. Ainsi, le choix a été pris de tester l'ensemble de la méthode sur plusieurs types de scènes agronomiques dont les deux végétaux suivants : des épis de blé et la plante Millepertuis. De plus, afin d'élargir les champs d'application, une séquence d'images d'un circuit intégré a permis le test des étapes deux et trois de la méthode dans le cadre d'analyse de défaillance de composants électroniques. A noter que ces résultats d'estimation de la profondeur n'ont pas vocation à permettre

une mesure qualitative de cette estimation mais de vérifier la cohérence de l'analyse du comportement des opérateurs réalisée. En effet, une analyse de la précision de l'estimation est dépendante d'une application particulière engendrant le choix d'un système d'acquisition avec ses limites en termes de profondeur de champ et d'un opérateur de mesure de netteté adapté. Notre problématique de départ n'entraîne pas un objectif de précision à atteindre sur une application particulière mais bien le développement d'un système d'estimation de la profondeur d'une scène naturelle. C'est pourquoi le protocole d'analyse du comportement des opérateurs de mesure de netteté sur un jeu de données de scènes naturelles a été mis en place. L'étude de ces comportements en amont de l'utilisation du système nous permet de déterminer les opérateurs adaptés à une majorité de scènes naturelles différentes. Cependant, pour une application bien particulière, la vérification de la cohérence de l'estimation de profondeur pourrait être envisagée par une comparaison entre la topographie mesurée par un scanner 3D et celle mesurée par notre méthode. Cette comparaison se basera sur une métrique du type erreur quadratique ou une mesure de corrélation entre les profondeurs estimées. Les résultats d'estimation de profondeur présentés dans ce chapitre ont donc pour objectif de confirmer la validité de notre approche d'étude du comportement des opérateurs. Ceci permet une utilisation du processus de reconstruction sur un grand nombre de scènes naturelles différentes dès lors qu'une étude en amont permettant de mettre en évidence les opérateurs adaptés est réalisée.

4.1 Application au dénombrement des épis de blé

Cette première mise en application de la méthode de reconstruction a été réalisée sur des images contenant des épis de blé. La volonté d'obtenir l'information de profondeur de ce type de scène fait suite aux travaux antérieurs de mon laboratoire axés sur la détection et le comptage des épis de blé en vue d'une estimation précoce du rendement [Cointault et al., 2008]. L'estimation du rendement étant réalisée au moment de la moisson par des capteurs embar-

qués au sein des machines agricoles, elle permet seulement l'établissement de cartes de rendements utilisées pour adapter la pratique des agriculteurs l'année suivante. Une estimation précoce de ce rendement permettrait d'adapter certaines pratiques en cours de saison et, éventuellement, améliorer le rendement en fin de saison et de mieux gérer les parcelles. La composante principale de l'évaluation du rendement est le nombre d'épis par mètre carré. Un comptage automatique par vision artificielle de ces épis est donc la première étape vers la conception d'un système d'évaluation précoce de rendement et a pour but d'éviter un comptage manuel. L'étape de détection des épis de blé développée précédemment passe par la recherche automatique d'un espace colorimétrique de représentation des images acquises permettant une mise en évidence claire des épis. L'information couleur n'étant pas suffisante, une information de texture lui est adjointe et définissant un nouvel espace de représentation appelé espace hybride. L'espace hybride adapté mutualise l'information de l'espace colorimétrique et des paramètres d'analyse de texture facilitant cette mise en évidence. Une fois les épis détectés par cette étape, une seconde étape de nettoyage par morphologie mathématique des images segmentées permet d'éliminer les éventuels artefacts. La dernière étape est celle du comptage des épis pour laquelle un problème se pose lors du recouvrement de plusieurs épis dans l'image acquise. Bien que le recours à des algorithmes, tels que la squelettisation, permettent de différencier ces épis dans certains cas, l'aide de l'information de profondeur est plus adaptée. En effet, un recouvrement signifie que les épis ne se situent pas dans le même plan spatial et être en possession de l'information de profondeur permet de discriminer ceux-ci lors du comptage. C'est donc cette problématique liée aux travaux antérieurs qui est à l'origine de ce projet de recherche.

4.1.1 Acquisitions par le système expérimental

Dans un premier temps, l'acquisition a été effectuée par le système expérimental (présenté partie 2.2) avec un pas entre chaque image de 1 cm. Un extrait d'images de la séquence originellement constituée de 30 images est illustré figure 4.1.1.



Figure 4.1.1: Extrait d'images de la séquence issue du système expérimental.

La mesure de netteté utilisée pour cette séquence d'images est celle basée sur la variance de l'amplitude du gradient (FM_{tenvar}) avec un voisinage de mesure de 11×11 pixels. Les résultats imagés de l'évaluation de la profondeur sont illustrés figure 4.1.2.

On notera l'application d'un lissage de la carte de profondeur avant projection lors de la présence de pics correspondant à une erreur ponctuelle d'estimation de la profondeur. Ce lissage est effectué par filtrage médian, suggéré par [Niederöst et al., 2002], des artefacts comme ceux entourés sur la carte de profondeur figure 4.1.2a.

4.1.2 Acquisitions par le système final

Comme présenté dans la section 2.3, le second système, constitué d'une caméra et d'un objectif motorisé, permet de s'affranchir du léger déplacement

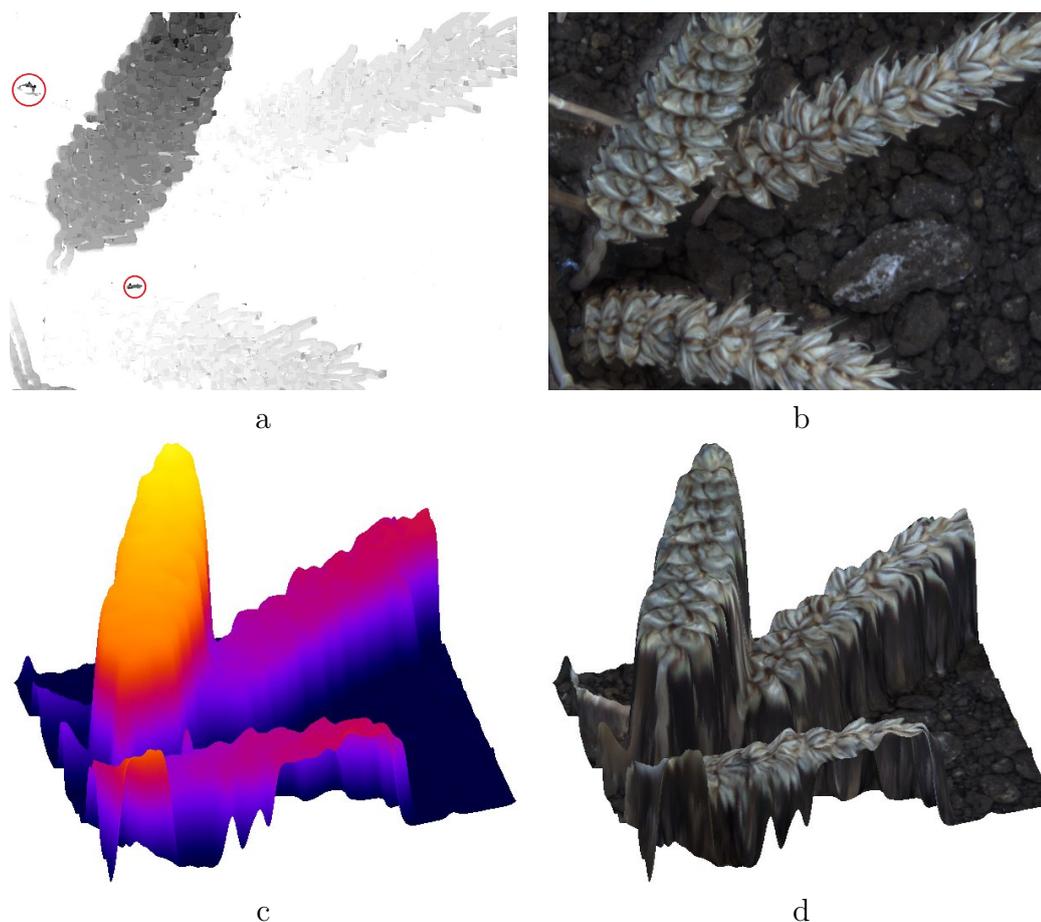


Figure 4.1.2: Reconstruction 3D par le système expérimental. a. Carte de profondeur. b. Image nette. c. Projection en fausses couleurs. d. Projection avec placage de texture.

du centre optique généré par les vibrations lors du déplacement. De plus, et s'inscrivant dans l'aspect automatique recherché de l'ensemble du processus de reconstruction, les images constituant la séquence sont acquises automatiquement et bien plus rapidement qu'avec le premier système. Ceci est dû au fait qu'une légère rotation de la bague de mise au point motorisée est plus rapide qu'un déplacement linéaire de l'ensemble du bloc optique. Du fait de la focale de 35 mm de cet objectif, le champ de vue est plus important que pour le premier système composé d'un objectif 50 mm.

En accord avec les résultats de l'étude du comportement des différents opérateurs de mesure de netteté présentés chapitre 3, on détaille les résul-

tats obtenus par l'opérateur basé sur la variance de l'amplitude du gradient (FM_{tenvar}). Les résultats obtenus par l'opérateur FM_{DGFsom} , notre opérateur basé sur les descripteurs généralisés de Fourier, et l'opérateur $FM_{entropy}$ mesurant la quantité d'information locale sont également illustrés afin d'avoir matière à comparaison. Les voisinages considérés pour les mesures de netteté sont 9×9 et 33×33 pixels afin de visualiser des résultats issus d'une petite zone d'intérêt et ceux d'une zone relativement importante. La différence d'un pixel par rapport aux voisinages utilisés lors du processus de comparaison (8×8 et 32×32 pixels) vient du fait que, au vu du nombre de calculs réalisés (19800 combinaisons de paramètres), l'incrémentatation n'était pas de un pixel mais d'un nombre correspondant au voisinage utilisé afin d'accélérer les calculs. Pour la comparaison, ceci ne pose pas de problème puisque l'intérêt n'est pas le rendu de la carte de profondeur mais la quantification des erreurs (taux de pixels faux dans les cartes de profondeur). Cependant, dans le cas de reconstruction 3D de scène réelle, l'incrémentatation est bien d'un pixel pour obtenir une estimation de profondeur propre à chaque pixel.

La séquence d'images acquises est constituée de 27 images, chacune espacée d'un pas de 1 cm dont un extrait est fourni figure 4.1.3. Les cartes de profondeur et les projections 3D associées pour un voisinage de 9×9 pixels sont illustrées figure 4.1.4 et figure 4.1.5 pour un voisinage 33×33 pixels.

Les résultats sont obtenus avec une interpolation gaussienne des courbes de netteté pour chaque pixel. L'information utile se retrouve à travers une matrice de valeurs pour chaque point de la scène. Ainsi, en fonction de l'application, les valeurs de cette matrice sont, par exemple, la distance entre le système d'acquisition et le point considéré ou encore la distance entre le sol et le point pour une acquisition verticale.

Le 1^{er} constat au regard des images résultats est la différence visuelle entre les résultats associés à une petite taille de voisinage (9×9 pixels) et un voisinage de plus grande dimension (33×33 pixels). En effet, comme stipulé dans le chapitre précédent, on remarque qu'une taille de voisinage de plus grande dimension entraîne moins d'artefacts d'erreur mais a pour effet de lisser la carte de profondeur. Ainsi, un plus petit voisinage, même s'il est sujet à plus d'artefacts permet une estimation plus fine des détails de la scène comme

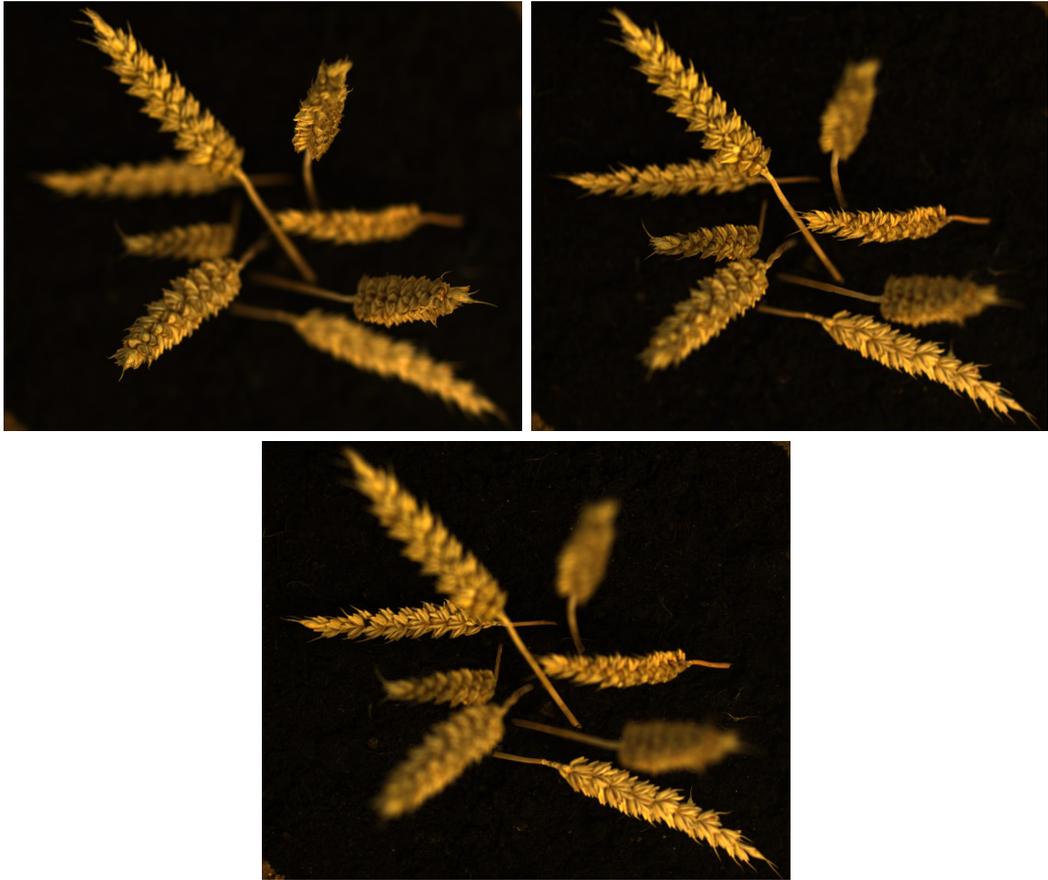


Figure 4.1.3: Extrait d'images de la séquence d'épis de blé issue du système final.

les grains ici. Dans notre cas, cette estimation plus fine met en évidence les grains de blé pour un petit voisinage contrairement à un voisinage élevé où la topographie est plus grossière mais moins soumise aux erreurs. Distinguer les grains se révèle intéressant par rapport à la problématique d'évaluation du rendement qui passe par le dénombrement des grains et l'étude de leur volume.

Globalement, on remarque visuellement que les différents opérateurs se comportent en accord avec les résultats des études statistiques. Ainsi, l'opérateur FM_{tenvar} fait état des meilleurs résultats pour les 2 tailles de voisinage différentes. L'opérateurs $FM_{DGF_{som}}$, malgré quelques erreurs plus importantes notamment au niveau des bordures, révèle un comportement relativement si-

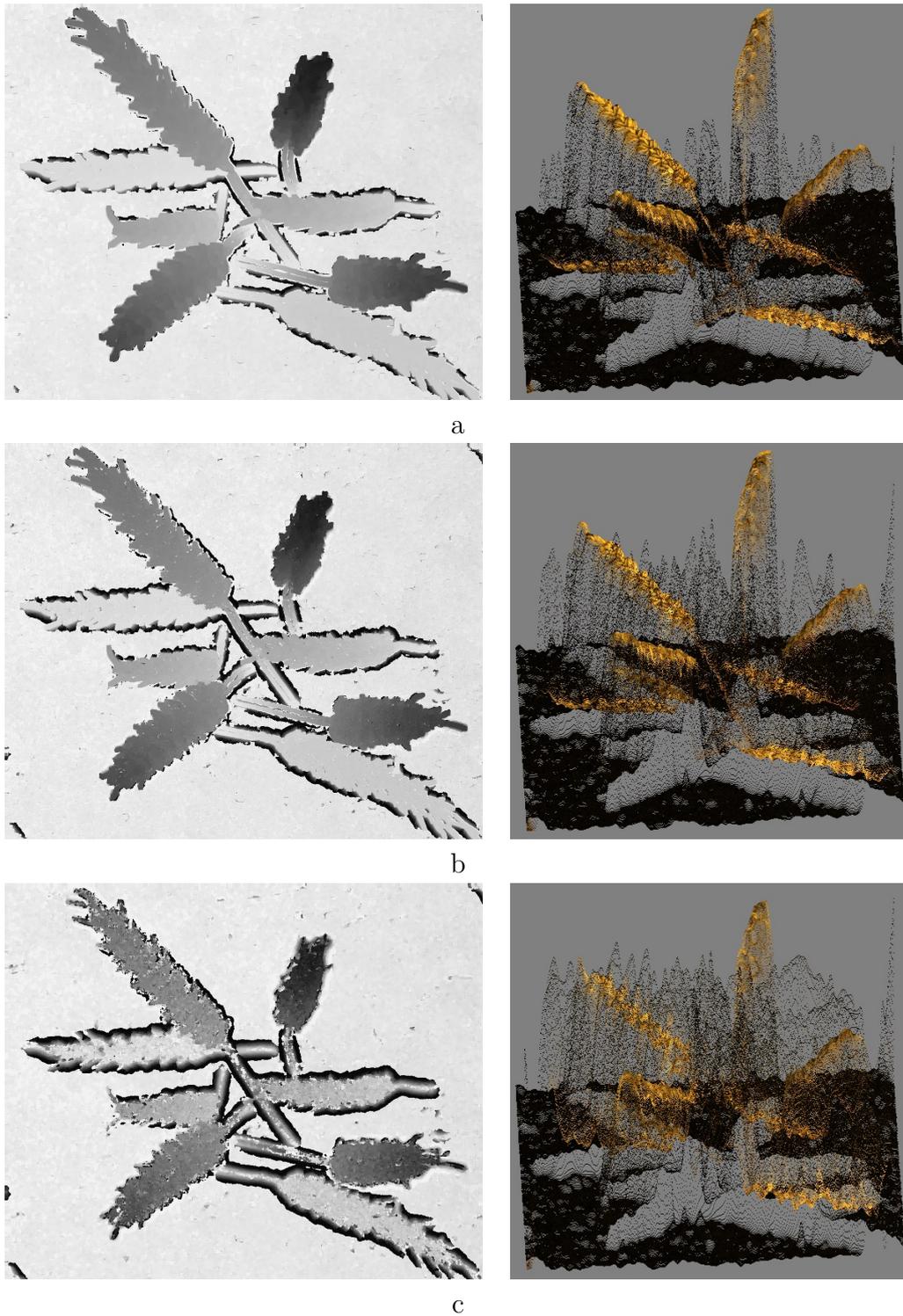


Figure 4.1.4: Cartes de profondeur et projection 3D d'une scène d'épis de blé pour un voisinage de 9×9 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$.

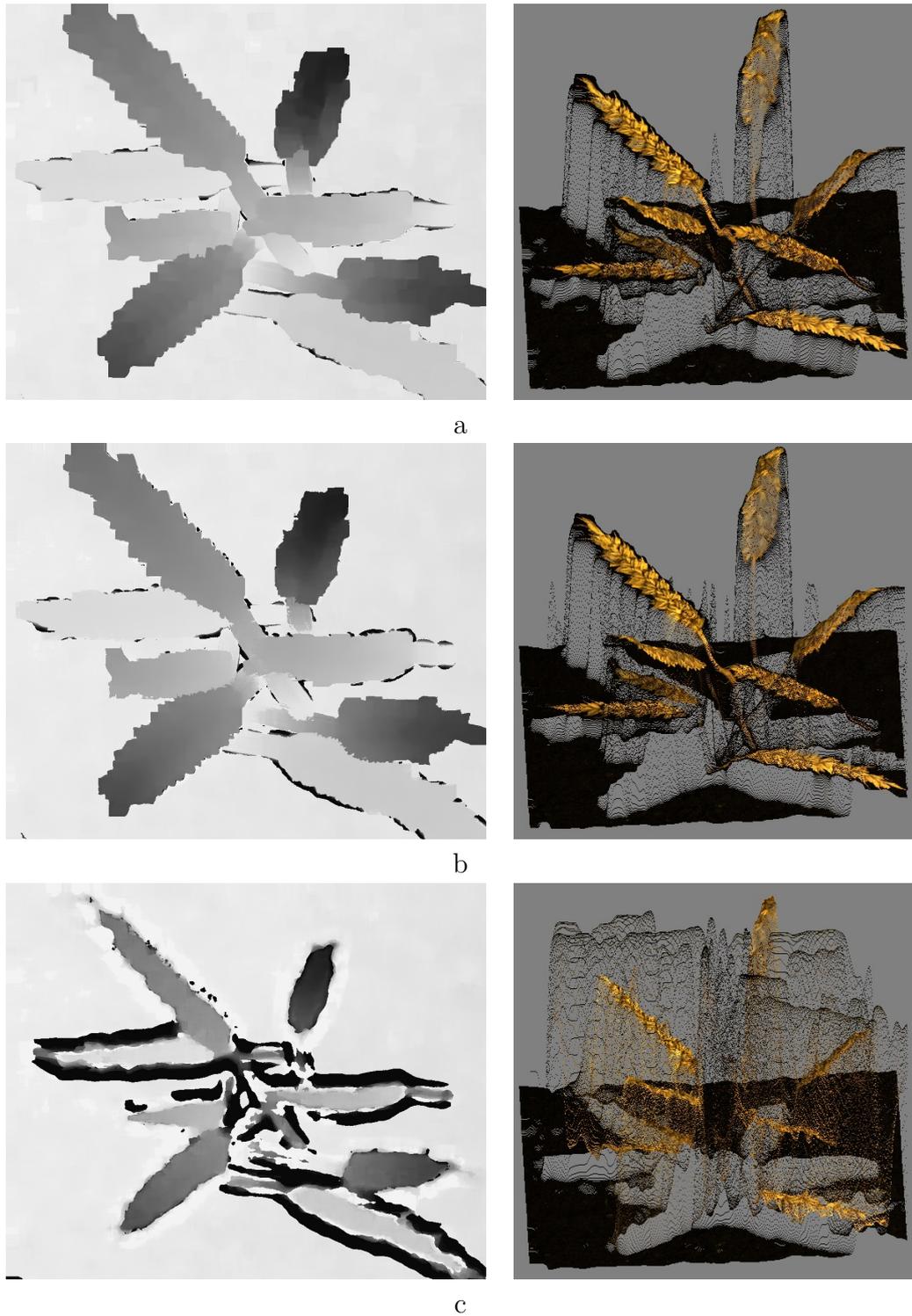


Figure 4.1.5: Cartes de profondeur et projection 3D d'une scène d'épis de blé pour un voisinage de 33×33 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$.

milaire à FM_{tenvar} . Le 3^{ème} opérateur $FM_{entropy}$, considéré médiocre par les études statistiques, présente un comportement effectivement bien moins adapté à une bonne estimation de la profondeur de la scène. La différence de résultats est moins flagrante pour un petit voisinage de mesure pour lequel on remarque surtout des erreurs importantes au niveau des bordures. Cependant, pour un voisinage de taille plus importante, beaucoup d'erreurs apparaissent entraînant une estimation imprécise de plusieurs zones de la scène considérée. Afin de visualiser plus facilement ces erreurs d'estimation sur les bordures, les superpositions des images nettes et des cartes de profondeur sont illustrées figure 4.1.6.

4.2 Application à l'étude de croissance de plante

Les problématiques récurrentes lors de l'étude des plantes sont la détection de maladies, le suivi de croissance ou encore l'estimation d'un volume foliaire. Ainsi, les scènes étudiées dans ce contexte sont plutôt des plantes en processus de croissance et constituées en partie de feuilles. Pour ce second type de scène, le choix s'est porté sur la plante Millepertuis dont un extrait d'images acquises par le système final se trouve figure 4.2.1. L'intérêt de cette scène réside dans la présence de feuilles à différentes profondeurs tout au long de la tige.

La séquence originale se compose de 22 images avec un espacement d'un centimètre. De la même façon que pour la séquence d'épis de blé, les résultats de 3 opérateurs de mesure de netteté sont comparés. Ainsi, les cartes de profondeur et les projections 3D obtenues pour un voisinage de 9×9 pixels sont illustrées figure 4.2.2. Pour un voisinage de 33×33 pixels, les résultats sont présentés figure 4.2.3.

Contrairement à la scène constituée d'épis de blé, celle-ci est nettement moins texturée et donc, présente moins de composantes hautes fréquences. Cependant, on constate que l'estimation de profondeur reste cohérente face à la scène visualisée et le dégradé de profondeur des différentes feuilles. Un suivi de croissance automatique de plantes est donc envisageable. On remarque

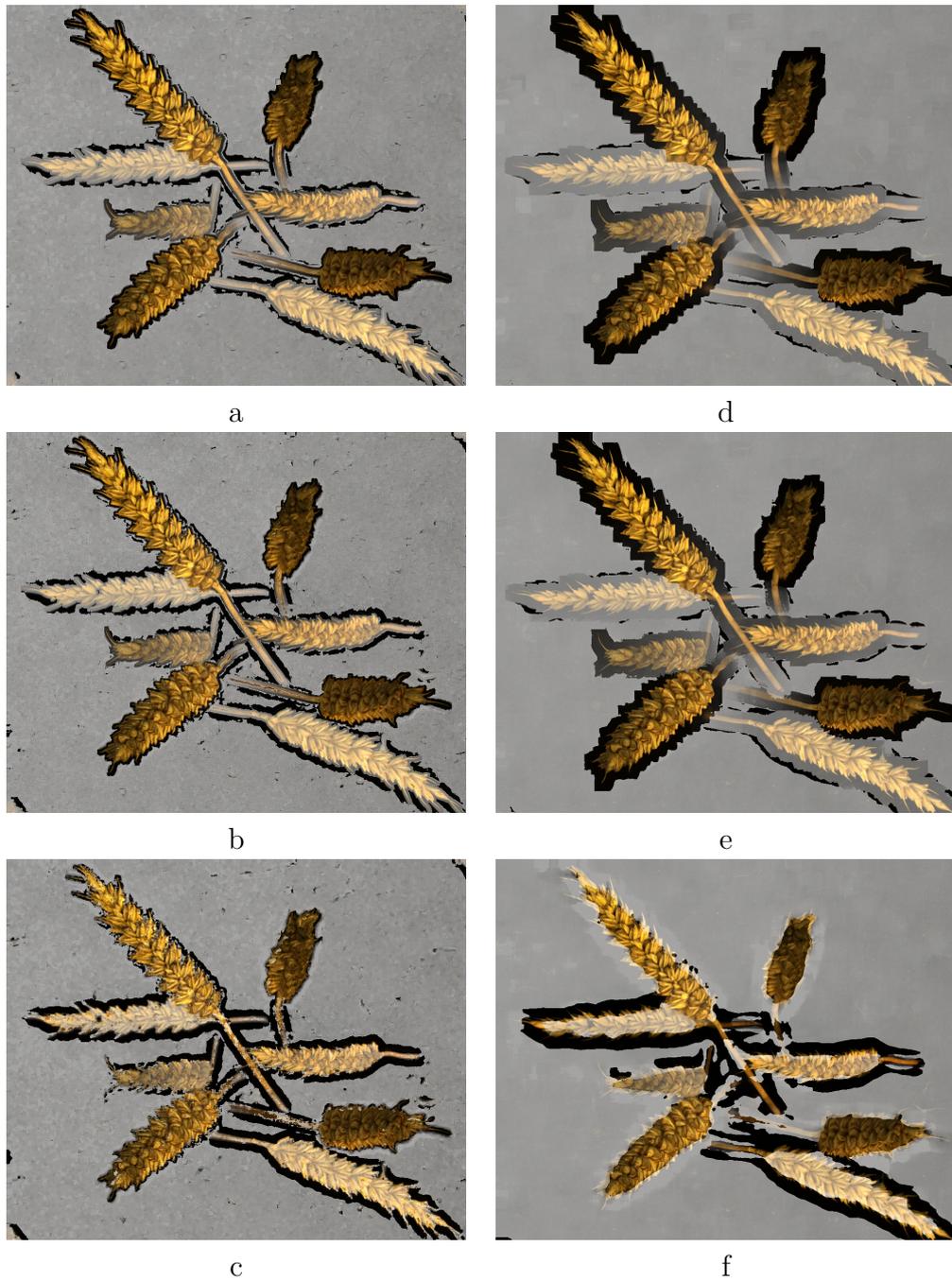


Figure 4.1.6: Superposition de l'image nette à la carte de profondeur. a. FM_{tenvar} et voisinage 9×9 . b. $FM_{DGF_{som}}$ et voisinage 9×9 . c. $FM_{entropy}$ et voisinage 9×9 . d. FM_{tenvar} et voisinage 33×33 . e. $FM_{DGF_{som}}$ et voisinage 33×33 . f. $FM_{entropy}$ et voisinage 33×33 .



Figure 4.2.1: Extrait d'images de la séquence de la plante Millepertuis issue du système final.

que, quelle que soit la taille de voisinage, cela entraîne moins de différences sur le résultat que pour une scène plus texturée. Toutefois, on note toujours une réduction des artefacts d'erreur et des contours de feuilles moins bien discriminés pour une taille de voisinage plus élevée (33×33 pixels) facilement visualisables figure 4.2.4.

Pour ce qui est de la comparaison visuelle des résultats de chaque opérateur, les résultats sont toujours en accord avec les comportements mis en avant par les études statistiques. Ainsi, l'opérateur $FM_{entropy}$ est nettement moins efficace pour un voisinage important mais plus proche de la réalité pour un petit voisinage. Les résultats sont cependant moins bons que pour les deux autres opérateurs. De même que pour la séquence des épis de blé, notre opérateur $FM_{DGF_{som}}$ reste très proche dans ses estimations de la topographie de l'opérateur FM_{tenvar} identifié comme étant le plus performant.

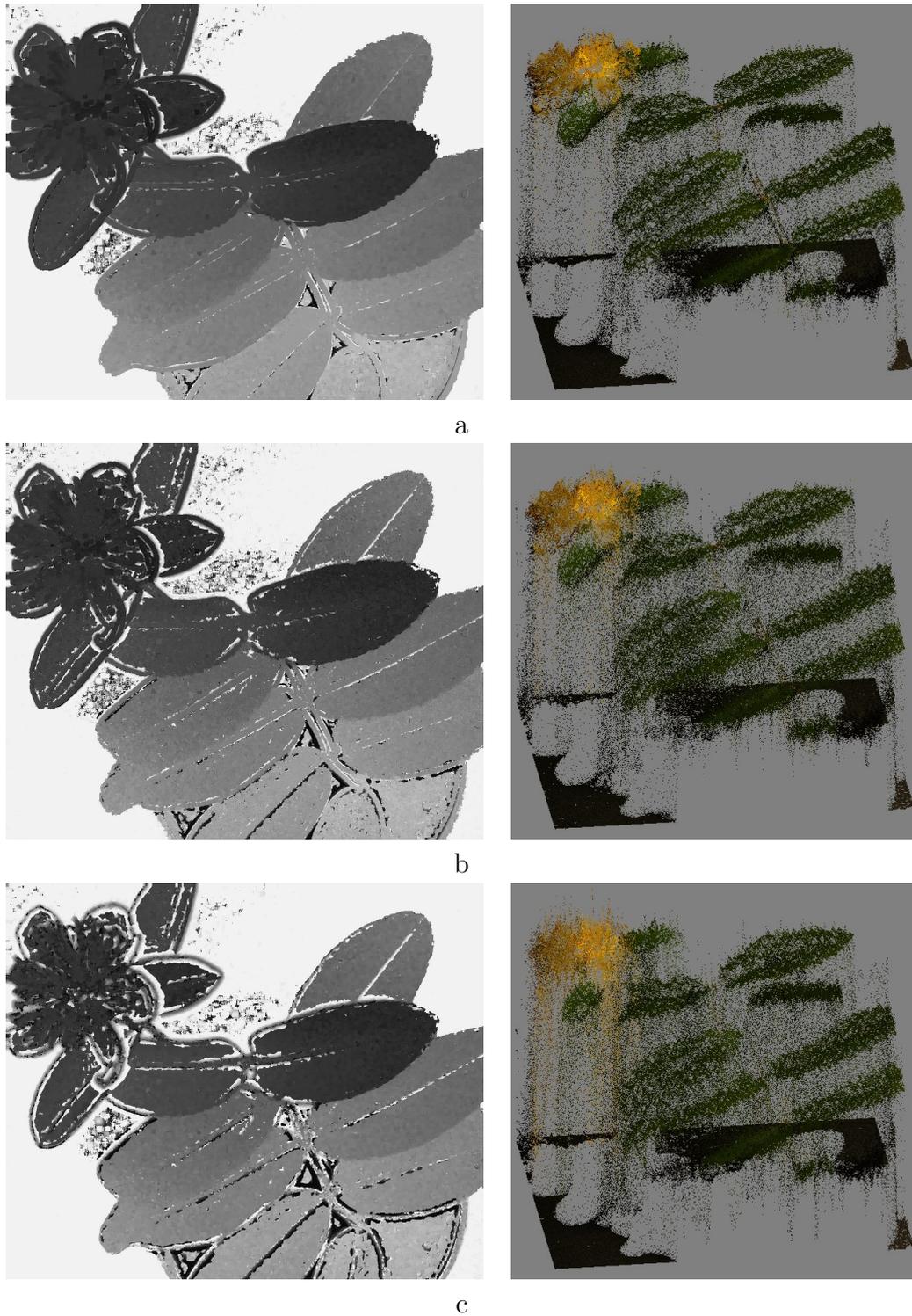


Figure 4.2.2: Cartes de profondeur et projection 3D scène Millepertuis pour un voisinage de 9×9 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$.

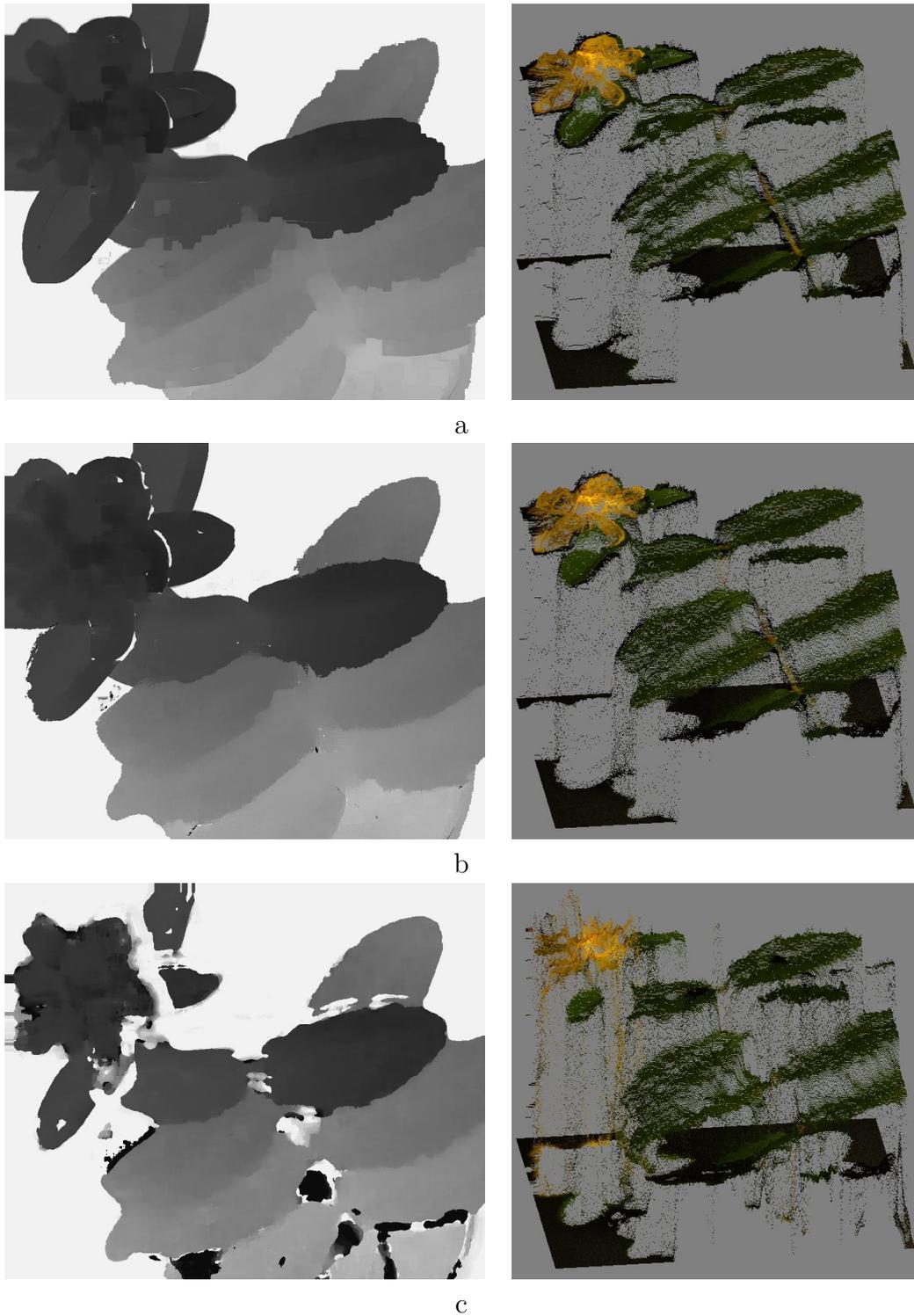


Figure 4.2.3: Cartes de profondeur et projection 3D scène Millepertuis pour un voisinage de 33×33 pixels. a. FM_{tenvar} . b. FM_{DGFsom} . c. $FM_{entropy}$.

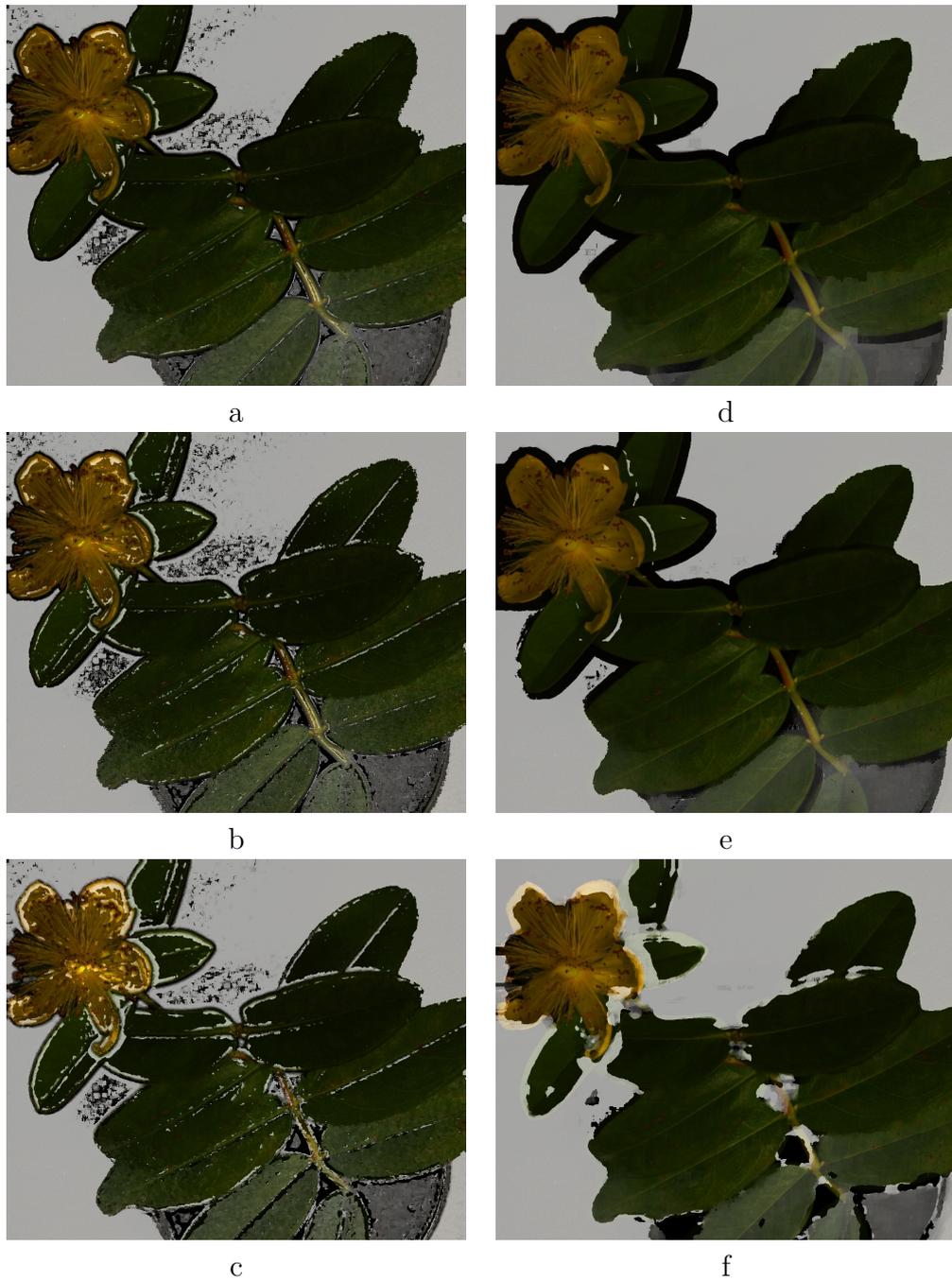


Figure 4.2.4: Superposition de l'image nette à la carte de profondeur. a. FM_{tenvar} et voisinage 9×9 . b. $FM_{DGF_{som}}$ et voisinage 9×9 . c. $FM_{entropy}$ et voisinage 9×9 . d. FM_{tenvar} et voisinage 33×33 . e. $FM_{DGF_{som}}$ et voisinage 33×33 . f. $FM_{entropy}$ et voisinage 33×33 .

4.3 Application à l'analyse de défaillances de circuits intégrés

Initialement, la méthode « Shape from Focus » était principalement utilisée pour des applications microscopiques. En effet, la faible profondeur de champ due à l'utilisation de microscopes nécessite l'emploi de cette méthode que ce soit pour générer une profondeur de champ étendue de la scène considérée ou estimer la topographie d'objets microscopiques. Ce besoin de profondeur de champ étendue apparaît dans le domaine de l'analyse de défaillances de circuits intégrés. L'analyse de défaillances d'un circuit correspond au processus intervenant en aval des tests de robustesse et de qualité de conception si ceux-ci ne sont pas concluants. Le constat de défaillance issu des tests entraîne le besoin de déterminer si le circuit considéré est un cas isolé ou un problème récurrent. L'analyse de défaillances du circuit revient à extraire des informations pertinentes telles que la localisation des défauts. Cette dernière s'appuie sur une image physique du composant (nommée *pattern*) sur laquelle est superposée l'information mesurée (telle que les zones d'émissions caractérisant un défaut). Cette image de *pattern* permet non seulement à l'utilisateur de se repérer dans le composant mais également de réaliser la mise au point sur la zone visualisée.

La localisation des défauts est rendue possible par l'utilisation d'outils d'observation adaptés tels que la microscopie à émission de lumière par le biais de systèmes dédiés à l'analyse de défaillances comme le système TriPHEMOS (Hamamatsu Photonics, Hamamatsu, Japon). Ce type d'appareil intègre plusieurs méthodes de localisation de défauts qui nécessitent toujours une étape de préparation du circuit à étudier. Cette phase passe par différentes actions qui, du fait de la faible profondeur de champ du système d'acquisition, exigent une profondeur de champ étendue afin d'obtenir une visualisation nette de la zone de localisation des défauts. Ces différentes actions créant un flou sur l'image de *pattern* sont :

1. **Polissage de la face arrière.** Cette première étape a pour but de réduire l'épaisseur de substrat en silicium afin de permettre une meilleure

transmission des rayons infrarouges. En effet, le silicium a la particularité de se comporter comme un milieu transparent pour cette gamme de longueurs d'onde. Cependant, la traversée du substrat par le faisceau engendre tout de même des perturbations des signaux réfléchis donc il convient de réduire ce substrat afin de limiter les perturbations. L'inconvénient de ce polissage est le fait que la surface polie n'est pas parfaitement plane. Ce défaut peut entraîner l'apparition de zones floues dans les images du circuit étudié.

2. **Couplage du circuit à une carte de stimulation électrique.** La localisation de défauts nécessitant un circuit en cours de fonctionnement, celui-ci est connecté à une carte de stimulation. Cependant, à l'image des défauts de polissage, cette carte peut former des zones floues du fait que l'ensemble de la scène visualisée n'est pas parfaitement perpendiculaire au système d'acquisition.
3. **Mise au point sur la zone étudiée.** Suite aux étapes précédentes, le circuit est placé dans l'appareil d'analyse et son positionnement est effectué visuellement de façon à considérer la zone étudiée du circuit. Une fois positionné, la faible profondeur de champ induite par le système optique ne permet pas l'obtention d'une zone nette en tout point de la surface analysée du fait de l'épaisseur variable des éléments constituant la partie du circuit analysé.

Cette dernière étape nécessite une visualisation qui est rendue possible, dans le cas présent, par une caméra basée sur un capteur InGaAs (Arséniure d'indium-gallium) disposant d'une sensibilité spectrale en proche infrarouge de 900 nm à 1550 nm. La définition des images acquises est de 640 par 512 pixels avec des photo-éléments de 20 μm de large. Cette caméra a permis l'acquisition de la séquence d'images mise à ma disposition par le laboratoire d'expertise du CNES (Centre National d'Etudes Spatiales). L'objectif utilisé conjointement à la caméra permet un grossissement 20 \times et une variation de la distance de mise au point entre chaque acquisition génère la séquence d'images. Un extrait de cette dernière constituée initialement de 35 images est illustré figure 4.3.1 et représente une partie d'un microcontrôleur (ARM Cortex M4 STM32).

On remarque sur ces extraits, un déplacement de la zone de netteté du

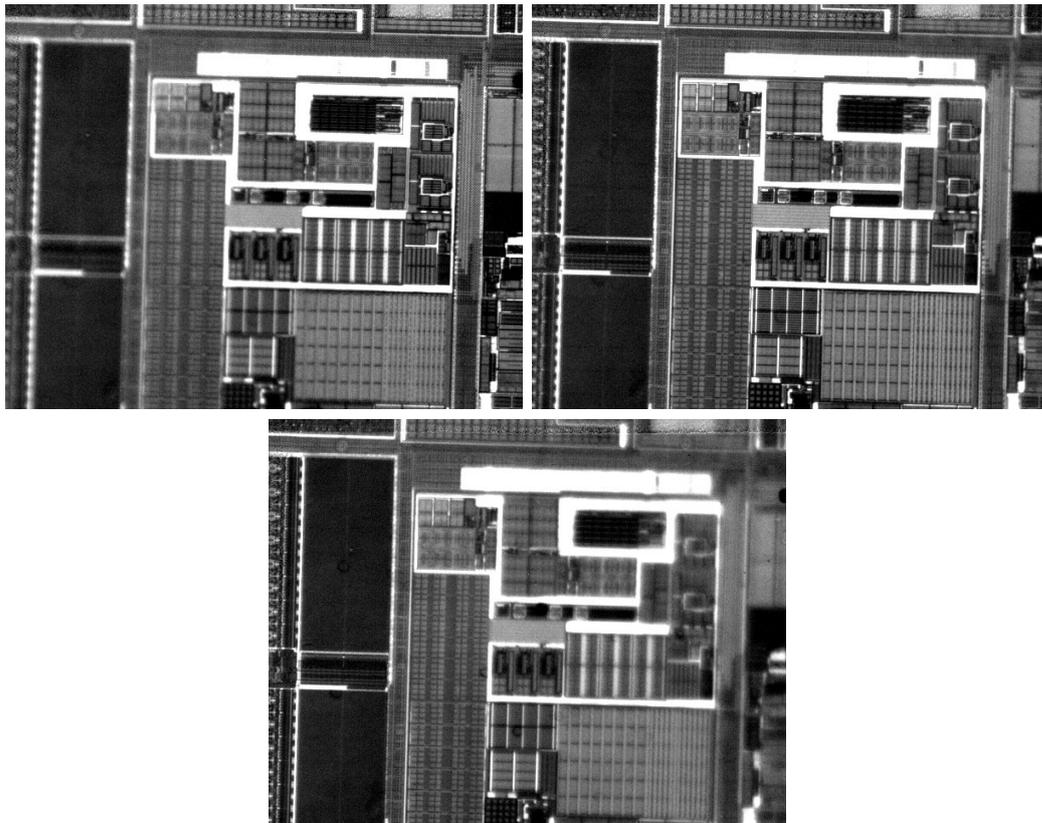


Figure 4.3.1: Extrait d'images de la séquence du microcontrôleur.

coin en haut à gauche pour la 1^{ère} image jusqu'au coin en bas à droite sur la 3^{ème}. De façon à obtenir une image entièrement nette de la scène et donc une extension de la profondeur de champ, les étapes 2 et 3 de notre processus d'estimation de la profondeur (figure 4.0.1) sont mises en œuvre. De la même façon que pour les applications précédentes, les résultats de l'image de profondeur de champ étendue issus des 3 opérateurs de mesure de netteté sont présentés pour un voisinage de 9×9 et 33×33 (figure 4.3.2). Cette scène est bien différente des scènes précédemment étudiées du fait de l'aspect artificiel et manufacturé de celle-ci. Cependant, le caractère texturé est toujours présent mais très variable d'une zone à l'autre du circuit.

L'opérateur FM_{tenvar} génère une image entièrement nette faisant état de peu d'artefacts et ceci pour les 2 tailles de voisinage. Notre opérateur $FM_{DGF_{som}}$ produit une image nette proche de celle obtenue par l'opérateur FM_{tenvar} pour

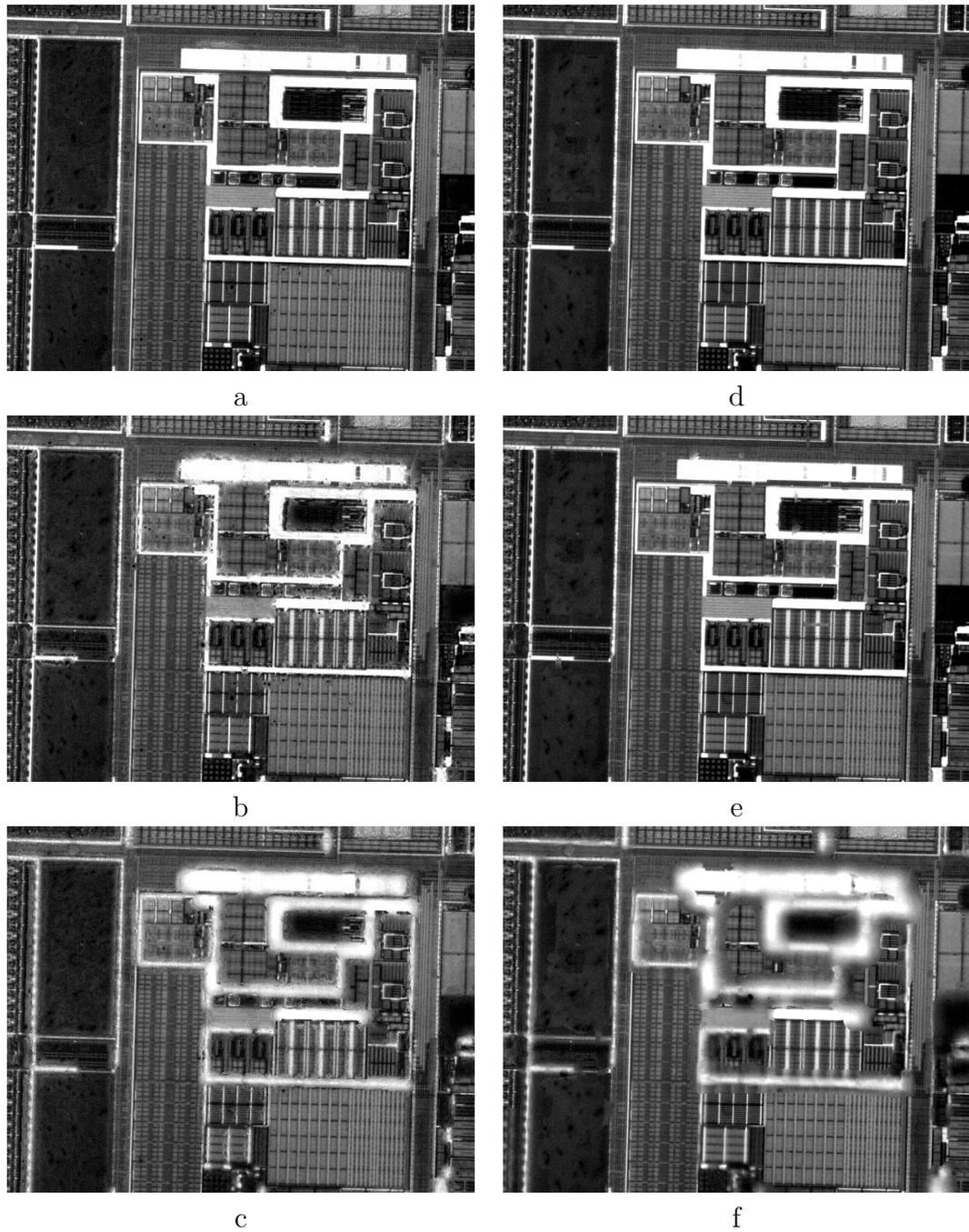


Figure 4.3.2: Images de profondeur de champ étendue d'une partie de micro-contrôleur. a. FM_{tenvar} et voisinage 9×9 . b. FM_{DGFsom} et voisinage 9×9 . c. $FM_{entropy}$ et voisinage 9×9 . d. FM_{tenvar} et voisinage 33×33 . e. FM_{DGFsom} et voisinage 33×33 . f. $FM_{entropy}$ et voisinage 33×33 .

un voisinage de 33×33 pixels mais ce qui n'est plus le cas pour un voisinage plus petit où la présence d'erreurs d'estimation est plus importante. Pour ce qui est de l'opérateur $FM_{entropy}$, les résultats reflètent encore un manque de justesse de l'estimation des pixels nets par rapport aux autres opérateurs. Même si le type de scène change, les comportements des opérateurs suivent toujours ceux mis en évidence dans le chapitre précédent.

Une image entièrement nette de la scène analysée est donc réalisable par le processus de reconstruction. Cette extension de la profondeur de champ permet donc de pallier aux différents défauts induits lors de la préparation du circuit intégré étudié et de proposer une visualisation nette en tout point à l'utilisateur.

4.4 Conclusion

Ce chapitre présente tout d'abord l'ensemble du processus d'estimation de la profondeur mis en place basé sur la méthode « Shape from Focus ». Deux applications agronomiques correspondant à 2 scènes différentes sont présentées avec les résultats obtenus pour 3 opérateurs de mesure de netteté et 2 tailles de voisinage différents. Le processus d'estimation de la profondeur est ensuite appliqué à une problématique d'analyse de défaillances de circuit intégré qui nécessite une profondeur de champ étendue. Cette dernière est rendue possible par la création d'une image entièrement nette de la zone du circuit étudié à partir d'une séquence d'images à différentes distances de mise au point.

Quels que soit les résultats des applications, on constate que le classement des 3 opérateurs étudiés est toujours le même. Les comportements réels des opérateurs valident le processus d'étude statistique mis en place au chapitre précédent permettant une prédiction du comportement. Ainsi, l'opérateur basé sur la mesure de variance de l'amplitude du gradient (FM_{tenvar}) reste l'opérateur présentant les meilleurs résultats. Notre opérateur basé sur les descripteurs généralisés de Fourier $FM_{DGF_{som}}$ fait cependant état de bonnes performances notamment pour les scènes agronomiques. Comme prévu par les études statistiques, l'opérateur mesurant la quantité locale d'information ($FM_{entropy}$) n'est pas un opérateur permettant une bonne estimation de la profondeur. Le protocole d'étude du comportement des opérateurs présenté au chapitre précédent, s'il est effectué pour le type de scène considérée par l'utilisateur, permet de déterminer le ou les opérateurs adaptés au processus d'estimation de la profondeur. L'avantage principal est une prédiction en amont permettant une estimation de la profondeur de scènes inconnues et de pallier au manque de vérité terrain.

Les résultats obtenus pour différents types de scène confirment le postulat de départ qui est que la méthode « Shape from Focus » est pertinente tant que la scène est texturée. En effet, l'estimation de profondeur sur une scène texturée comme celle comportant les épis de blé est moins affectée par les erreurs qu'une scène moins texturée comme la plante Millepertuis. Cependant, les images résultantes de l'estimation de profondeur de cette dernière sont tout

de même cohérentes à la scène visualisée, ce qui confirme la faisabilité de la méthode sur ce type de scène.

A noter que les résultats présentés n'ont pas fait l'objet d'un post-traitement visant à diminuer les erreurs afin de visualiser l'efficacité des opérateurs. Ceci permet également de mettre en évidence le compromis à faire entre une petite taille de voisinage impliquant une estimation de la profondeur moins lissée et une taille plus importante diminuant les erreurs ponctuelles d'estimation. Bien entendu, en fonction de la qualité des résultats obtenus pour une scène particulière avec la taille de voisinage ainsi que l'opérateur adapté, la mise en place de traitements de minimisation des erreurs peut être envisagée. On peut, par exemple, envisager des traitements du type morphologie mathématique permettant une suppression des artefacts ponctuels apparaissant en cas d'erreur. Cependant, il convient de toujours garder à l'esprit que de tels traitements créent une modification de l'information pouvant dégrader la qualité globale des données initiales.

Conclusion générale et perspectives

Conclusion générale

Ce mémoire de doctorat présente le travail effectué sur la conception d'un système de reconstruction 3D, basé sur la méthode « Shape from Focus », pour lequel les applications visées sont en priorité agronomiques. Une notion récurrente apparaît tout au long du processus de réflexion ayant permis d'aboutir au système final : la notion de compromis. C'est en effet au travers de cette notion que sont retrouvés les différents apports de cette recherche.

Le premier compromis concerne le choix de la méthode de reconstruction 3D utilisée. Ce choix a nécessité bien entendu une étude approfondie des différentes méthodes existantes permettant l'obtention de l'information de profondeur, étude qui constitue l'état de l'art présenté dans ce manuscrit. Les avantages et inconvénients des méthodes ainsi que les contraintes liées aux scènes naturelles, telles que la présence d'occlusions, ont permis de faire ressortir la méthode « Shape from Focus » considérée comme la plus adaptée dans notre cas. De plus, cette méthode se révèle intéressante d'un point de vue recherche, de par sa mise en œuvre peu courante à une échelle macroscopique. Cet état de l'art permet ainsi une vision plus claire des différents moyens d'accès à l'information de profondeur pour les personnes non spécialisées dans ce domaine. De plus, il convient de rappeler que le choix de la méthode à utiliser est principalement lié aux contraintes induites par l'application visée telles que la taille de la scène analysée, la vitesse d'acquisition souhaitée ou

l'encombrement du système conçu.

Le second compromis intervient lors du dimensionnement du système d'acquisition à utiliser. La précision de l'estimation de profondeur est directement liée à la profondeur de champ induite par le système d'acquisition. En effet, plus celle-ci est faible et plus l'incertitude quant à la position nette d'un point particulier dans la séquence d'images est faible. En contrepartie, une forte diminution de la profondeur de champ va engendrer une diminution des dimensions de la scène visualisée. Ceci vient du fait que la variation de plusieurs paramètres du système d'acquisition dont la distance focale et la distance de mise au point entraînent une réduction de la profondeur de champ mais également une diminution du champ de vue. La réduction du champ de vue est liée à ces deux derniers paramètres dès lors qu'une faible profondeur de champ est requise. Ces notions sont détaillées dans le deuxième chapitre où une mise en évidence de ce lien entre champ de vue et profondeur de champ est illustré et permet d'orienter l'utilisateur pour le choix de l'objectif adapté à son application.

La diminution de la profondeur de champ engendre un autre compromis entre le nombre d'images acquises et la vitesse d'acquisition. Plus la profondeur de champ est faible, plus la précision du système est importante mais plus le nombre d'images de la scène à acquérir augmente afin de couvrir une même profondeur dans la scène considérée.

L'étape d'acquisition de la séquence d'images étant fondamentale pour une estimation correcte de la profondeur, deux systèmes d'acquisition permettant l'acquisition de la séquence d'images de la scène ont été conçus et présentés dans le chapitre 2. Ils possèdent tous les deux des points forts et des points faibles et, de fait, l'utilisation de l'un ou l'autre est défini en fonction de l'application visée. Le premier système à déplacement linéaire du système d'acquisition permet un changement d'objectif en fonction de la scène visualisée ou de la précision en profondeur requise. Cependant, le déplacement linéaire exige un temps d'acquisition de la séquence d'images long, notamment pour des scènes profondes, mais également des vibrations nécessitant un recalage des images de façon à ce qu'à chaque point de la scène soit associée la même coordonnée de pixel pour chaque image. Le second système à objectif motorisé

s'affranchit de ce problème de vibrations tout en accélérant la procédure d'acquisition de la séquence. De plus, son encombrement fortement réduit permet un transport et une utilisation plus aisée sur le terrain. Cependant, le choix des caractéristiques optiques pour ce type d'objectif est plus restreint que pour des objectifs standards. Il existe des objectifs motorisés permettant une variation de la distance focale mais la complexité de conception de ces objectifs est souvent liée à un diamètre d'ouverture plus faible. Ceci a pour effet de produire une profondeur de champ plus importante et donc de diminuer la précision. Les deux systèmes d'acquisition conçus sont donc complémentaires et permettent une adaptabilité face aux applications futures.

La seconde étape fondamentale du processus d'estimation de la profondeur est l'évaluation de la netteté locale de chaque pixel de la séquence d'images. Le chapitre 3 y est consacré et présente en premier lieu un état de l'art des différentes familles d'opérateurs de mesure de netteté. À ces opérateurs de mesure de netteté existants sont ajoutés deux nouveaux opérateurs développés durant ces travaux de recherche et basés sur les descripteurs généralisés de Fourier. L'idée principale dans le cadre du « Shape from Focus » est l'estimation de la netteté d'un pixel de mêmes coordonnées dans chaque image correspondant à un point unique de la scène. Pour un opérateur, une courbe de netteté est obtenue dont le maximum renseigne sur la position de l'image de la séquence où le point de la scène visualisé est le plus net. La profondeur réelle de ce point dans l'espace peut donc être assimilée à la profondeur de cette image de la séquence, connue par l'étape de calibrage du système d'acquisition. Pour une estimation de la profondeur plus fine, une étape d'interpolation de la courbe de netteté permet d'approximer la position du maximum de netteté.

Du fait du grand nombre d'opérateurs existants, une comparaison de l'efficacité de ceux-ci est nécessaire afin de déterminer lesquels sont les plus adaptés pour un cas particulier. Le type de scène considéré dans notre problématique, à savoir les scènes agronomiques, ne s'apparente pas aux images habituellement utilisées pour la comparaison. Un protocole d'étude du comportement des opérateurs basé sur un plan d'expériences factoriel est présenté. Contrairement aux comparaisons usuelles de la littérature, cette étude permet une analyse du comportement des opérateurs face aux variations de deux para-

mètres affectant la mesure de netteté. Ces deux paramètres sont la taille de voisinage locale considérée pour la mesure et la présence ou non de bruit additif au voisinage. En effet, les comparaisons usuelles renseignent sur la robustesse des opérateurs face à une variation graduelle et simultanée des deux paramètres. Notre protocole d'étude permet une mise en évidence de la robustesse des opérateurs face à l'un ou l'autre des paramètres. L'intérêt est de pouvoir constater qu'un opérateur peut être robuste à un paramètre mais sensible à un autre et sera plus adapté dans un cas particulier qu'un opérateur de robustesse moyenne à tous les paramètres. L'étude est effectuée sur une base d'images adaptée à l'application visée. Dans notre cas, la base d'images de texture de scènes naturelles VisTex est utilisée [VisTex, 2002]. Ceci rend possible une étude du comportement des opérateurs objective en fonction de l'application et non une comparaison générique sur une base d'images de synthèse. À noter que la notion de compromis se retrouve dans le choix de la taille de voisinage car une petite taille autorise une résolution spatiale plus fine de l'estimation mais entraîne cependant plus d'erreurs qu'une taille de voisinage plus importante.

Dans le cadre d'une automatisation du processus d'estimation de la profondeur, deux méthodes (non supervisée et supervisée) d'évaluation de la netteté sont proposées. La première se base sur la fusion de l'estimation de profondeur issue de plusieurs opérateurs différents afin de pallier aux erreurs d'estimation des uns et des autres. La seconde se base sur une analyse approfondie de l'efficacité des opérateurs sur la base d'images VisTex afin de proposer l'opérateur le plus adapté en fonction des paramètres de taille de voisinage et de présence ou non de bruit additif. Cette méthode est très importante dans notre cas. En effet, le processus d'estimation de la profondeur est voué à une utilisation sur des scènes naturelles ponctuelles et uniques. Ceci entraîne un manque de vérité terrain permettant de quantifier l'efficacité des différents opérateurs de mesure de netteté. L'analyse approfondie du comportement des opérateurs permet de pallier à ce manque de vérité terrain par un choix en amont de l'opérateur à utiliser en fonction du type de scène, de la taille de voisinage souhaitée et du taux de bruit présent dans la séquence d'images.

Enfin, le chapitre 4 illustre les résultats d'estimation de la profondeur pour

plusieurs types de scènes différentes propres à des applications particulières. L'illustration de ces résultats met en évidence un comportement analogue des opérateurs de mesure de netteté entre les résultats de la reconstruction 3D et les constats effectués lors de l'analyse approfondie du comportement des opérateurs. Ainsi, un de nos opérateurs basés sur les descripteurs généralisés de Fourier, même s'il n'est pas considéré comme le plus efficace des opérateurs étudiés, fait état d'une efficacité supérieure à la majorité des autres opérateurs. Il se positionne ainsi comme une alternative ou un complément lorsque l'opérateur de référence fait état d'une diminution de l'efficacité lors de l'estimation de netteté.

Pour résumer, les apports principaux de cette thèse sont la conception de deux systèmes d'acquisition de séquences d'images, le développement de deux nouveaux opérateurs de mesure de netteté ainsi que la mise en place d'un protocole de comparaison de ces opérateurs original et adapté à l'application. Ces différents apports contribuent au processus complet et automatisé de reconstruction 3D et permettent de disposer d'un dispositif entièrement fonctionnel.

Perspectives

Les perspectives de ces travaux sont non seulement l'amélioration du processus de reconstruction actuel mais également son utilisation à des applications spécifiques.

Concernant l'amélioration de l'existant, cela consiste à surmonter les limites de la méthode de reconstruction. Les principales sont le champ de vue réduit, dès lors qu'une profondeur de champ très faible est requise, ainsi que le temps de traitement élevé, causé par la mesure de netteté appliquée à chaque pixel de chaque image de la séquence. Dès lors que le champ de vue n'est pas suffisamment étendu, une solution est la mutualisation des systèmes d'acquisition et la reconstruction de panoramas afin de recréer une séquence d'images de la scène. Un autre solution est liée à l'évolution des objectifs permettant une diminution de la focale tout en garantissant un diamètre d'ouverture élevé pour

réduire la profondeur de champ. C'est le cas d'un objectif de focale 25 mm et d'ouverture relative $f0.95$, à l'état de prototype au début de mes recherches, qui est maintenant commercialisé par différents fabricants (Thalès Angénieux, France et Voigtlander, Allemagne). Cet objectif permet donc un champ de vue important de part sa focale de 25 mm tout en limitant la profondeur de champ grâce à une ouverture relative minimum très faible (une ouverture $f0.95$ entraîne deux fois plus de lumière reçue par le capteur qu'une ouverture $f1.4$, la profondeur de champ s'en retrouve donc réduite). Pour ce qui est du temps de traitement élevé de la séquence d'images, la solution passe par une implémentation parallélisée des algorithmes. On pourra s'orienter vers une implémentation de type GPGPU (General-purpose Processing on Graphics Processing Units) afin de bénéficier de l'architecture des cartes graphiques plus adaptée au traitement de données telles que les images. À noter que notre processus de reconstruction est très facilement parallélisable puisque le même traitement est appliqué à chaque image de la séquence. Une implémentation parallèle est donc bien plus adaptée dans ce cas qu'une implémentation séquentielle.

Outre l'utilisation des systèmes conçus au sein du laboratoire dans le cadre d'applications futures, on notera une adaptabilité du processus de reconstruction à d'autres systèmes d'acquisition. C'est le cas des plateformes de phénotypage succinctement exposées en introduction, où les systèmes d'acquisition reposent pour la plupart sur une base « caméra et objectif motorisé ». Une implémentation de la méthode est donc réalisable directement, permettant l'estimation de la profondeur en tout point de la scène visualisée sans ajout de matériel. Cette information de profondeur est nécessaire dans le cadre de phénotypage de plante afin de proposer un suivi de croissance ou de renseigner sur la disposition spatiale des différents éléments des plantes analysées. De plus, en prenant l'exemple de la plateforme de phénotypage de Dijon, chaque cellule d'acquisition est équipée de deux systèmes d'acquisition. On pourrait ainsi envisager une estimation de la profondeur de la plante en vues de dessus et de côté puis une fusion de l'information 3D afin de pallier aux limites de visibilité liées à l'utilisation d'un seul système (plusieurs vues entraînent une diminution des zones cachées). Cette fusion des données de profondeur passe par l'utilisation d'algorithmes de type ICP (Iterative Closest Point) [Besl et

[McKay, 1992](#)] et d'autres méthodes dont un état de l'art est réalisé par [[Matbosch et al., 2005](#)].

Publications

Revues

- "3D Image Acquisition System Based on Shape from Focus Technique", Bastien Billiot, Frédéric Cointault, Ludovic Journaux, Jean-Claude Simon, Pierre Gouton, Sensors, 2013.

Conférences internationales et nationales

- "Mesure de netteté basée sur les descripteurs généralisés de Fourier appliquée à la reconstruction 3D par Shape from Focus", Bastien Billiot, Frédéric Cointault, Pierre Gouton, 24ème colloque GRETSI, Brest, France, 2013.
- "3D acquisition system applied to agronomic scenes", Bastien Billiot, Pierre Gouton, Frédéric Cointault, International Conference on Precision Agriculture, Indianapolis, États-Unis, 2012.
- « Reconstruction monoculaire 3D de scènes agronomiques », Bastien Billiot, Frédéric Cointault, Pierre Gouton, ORASIS - Congrès des jeunes chercheurs en vision par ordinateur, Praz-sur-Arly, France, 2011.

Workshop

- « Imagerie 3D – PPHD Dijon, Présentation & Applications », Bastien Billiot, Frédéric Cointault, Colloque INRA, Montpellier, France, 2012.

-
- « L'utilisation de la proxi-détection 3D pour la caractérisation du blé », Bastien Billiot, Frédéric Cointault, Atelier proxi-détection, Montpellier, France, 2012.
 - « Système d'acquisition dédié à la reconstruction 3D de scènes agronomiques », Bastien Billiot, Forum des Jeunes Chercheurs, Dijon, France, 2011.

Annexe



Figure 4.4.1: Base de texture VisTex

Bibliographie

- [Adams, 2002] Adams, M. (2002). Coaxial range measurement-current trends for mobile robotic applications. *Sensors Journal, IEEE*, 2(1) :2–13. *Cité page 20*
- [AER, 2013] AER (2013). <http://www.cmm.com.sg>. *2 citations pages xviii et 11*
- [Ahmad et Choi, 2007] Ahmad, M. B. et Choi, T. S. (2007). Application of three dimensional shape from image focus in lcd/tft displays manufacturing. *Consumer Electronics, IEEE Transactions on*, 53(1) :1–4. *Cité page 70*
- [An et al., 2008] An, Y., Kang, G., Kim, I.-J., Chung, H.-S., et Park, J. (2008). Shape from focus through laplacian using 3d window. In *Future Generation Communication and Networking, 2008. FGCN'08. Second International Conference on*, volume 2, pages 46–50. IEEE. *Cité page 70*
- [Asif et Choi, 2001] Asif, M. et Choi, T. (2001). Shape from focus using multilayer feedforward neural networks. *Image Processing, IEEE Transactions on*, 10(11) :1670–1675. *Cité page 80*
- [Bailey, 2002] Bailey, T. (2002). *Mobile Robot Localisation and Mapping in Extensive Outdoor Environments*. PhD thesis. *Cité page 24*
- [Barnard et Fischler, 1982] Barnard, S. et Fischler, M. (1982). Computational stereo. *ACM Computing Surveys (CSUR)*, 14(4) :553–572. *Cité page 20*
- [Batlle et al., 1998] Batlle, J., Mouaddib, E., et Salvi, J. (1998). Recent progress in coded structured light as a technique to solve the correspondence problem : a survey. *Pattern recognition*, 31(7) :963–982. *Cité page 16*

- [Besl et McKay, 1992] Besl, P. J. et McKay, N. D. (1992). Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics. *Cité page 133*
- [Bouguet, 2010] Bouguet, J. (2010). Camera calibration toolbox. *Cité page 56*
- [Brenner et al., 1976] Brenner, J., Dew, B., Horton, J., King, T., Neurath, P., et Selles, W. (1976). An automated microscope for cytologic research a preliminary evaluation. *Journal of Histochemistry & Cytochemistry*, 24(1) :100. *Cité page 70*
- [Carrhill et Hummel, 1985] Carrhill, B. et Hummel, R. (1985). Experiments with the intensity ratio depth sensor. *Computer Vision, Graphics, and Image Processing*, 32(3) :337–358. *3 citations pages xiii, 18 et 19*
- [Caspi et al., 1998] Caspi, D., Kiryati, N., et Shamir, J. (1998). Range imaging with adaptive color structured light. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(5) :470–480. *Cité page 17*
- [Chane et al., 2012] Chane, C. S., Mansouri, A., Marzani, F. S., et Boochs, F. (2012). Integration of 3d and multispectral data for cultural heritage applications : Survey and perspectives. *Image and Vision Computing*. *Cité page 15*
- [Chéné et al., 2012] Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., Belin, É., et Chapeau-Blondeau, F. (2012). On the use of depth camera for 3d phenotyping of entire plants. *Computers and Electronics in Agriculture*, 82 :122–127. *Cité page 17*
- [Cheung et al., 2003] Cheung, G., Baker, S., et Kanade, T. (2003). Visual hull alignment and refinement across time : A 3d reconstruction algorithm combining shape-from-silhouette with stereo (pdf). *Cité page 24*
- [Choi et Yun, 2000] Choi, T. et Yun, J. (2000). Three-dimensional shape recovery from the focused-image surface. *Optical Engineering*, 39 :1321. *Cité page 80*
- [Clerc et Mallat, 2002] Clerc, M. et Mallat, S. (2002). The texture gradient equation for recovering shape from texture. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4) :536–549. *2 citations pages xiii et 22*

- [Cointault et al., 2008] Cointault, F., Guérin, D., Guillemin, J., et Chopinet, B. (2008). In-field wheat ears counting using colour-texture image analysis. *New Zealand Journal of Crop and Horticultural Science*, 36(2) :117–130. *Cité page 106*
- [Courteille, 2006] Courteille, F. (2006). *Vision monoculaire : contributions théoriques et application à la numérisation des documents*. PhD thesis. *Cité page 23*
- [Darrell et Wohn, 1988] Darrell, T. et Wohn, K. (1988). Pyramid based depth from focus. pages 504–509. *Cité page 58*
- [Doyon, 2000] Doyon, D. (2000). *Scanner à rayons X : tomodensitométrie*. Elsevier Masson. *Cité page 12*
- [Doyon, 2004] Doyon, D. (2004). *IRM : imagerie par résonance magnétique*. Elsevier Masson. *Cité page 12*
- [Falco, 2010] Falco, P. (2010). La photographie. *Cité page 42*
- [Faugeras, 1993] Faugeras, O. (1993). *Three-dimensional computer vision : a geometric viewpoint*. the MIT Press. *Cité page 21*
- [Favaro et al., 2003] Favaro, P., Mennucci, A., et Soatto, S. (2003). Observing shape from defocused images. *International Journal of Computer Vision*, 52(1) :25–43. *Cité page 25*
- [Favaro et Soatto, 2002] Favaro, P. et Soatto, S. (2002). Learning shape from defocus. In *Computer Vision—ECCV 2002*, pages 735–745. Springer. *2 citations pages xviii et 26*
- [Firestone et al., 1991] Firestone, L., Cook, K., Culp, K., Talsania, N., et Preston Jr, K. (1991). Comparison of autofocus methods for automated microscopy. *Cytometry*, 12(3) :195–206. *2 citations pages 72 et 80*
- [Forsyth, 2002] Forsyth, D. (2002). Shape from texture without boundaries. page III : 225 ff. *Cité page 21*
- [Gauthier et al., 1991] Gauthier, J.-P., Bornard, G., et Silbermann, M. (1991). Motions and pattern analysis : harmonic analysis on motion groups and their homogeneous spaces. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(1) :159–172. *Cité page 75*

- [Gholipour et al., 2007] Gholipour, A., Kehtarnavaz, N., Briggs, R., Devous, M., et Gopinath, K. (2007). Brain functional localization : a survey of image registration techniques. *Medical Imaging, IEEE Transactions on*, 26(4) :427–451. *Cité page 52*
- [Grimson, 1981] Grimson, W. E. L. (1981). A computer implementation of a theory of human stereo vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, pages 217–253. *Cité page 7*
- [Groen et al., 1985] Groen, F., Young, I., et Ligthart, G. (1985). A comparison of different focus functions for use in autofocus algorithms. *Cytometry*, 6(2) :81–91. *Cité page 72*
- [Haber-Pohlmeier et Pohlmeier, 2010] Haber-Pohlmeier, S. et Pohlmeier, A. (2010). A new window to the rhizosphere by magnetic resonance imaging. In *EGU General Assembly Conference Abstracts*, volume 12, page 15371. *Cité page 12*
- [Haralick et al., 1973] Haralick, R. M., Shanmugam, K., et Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6) :610–621. *Cité page 75*
- [Hartley, 2008] Hartley, R. (2008). *Multiple view geometry in computer vision*. Cambridge University Press. *Cité page 20*
- [Hasinoff et Kutulakos, 2009] Hasinoff, S. W. et Kutulakos, K. N. (2009). Confocal stereo. *Int. J. Comput. Vision*, 81 :82–104. *Cité page 28*
- [Helmli et Scherer, 2001] Helmli, F. et Scherer, S. (2001). Adaptive shape from focus with an error estimation in light microscopy. In *Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on*, pages 188–193. IEEE. *Cité page 74*
- [Horn, 1970] Horn, B. K. (1970). Shape from shading : A method for obtaining the shape of a smooth opaque object from one view. Technical report, Cambridge, MA, USA. *Cité page 22*
- [Inokuchi et al., 1984] Inokuchi, S., Sato, K., et Matsuda, F. (1984). Range imaging system for 3-d object recognition. In *Proceedings of the International Conference on Pattern Recognition*, pages 806–808. *Cité page 16*

- [Ishii et al., 2007] Ishii, I., Yamamoto, K., Tsuji, T., et al. (2007). High-speed 3d image acquisition using coded structured light projection. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 925–930. IEEE. *Cité page 17*
- [Jolliffe, 2005] Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library. *Cité page 77*
- [Journaux et al., 2011] Journaux, L., Simon, J.-C., Destain, M.-F., Cointault, F., Miteran, J., et Piron, A. (2011). Plant leaf roughness analysis by texture classification with generalized fourier descriptors in different dimensionality reduction context. *Precision Agriculture*, 12(3) :345–360. *Cité page 75*
- [Kanatani et Chou, 1989] Kanatani, K. et Chou, T. (1989). Shape from texture : General principle. 38(1) :1–48. *Cité page 21*
- [Kirk, 1995] Kirk, R. E. (1995). *Experimental design : procedures for the behavioral sciences*. *Cité page 95*
- [Krähenbühl et al., 2012] Krähenbühl, A., Kerautret, B., Debled-Rennesson, I., Longuetaud, F., et Mothe, F. (2012). Knot detection in x-ray ct images of wood. In *Advances in Visual Computing*, pages 209–218. Springer. *Cité page 12*
- [Kristan et al., 2006] Kristan, M., Pers, J., Perse, M., et Kovacic, S. (2006). A bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform. *Pattern Recognition Letters*, 27(13) :1431–1439. *Cité page 73*
- [Krotkov, 1987] Krotkov, E. (1987). Focusing. *International journal of computer vision*, 1(3) :223–237. *2 citations pages 71 et 72*
- [Kusada, 1996] Kusada, Y. (1996). Focus detecting device having light distribution detection. US Patent 5,485,003. *Cité page 68*
- [Laurentini, 1994] Laurentini, A. (1994). The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 150–162. *Cité page 24*
- [Le Chevalier, 2002] Le Chevalier, F. (2002). *Principles of radar and sonar signal processing*. Artech House Publishers. *Cité page 13*

- [Lippmann, 1908] Lippmann, G. (1908). Epreuves reversibles. photographies integrals. *Comptes-Rendus Academie des Sciences*, 146 :446–451.
Cité page 29
- [Mahmood et Choi, 2008] Mahmood, M. et Choi, T. (2008). A feature analysis approach to estimate 3d shape from image focus. In *Image Processing, 2008. ICIIP 2008. 15th IEEE International Conference on*, pages 3216–3219. IEEE.
Cité page 73
- [Malik et Choi, 2008] Malik, A. et Choi, T. (2008). A novel algorithm for estimation of depth map using image focus for 3d shape recovery in the presence of noise. *Pattern Recognition*, 41(7) :2200–2225.
3 citations pages xv, 73 et 83
- [Marr et Poggio, 1977] Marr, D. et Poggio, T. (1977). A theory of human stereo vision. Technical report, Cambridge, MA, USA. Cité page 7
- [Maruyama et Abe, 1993] Maruyama, M. et Abe, S. (1993). Range sensing by projecting multiple slits with random cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(6) :647–651.
3 citations pages xiii, 17 et 18
- [Matabosch et al., 2005] Matabosch, C., Salvi, J., Fofi, D., et Meriaudeau, F. (2005). Range image registration for industrial inspection. In *Electronic Imaging 2005*, pages 216–227. International Society for Optics and Photonics.
Cité page 133
- [Mattia et al., 2003] Mattia, F., Le Toan, T., Picard, G., Posa, F. I., D’Alessio, A., Notarnicola, C., Gatti, A. M., Rinaldi, M., Satalino, G., et Pasquariello, G. (2003). Multitemporal c-band radar measurements on wheat fields. *Geoscience and Remote Sensing, IEEE Transactions on*, 41(7) :1551–1560.
Cité page 13
- [McBratney, 2000] McBratney, A. (2000). Pv or not pv? In *5th International Symposium on Cool Climate Viticulture and Oenology, 16-20 January 2000, Melbourne, Australia [Archivo de ordenador] : proceedings of the International Symposium on Grapevine Phylloxera Management*. Winetitles.
Cité page 2
- [Minhas et al., 2009] Minhas, R., Mohammed, A. A., Wu, Q. J., et Sid-Ahmed, M. A. (2009). 3d shape from focus and depth map computation

- using steerable filters. In *Image Analysis and Recognition*, pages 573–583. Springer. *Cité page 74*
- [Miyasaka et al., 2000] Miyasaka, T., Kuroda, K., Hirose, M., et Araki, K. (2000). High speed 3-d measurement system using incoherent light source for human performance analysis. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING*, 33(B5/2; PART 5) :547–551. *Cité page 18*
- [Nair et Stewart, 1992] Nair, H. et Stewart, C. (1992). Robust focus ranging. In *Computer Vision and Pattern Recognition, 1992. Proceedings CV-PR'92., 1992 IEEE Computer Society Conference on*, pages 309–314. IEEE. *Cité page 60*
- [Nayar, 1989] Nayar, S. K. (1989). Shape from focus. Technical report. *3 citations pages 26, 71 et 79*
- [Ng, 2006] Ng, R. (2006). *Digital light field photography*. PhD thesis, stanford university. *Cité page 29*
- [Niederöst et al., 2002] Niederöst, M., Niederöst, J., et Scucky, J. (2002). Automatic 3d reconstruction and visualization of microscopic objects from a monoscopic multifocus image sequence. In *International Workshop on Visualization and Animation of Reality based 3D Models*. Citeseer. *2 citations pages 80 et 108*
- [Noguchi et Nayar, 1994] Noguchi, M. et Nayar, S. (1994). Microscopic shape from focus using active illumination. pages A :147–152. *Cité page 69*
- [Omasa et al., 2007] Omasa, K., Hosoi, F., et Konishi, A. (2007). 3d lidar imaging for detecting and understanding plant responses and canopy structure. *Journal of Experimental Botany*, 58(4) :881–898. *Cité page 20*
- [Pajes et Forest, 2004] Pajes, Jordi, S. J. et Forest, J. (2004). A new optimised de bruijn coding strategy for structured light patterns. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 284–287. IEEE. *3 citations pages xiii, 17 et 18*
- [Pech-Pacheco et al., 2000] Pech-Pacheco, J., Cristobal, G., Chamorro-Martinez, J., et Fernandez-Valdivia, J. (2000). Diatom autofocusing in brightfield microscopy : a comparative study. In *icpr*, page 3318. Published by the IEEE Computer Society. *Cité page 71*

- [Pentland, 1987] Pentland, A. (1987). A new sense for depth of field. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (4) :523–531.
2 citations pages 25 et 38
- [Pentland et al., 1994] Pentland, A., Scherock, S., Darrell, T., et Girod, B. (1994). Simple range cameras based on focal error. *JOSA A*, 11(11) :2925–2934.
Cité page 25
- [Pertuz et al., 2012] Pertuz, S., Puig, D., et Angel Garcia, M. (2012). Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*.
Cité page 81
- [Pollefeys et al., 2004] Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., et Koch, R. (2004). Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59 :207–232.
Cité page 23
- [Posdamer et Altschuler, 1982] Posdamer, J. et Altschuler, M. (1982). Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1) :1–17. 2 citations pages xiii et 16
- [Pribanić et al., 2009] Pribanić, T., Džapo, H., et Salvi, J. (2009). Efficient and low-cost 3d structured light system based on a modified number-theoretic approach. *EURASIP Journal on Advances in Signal Processing*, 2010(1) :474389.
3 citations pages xiii, 17 et 18
- [Rocchini et al., 2001] Rocchini, C., Cignoni, P., Montani, C., Pingi, P., et Scopigno, R. (2001). A low cost 3d scanner based on structured light. In *Computer Graphics Forum*, volume 20, pages 299–308. Wiley Online Library.
Cité page 16
- [Rovira-Más et al., 2008] Rovira-Más, F., Zhang, Q., et Reid, J. F. (2008). Stereo vision three-dimensional terrain maps for precision agriculture. *Computers and Electronics in Agriculture*, 60(2) :133–143. Cité page 21
- [Salvi et al., 2010] Salvi, J., Fernandez, S., Pribanic, T., et Llado, X. (2010). A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8) :2666–2680. Cité page 16
- [Salvi et al., 2004] Salvi, J., Pages, J., et Batlle, J. (2004). Pattern codification strategies in structured light systems. 37(4) :827–849.
2 citations pages 16 et 17

- [Seugling, 2000] Seugling, R. M. (2000). Coordinate measuring machines. Technical report. *Cité page 10*
- [Shen et Chen, 2006] Shen, C. et Chen, H. (2006). Robust focus measure for low-contrast images. In *Consumer Electronics, 2006. ICCE'06. 2006 Digest of Technical Papers. International Conference on*, pages 69–70. IEEE. *2 citations pages 73 et 74*
- [Smach et al., 2008] Smach, F., Lemaître, C., Gauthier, J., Miteran, J., et Atri, M. (2008). Generalized fourier descriptors with applications to objects recognition in svm context. *Journal of Mathematical Imaging and Vision*, 30(1) :43–71. *Cité page 75*
- [Subbarao, 1988] Subbarao, M. (1988). Parallel depth recovery by changing camera parameters. In *Proceedings of the International Conference on Computer Vision*, pages 149–155. *Cité page 25*
- [Subbarao et Choi, 1995] Subbarao, M. et Choi, T. (1995). Accurate recovery of three-dimensional shape from image focus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(3) :266–274. *Cité page 80*
- [Subbarao et al., 1993] Subbarao, M., Choi, T., et Nikzad, A. (1993). Focusing techniques (journal paper). *Optical Engineering*, 32(11) :2824–2836. *Cité page 70*
- [Tajima et Iwakawa, 1990] Tajima, J. et Iwakawa, M. (1990). 3-d data acquisition by rainbow range finder. In *Pattern Recognition, 1990. Proceedings., 10th International Conference on*, volume 1, pages 309–313. IEEE. *3 citations pages xviii, 18 et 19*
- [Tenenbaum, 1970] Tenenbaum, J. (1970). *Accommodation in computer vision*. PhD thesis. *Cité page 70*
- [Thévenaz et al., 1998] Thévenaz, P., Ruttimann, U., et Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing*, 7(1) :27–41. *Cité page 52*
- [Tsai, 1987] Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of*, 3(4) :323–344. *Cité page 56*

- [VisTex, 2002] VisTex (2002). Vision texture database. Maintained by the Vision and Modeling group at the MIT Media Lab. web page :
<http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>.
2 citations pages 82 et 130
- [Vollath, 1987] Vollath, D. (1987). Automatic focusing by correlative methods. *J. Microsc.*, 147(3) :279–288. *Cité page 74*
- [Vuylsteke et Oosterlinck, 1990] Vuylsteke, P. et Oosterlinck, A. (1990). Range image acquisition with a single binary-encoded light pattern. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(2) :148–164.
3 citations pages xiii, 17 et 18
- [Watanabe et Nayar, 1995] Watanabe, M. et Nayar, S. K. (1995). Telecentric optics for constant magnification imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19 :1360–1365. *Cité page 59*
- [Willson et Shafer, 1991] Willson, R. et Shafer, S. (1991). Dynamic lens compensation for active color imaging and constant magnification focusing. Technical Report CMU-RI-TR-91-26, Robotics Institute, Pittsburgh, PA.
3 citations pages xiv, 58 et 59
- [Xie et al., 2006] Xie, H., Rong, W., et Sun, L. (2006). Wavelet-based focus measure and 3-d surface reconstruction method for microscopy images. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 229–234. IEEE. *Cité page 73*
- [Yang et Nelson, 2003] Yang, G. et Nelson, B. (2003). Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 3, pages 2143–2148. IEEE. *Cité page 73*
- [Zhang et al., 1999] Zhang, R., Tsai, P., Cryer, J., et Shah, M. (1999). Shape-from-shading : a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21 :690–706. *Cité page 22*
- [Zhang, 2000] Zhang, Z. (2000). A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11) :1330–1334. *Cité page 56*

[Zitova et Flusser, 2003] Zitova, B. et Flusser, J. (2003). Image registration methods : a survey. *Image and vision computing*, 21(11) :977–1000.

Cité page 52

[Zwaenepoel et al., 1997] Zwaenepoel, P., Le Bars, J., et al. (1997). L'agriculture de précision.

Cité page 1

Résumé

Dans le cadre de l'acquisition de l'information de profondeur de scènes texturées, un processus d'estimation de la profondeur basé sur la méthode de reconstruction 3D « Shape from Focus » est présenté dans ce manuscrit. Les deux étapes fondamentales de cette approche sont l'acquisition de la séquence d'images de la scène par sectionnement optique et l'évaluation de la netteté locale pour chaque pixel des images acquises. Deux systèmes d'acquisition de cette séquence d'images sont présentés ainsi que les traitements permettant d'exploiter celle-ci pour la suite du processus d'estimation de la profondeur. L'étape d'évaluation de la netteté des pixels passe par la comparaison des différents opérateurs de mesure de netteté. En plus des opérateurs usuels, deux nouveaux opérateurs basés sur les descripteurs généralisés de Fourier sont proposés. Une méthode nouvelle et originale de comparaison est développée et permet une analyse approfondie de la robustesse à différents paramètres des divers opérateurs. Afin de proposer une automatisation du processus de reconstruction, deux méthodes d'évaluation automatique de la netteté sont détaillées. Finalement, le processus complet de reconstruction est appliqué à des scènes agronomiques, mais également à une problématique du domaine de l'analyse de défaillances de circuits intégrés afin d'élargir les domaines d'utilisation.

Mots clés : Reconstruction 3D ; Shape from Focus ; Mesure de netteté ; Descripteurs généralisés de Fourier ; Evaluation de robustesse.

Abstract

In the context of the acquisition of depth information for textured scenes, a depth estimation process based on a 3D reconstruction method called "shape from focus" is proposed in this thesis. The two crucial steps of this approach are the image sequence acquisition of the scene by optical sectioning and the local sharpness evaluation for each pixel of the acquired images. Two acquisition systems have been developed and are presented as well as different image processing techniques that enable the image exploitation for the depth estimation process. The pixel sharpness evaluation requires comparison of different focus measure operators in order to determine the most appropriate ones. In addition to the usual focus measure operators, two news operators based on generalized Fourier descriptors are presented. A new and original comparison method is developed and provides a further analysis of the robustness to various parameters of the focus measure operators. In order to provide an automatic version of the reconstruction process, two automatic sharpness evaluation methods are detailed. Finally, the whole reconstruction process is applied to agronomic scenes, but also to a problematic in failure analysis domain aiming to expand to other applications.

Keywords : 3D reconstruction ; Shape from Focus ; Focus measure ; Generalized Fourier descriptors ; Robustness evaluation.

The logo for SPIM (École doctorale SPIM) features the letters 'S', 'P', 'I', and 'M' in a large, white, sans-serif font. The 'S' is stylized with a thick orange horizontal bar to its left. The letters are set against a white background.