



Reconstruction de surface 3D d'objets vivants

Rémi Slys

► To cite this version:

Rémi Slys. Reconstruction de surface 3D d'objets vivants. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université de Technologie de Troyes, 2014. Français. NNT : 2014TROY0022 . tel-03357128

HAL Id: tel-03357128

<https://theses.hal.science/tel-03357128>

Submitted on 28 Sep 2021

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

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

Thèse
de doctorat
de l'UTT

Rémi SLYSZ

Reconstruction de surface 3D d'objets vivants

Spécialité :
Systèmes Mécaniques et Matériaux

2014TROY0022

Année 2014

THESE

pour l'obtention du grade de

DOCTEUR de l'UNIVERSITE DE TECHNOLOGIE DE TROYES Spécialité : SYSTEMES MECANQUES ET MATERIAUX

présentée et soutenue par

Rémi SLYSZ

le 10 juin 2014

Reconstruction de surface 3D d'objets vivants

JURY

M. J. PONCE	PROFESSEUR DES UNIVERSITES	Président
Mme I. BLOCH	PROFESSEUR TELECOM PARISTECH	Rapporteur
M. H. BOROUCAKI	PROFESSEUR DES UNIVERSITES	Directeur de thèse
M. R. CHAINE	PROFESSEUR DES UNIVERSITES	Rapporteur
Mme L. GIRAUD-MOREAU	MAITRE DES CONFERENCES	Directeur de thèse
M. M. ROUX	DOCTEUR	Examineur

Remerciements

Je tiens tout d'abord à exprimer ma gratitude aux personnes qui m'ont fait l'honneur de participer au jury de cette thèse :

Monsieur Jean Pons, Professeur des Universités (ENS) qui m'a fait l'honneur de présider ce jury,

Madame Isabelle Bloch, Professeur à Télécom ParisTech (ENST), Rapporteur de cette thèse

Madame Raphaëlle Chainé (prof), Professeur des Universités (LIRIS), Rapporteur de cette thèse,

Monsieur Michel Roux, Docteur Ingénieur (ENST), Examineur de cette thèse

Madame Laurence Giraud-Moreau et Monsieur Houman Borouchaki, qui m'ont accueilli au sein de l'équipe de recherche GAMMA3 de l'institut Charles Delauney de l'Université de Technologie de Troyes, et m'ont permis de travailler dans les meilleures conditions en assumant tous deux le rôle de Directeur de Thèse et en me soutenant dans les moments difficiles.

Je suis tout particulièrement reconnaissant à Véronique Banse pour son formidable travail au secrétariat du pôle, ses aides précieuses et toutes nos discussions pleines de convivialité.

Je remercie Rémi Cogranne pour son aide et le temps qu'il a accordé à nos co-rédactions.

Toute ma sympathie va aussi aux membres de l'équipe de recherche GAMMA3, ainsi qu'à tout le personnel de l'UTT et de l'INRIA que j'ai été amené à côtoyer.

Je tiens également à remercier les amis que j'ai rencontrés au cours de mon passage à Troyes, tout particulièrement Cathel, Jawad, Léa, Madhi, Rony et Sabrina qui ont largement contribué à cette thèse par des moments de détente nécessaire.

Enfin, je ne remercierai jamais assez mes parents et mes proches pour leur soutien inconditionnel, leurs encouragements répétés, et leur disponibilité. Et parmi eux, je remercie tout spécialement ma compagne Sandra pour son indéfectible soutien moral et grammatical.

Finalement, j'aimerais remercier Gipsy, mon chat, qui m'a prouvé et me prouve encore que la *ronron thérapie* est le meilleur des antidépresseurs.

Table des matières

Table des figures	ix
Liste des tableaux	xiii
Notations	xv
Introduction générale	1
I État de l'art sur les méthodes de reconstruction	5
1 Méthodes par contact	6
2 Méthodes basées sur des propriétés physiques	6
2.1 Le temps de vol	7
2.2 Le décalage de phase	8
2.3 L'holographie conoscopique	9
2.4 L'imagerie radar	9
3 Méthodes utilisant des capteurs CCD	10
3.1 La lumière modulée	10
3.2 Les silhouettes	11
3.3 La triangulation	11
3.4 La lumière structurée	13
3.5 La stéréoscopie	14
4 Conclusion	15

Table des matières

II	Mise en place d'un dispositif d'acquisition expérimental	17
1	Choix et disposition des appareils de captures	17
1.1	Appareils de captures d'images	18
1.2	Appareils de projection	21
1.3	Assemblage	25
2	Interaction cabine - ordinateur	27
2.1	Interface	28
2.2	Système de déclenchement	29
2.3	Flux de récupération des données	31
3	<i>Calibration</i> du système	33
4	Conclusion	37
III	Corrélation point à point	39
1	Modèle géométrique de formation des images	40
2	Géométrie épipolaire	45
2.1	Relations géométriques	45
2.2	Relations analytiques	46
3	Méthodes courantes d'appariement de point d'intérêt	47
3.1	Notions générales	48
3.2	Disparité et fonction de disparité	52
3.3	Contraintes	52
3.3.1	Contrainte épipolaire et rectification épipolaire	53
3.3.2	Contrainte d'unicité	55
3.3.3	Contrainte d'ordre	55
3.3.4	Contrainte de symétrie et contrainte de consistance forte	56
3.3.5	Contrainte de consistance faible	56
3.3.6	Contrainte sur le gradient de disparité	57
3.3.7	Contrainte de rang	58
3.3.8	Contrainte de continuité figurale	58
3.3.9	Conclusion sur les contraintes	58
3.4	Méthodes accélératrices	59
3.4.1	Réduction de la zone de recherche	59
3.4.2	Caractérisation locale des points remarquables	60
3.4.3	Conclusion sur les méthodes accélératrices	62
3.5	Mesures de corrélation	63

	3.5.1	Corrélation croisée	64
	3.5.2	Statistique	65
	3.5.3	Dérivée	66
	3.5.4	Autres mesures	66
4		Évaluation	67
	4.1	Protocole expérimental	67
	4.2	Résultats	68
5		Conclusion	68
IV Corrélation appuyée sur un maillage			71
1		Un motif utilisé comme contrainte	72
	1.1	Proposition d'exploitation et implémentation	73
	1.2	Résultats obtenus	73
2		Intérêt d'un motif de type maillage	75
	2.1	Structure topologique d'une triangulation	76
	2.1.1	Liste des éléments	76
	2.1.2	Robustesse des éléments face à la transformation	76
	2.2	Construction d'une clef	77
	2.2.1	Rang d'une clef	78
	2.2.2	Utilisation des clefs	78
	2.3	Résultats	80
	2.3.1	Résultats numériques	80
	2.3.2	Preuve de la redondance	81
	2.4	Conclusion	83
3		Détecteur de contour élaboré pour l'extraction d'un motif de type maillage . . .	85
	3.1	Image 2D vue comme un assemblage de surface 3D	85
	3.2	Construction d'un détecteur de contour associé	88
	3.2.1	Implémentation	88
	3.2.2	Détection de la présence d'un contour	91
	3.2.3	Descripteur de contours	93
	3.3	Évaluation du détecteur	97
	3.3.1	Protocole expérimental	97
	3.3.2	Résultats	100
4		Extraction des maillages projetés	104
	4.1	Extraction du motif	104

Table des matières

4.1.1	Analyse du motif : recherche des coins	105
4.1.2	Analyse du motif : recherche des nœuds	109
4.2	Extraction du motif par propagation	111
5	Conclusion	114
V	Reconstruction tridimensionnel et raffinement	117
1	Reconstruction tridimensionnelle	118
1.1	Stratification des reconstructions tridimensionnelles	118
1.1.1	Reconstruction projective	119
1.1.2	Reconstruction affine	119
1.1.3	Reconstruction métrique	120
1.1.4	Reconstruction euclidienne	122
1.2	Rappel sur la calibration	122
1.3	Triangulation à partir de deux vues	123
1.4	Reconstruction par bloc de capteur	124
1.5	Assemblage des maillages surfaciques de bloc	125
2	Raffinement	126
2.1	Détection des zones nécessitant raffinement	126
3	Conclusion	129
	Conclusion générale et perspectives	131
	Annexe A : Détecteur de contours	137
	Annexe C : Documentation de la cabine	139
B.1	L'agencement de la cabine	139
B.1.1	Appareils de captures d'images	139
B.1.2	Appareils de projection	140
B.1.3	Système de déclenchement	141
B.1.4	Disposition	143
B.2	L'interface	144
B.2.1	IHM	144
B.2.2	<i>Calibration</i>	146
B.2.3	Identification automatique des appareils photographiques et arborescence	146
B.2.4	Flux de récupération des données	147

Bibliographie	149
---------------	-----

Table des matières

Table des figures

I.1	Système MicroScrib à gauche et système Patriot à droite.	6
I.2	Schéma du principe par temps de vol.	7
I.3	Schéma du principe par décalage de phase.	8
I.4	Acquisition en imagerie radar	10
I.5	Scanner tridimensionnel par triangulation.	12
I.6	Le projet Digital Michelangelo.	13
II.1	Disposition des appareils photographiques numériques	20
II.2	Schéma de calcul du cercle optique.	20
II.3	Technologie LCD	21
II.4	Effet <i>Screendor</i>	22
II.5	Technologie DLP	22
II.6	Effet “Arc-en-ciel” dû à la roue de couleurs.	23
II.7	Technologie LCoS.	23
II.8	Disposition des vidéoprojecteurs.	25
II.9	Structure de la cabine.	26
II.10	Cablage de la cabine.	28
II.11	Photographies de la cabine d’acquisition	29
II.12	Interface	30
II.13	Circuit de déclenchement pour un appareil photographique de type 600D.	31
II.14	Circuit de déclenchement simultané utilisant une interface Arduino	32
II.15	Mire 3D	34
II.16	Mire 2D	34

Table des figures

II.17	Mire 1D	35
II.18	Mire plane aléatoire pour résoudre les problèmes d'occultation. [OFN12] . . .	35
III.1	Stéréovision binoculaire	39
III.2	Formation d'une image au sein d'un appareil photographique	41
III.3	Formation d'image : lentille mince	41
III.4	Sténopé	42
III.5	Repère sténopé	42
III.6	Géométrie épipolaire.	45
III.7	Zones de recherche, d'agrégation et support	49
III.8	Rectification épipolaire.	54
III.9	Contrainte d'unicité non respectée.	55
III.10	Contrainte d'ordre non respectée.	56
III.11	Synthèse de toutes les contraintes géométrique.	57
III.12	Réduction de la zone de recherche en connaissant l'enveloppe.	60
III.13	Taches.	68
III.14	Nuage de points acquis	69
IV.1	Zones de test pour l'évaluation des mesures de corrélation à 60°	73
IV.2	Exemple de construction de clefs	79
IV.3	Exemple d'utilisation du motif proposé	80
IV.4	Distribution du degré des nœuds dans une triangulation aléatoire de Delaunay pour 100000 nœuds.	81
IV.5	Critère de Delaunay et Cocircularité	84
IV.6	Illustration du modèle de représentation proposé	85
IV.7	Cartes de contour	90
IV.8	Exemple trivial de contour	92
IV.9	Résultat de la méthode de reconnaissance de forme proposée élargi par une dilatation morphologique pour faciliter la visualisation.	94
IV.10	Résultat de la méthode de reconnaissance de forme proposée élargi par une dilatation morphologique pour faciliter la visualisation.	95
IV.11	Analyse de la variation des coefficients sur des exemples types.	96
IV.12	Évaluation visuelle du détecteur de contours et du détecteur de coins induit .	96
IV.13	Image de test utilisée pour l'évaluation du détecteur de contours et la carte de ses contours.	97
IV.14	Images de test	99

IV.15	Comparaison de la robustesse face au flou gaussien (<i>FoM</i>)	101
IV.16	Comparaison de la robustesse face au bruit gaussien (<i>FoM</i>)	101
IV.17	Illustration de l'influence des paramètres face au flou	102
IV.18	Illustration de l'influence des paramètres face au bruit	102
IV.19	Comparaison visuelle sur les images <i>Lena</i> et <i>boats</i>	103
IV.20	Ligne de partage des eaux : initialisation	107
IV.21	Ligne de partage des eaux : recherche dichotomique	109
IV.22	Positionnement des coins par rapport au nœud	110
IV.23	Extraction d'un triangle	113
V.1	Hierarchisation des reconstructions.	118
V.2	Triangulation spatiale	123
V.3	Triangulation spatiale au sein d'un bloc	125
V.4	Correction des fonctions de forme	127
A.1	<i>Brueghel : La tour de Babel, 1563</i> . Image utilisée pour la détection suivante, voire figure A.2	137
A.2	Les contours sont extraits à l'aide d'une quadrique dans un voisinage 3×3 depuis A.1	138
B.1	Disposition des appareils photographiques numériques	140
B.2	Disposition des vidéo-projecteurs.	141
B.3	Circuit de déclenchement simultané utilisant une interface Arduino et brochage de ses broches de type digitale (<i>digital pin</i> sur l'Arduino). "dec." correspond à la commande de déclenchement et "focus" à la commande de mise au point. Les appareils sont notés en fonction du numéro de leur colonne et de la hauteur sur celle-ci ("H" pour haut et "B" pour bas).	142
B.4	Structure de la cabine.	143
B.5	Cablage de la cabine.	144
B.6	Interface	145

Table des figures

Liste des tableaux

II.1	Étude comparative des différents appareils de captures envisagés	19
II.2	Étude comparative des différents objectifs.	21
IV.1	Mesures comparatives à 60 ° et distance à l'erreur	74
IV.2	Pourcentage de clefs différentes	81
IV.3	Table des majorants et condition d'unicité.	83
IV.4	Comparaison de la robustesse face au flou gaussien	99
IV.5	Comparaison de la robustesse face au bruit gaussien	100

Liste des tableaux

Notations

Élément géométrique de la scène et des capteurs :

$(O, \vec{x}, \vec{y}, \vec{z})$: Repère tridimensionnel (repère monde).

$(O_c, \vec{x}, \vec{y}, \vec{z})$: Repère tridimensionnel lié au capteur d'image c .

(I_c, \vec{u}, \vec{v}) : Repère image centré sur le pixel en haut à gauche, \vec{u} allant vers la droite et \vec{v} allant vers le bas.

I_i : Image i .

C_i : Position du centre d'un capteur d'image.

S : Scène à acquérir.

M : Point de surface tridimensionnel (sur S).

$M_{\text{Repère}}$: M exprimé dans un repère “Repère” spécifique (différent du repère monde).

m_i : Image du point M dans l'espace image i .

l_i : i -ème droite épipolaire.

e_i : i -ème épipôle.

Π : Plan épipolaire.

f : Distance focale.

p : Point principal.

k_u et k_v : Facteurs de conversion horizontal et vertical.

ϕ : Angle entre les vecteurs \vec{u} et \vec{v} .

K_{int} : Matrice encodant les paramètres intrinsèques.

F : Matrice fondamentale.

Relations :

\hat{v} : Le vecteur v exprimé en coordonnées homogènes.

\propto : “égal a un facteur près non-nul”. Utilisé conjointement aux coordonnées homogènes :
 $v = (x, y, z) \iff (\hat{v} = (ax, ay, az, a), \forall a \neq 0)$

$A \doteq B$: Relation symbolisant que A et B sont deux points homologues.

$A \not\doteq B$: Relation symbolisant que A et B sont deux points qui ne sont pas homologues.

$+$: Opérateur d'addition modulo le degré de η .
 $[\deg(\eta)]$

∇ : Opérateur Nabla. ∇I : Gradient de l'image I .

$\#A$: Cardinal de l'ensemble A .

Notations matricielles :

$Id_{d \times d}$: Matrice identité carrée de taille d .

R : Matrice de rotation.

$(R|t)$: Matrice 4×3 comprenant une matrice de rotation pour les trois premières colonnes et une matrice de translation dans la quatrième.

P : Matrice de projection (d'une caméra).

A^T : Matrice transposée de A .

A^+ : Matrice pseudo inverse de A . $A^+ = A^T(AA^T)^{-1}$

$[x]_{\times}$: Matrice carrée encodant le produit scalaire. $[v_1]_{\times} v_2 = v_1 \times v_2$. Elle se définit en dimension 3 par $[x]_{\times} = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}$

Notation des énergies :

E_{type} : Énergie de type “type”.

E_{type}^c : Énergie de type “type” pour un couple de primitives données c .

Notations des zones pour l'algorithme d'appariement :

$\mathbb{A}\mathbb{C}^c(P)$: Zone d'agrégation associée à la contrainte \mathbb{C} pour un couple de primitives données c dont l'un des membres est la primitive P (généralement, c'est un voisinage).

$\mathbb{A}(P)$: Zone d'agrégation locale à la primitive P .

\mathbb{S} : Support de l'appariement (ensemble des primitive requises pour le calcul).
 $\mathbb{Z}_{\text{considérée}}$: Zone dont on veut apparier les points avec ceux d'une autre image.
 $\mathbb{Z}_{\text{recherche}}$: Zone dans une autre image contenant les candidats à l'appariement.

Notations des différentes mesures de corrélations :

CC : Corrélation croisée.
NCC : Corrélation croisée normalisée.
ZNCC : Corrélation centrée croisée normalisée.
SAD : Somme des différences absolues.
SSD : Somme du carré des différences.
LSSD ou ZSSD : Somme du carré des différences localement centrées.
ZNSSD : Somme du carré des différences normalisées localement centrées.
VD : Variance des différences.
VAD : Variance des différences absolues
 K_4 : Kurtosis.
 SES_p : Norme L_p des différences des dérivées fournies par l'opérateur de Sobel.
 SEK_p : Norme L_p des différences des dérivées fournies par l'opérateur de Kirsh.
NIS : Mesure de Nishimura.
PRATT : Mesure de Pratt.
NACK : Mesure de Nack.
MAD : Déviation absolue des médianes.
 MR_6 : Estimateur sur les résidus des moindres carrés.

Notations au sein d'une triangulation :

\mathcal{N} : Ensemble des nœuds d'une triangulation. Ses éléments sont désignés par η ou n .
 $\mathcal{V}(\eta)$: Ensemble des nœuds d'une triangulation voisins de η .
 K^i : Clef de rang i .
 \mathbb{D} : Sous-intervalle discret de \mathbb{N} . $\mathbb{D} = \llbracket 0, \deg(\eta) \rrbracket$.
 $\deg(\eta)$: Nombre d'adjacence du nœud η .

Autres notations :

$d(.)$: Fonction de disparité (renvoie un vecteur le long de la ligne épipolaire correspondant au décalage).

$\text{PSF}(C)$: Fonction représentant le bruit d'acquisition du capteur d'image C .

$\text{IMP}(C)$: Fonction représentant les défauts du capteur C (pixels morts, pixels chauds).

Ω : Espace vectoriel. $\Omega = \mathbb{N}^t$ où t est la taille de la zone d'agrégation locale.

ω : Ensemble des primitives à utiliser dans un critère de similarité mis sous forme vectorielle (généralement un voisinage mis sous forme de vecteur).

$\bar{\omega}$: Moyenne des primitives contenues dans ω . Dans notre cas, nous l'appliquerons exclusivement à des primitives de type intensité lumineuse des pixels.

N_c : Nombre de contrainte.

$P_g^{i,j}$: Primitive en (i, j) dans I_g .

Introduction générale

La présente thèse se situe dans le cadre du projet BRAMMS soutenu par la région Champagne-Ardenne et le Fonds Européen de Développement Économique et Régional et a commencé le 1er octobre 2010.

La problématique initiale de BRAMMS est de mettre au point des méthodes et des outils pour fournir aux professionnels en relation avec le buste féminin les moyens de satisfaire leur clientèle par la performance de leurs articles. Les initiateur du projet sont des professionnels de la lingerie-corseterie habillant toutes les femmes quelque soit leur morphologie, ou encore celles ayant subies une chirurgie réparatrice et/ou de reconstruction (chirurgie plastique incluse). Les professionnels du secteur du balnéaire sont concernés à moindre titre, bien que ceux du domaine sportif soient aussi fortement demandeurs de performances puisque l'efficacité du maintien est le garant d'une protection des tissus impliqués.

Les attentes des utilisateurs sont multiples : nouveaux moyens d'acquisition permettant d'effectuer des campagnes de mensurations plus rapidement, nouveaux modèles de représentation de la structure interne du seins (actuellement très imprécis), nouveaux modèles permettant d'évaluer l'efficacité du maintien et donc de la performance des produits (alliant confort et protection des tissus).

Trois nouvelles démarches sont donc à mener.

Tout d'abord, de nouveaux moyens d'acquisition et de caractérisation de la morphologie du buste nu sont nécessaires. En effet, les informations résultantes (mesures et surface 3D) pourront alimenter les outils CAO¹ des produits. Ces moyens seront particulièrement utiles par exemple à la conception des articles de lingerie et à la réalisation de prothèse pour le domaine chirurgical.

1. Conception Assistée par Ordinateur

Au delà de ce projet devront aussi naître de nouveaux modèles numériques, permettant d'évaluer l'efficacité du maintien du buste lors de l'activité quotidienne ou sportive, et cela dès le cycle de conception via des outils CAO et de calcul éléments finis.

Enfin, il sera nécessaire de créer un simulateur afin de tester et évaluer les performances des produits développés.

Afin de réaliser toutes ses attentes, le projet BRAMMS réunit plusieurs acteurs dont le principal est l'Institut Français du Textile et de l'Habillement (IFTH), initiateur du projet.

Dans le cadre de cette thèse, deux tâches dénommées “Tache 1 : Acquisition et mesure de la morphologie du buste et du sein” et “Tache 2 : Reconstruction de la surface du buste, extraction de mesures, modélisation CAO paramétrée du buste” ont été réalisées. La première tâche consistait principalement à élaborer un système d'acquisition adapté à la problématique. L'IFTH ayant par le passé testé un système d'acquisition ne permettant pas d'obtenir de mesures précises, il était nécessaire d'étudier les dispositifs actuellement disponibles ou d'en concevoir un si rien n'était adapté. La seconde tâche consistait elle à concevoir et implémenter l'algorithme permettant d'effectuer les mesures à l'aide du dispositif choisi.

Une troisième tâche dénommée “Tache 3 : Réalisation d'une mini-campagne de mesure” a été effectuée en mai 2012 à l'IFTH et à l'UTT. Nous avons profité de cet événement pour valider la conception de la cabine d'acquisition conçue dans le cadre de mes travaux à l'UTT.

La présente thèse s'évertue donc à répondre aux problématiques soulevées par les deux premières tâches. L'ensemble des travaux ont été effectués sous la supervision de Laurence Moreau et d'Houman Borouchaki, tous deux faisant partie du groupe de recherche Génération Automatique de Maillages et Méthodes Avancées (GAMMA3) qui m'a accueilli au sein de l'UTT, ainsi qu'à l'INRIA² lors de ma dernière année. Ce groupe de recherche est spécialisé dans l'utilisation de maillages en tant que support de calcul et dans l'adaptation de ces maillages. Cela a fortement orienté mes recherches à plusieurs niveaux et nous a permis de penser à la meilleure façon d'obtenir un maillage déjà adapté à la surface qu'il représente.

Les principaux points soulevés lors de cette thèse portent donc sur l'élaboration d'un système d'acquisition de mesure tridimensionnelle adapté à des êtres humains, ou à des objets mobiles, qui peuvent être — et c'est le cas dans la grande majorité — de couleur et de texture uniforme. L'uniformité est très contraignante dans notre application et est abordée à plusieurs reprises tout au long de la thèse. La quasi-totalité des systèmes d'acquisition de mesure tridimensionnelle actuellement sur le marché sont difficilement applicables aux objets mobiles. C'est ce qui motivera en grande partie l'élaboration du dispositif d'acquisition. Et de par la concep-

2. Institut National de Recherche en Informatique et en Automatique

tion contrainte de la cabine (contraintes mécaniques et matérielles), l'algorithme classique de reconstruction devra être repensé pour ce dispositif.

Dans un premier temps, nous consacrerons donc les deux premiers chapitres à la Tache 1. Le premier chapitre sera entièrement consacré aux très nombreuses méthodes de mesures tridimensionnelles existantes, en gardant à l'esprit l'objet d'étude qui est ici le buste féminin. La construction de la cabine n'étant toutefois pas anodine, il était nécessaire de pouvoir aussi l'utiliser pour d'autres applications. Un état de l'art des techniques de métrologie y est donc ainsi présenté.

Le second chapitre s'attèlera à montrer le raisonnement impliqué dans la conception de la cabine et ainsi à expliquer nos choix technologiques. Une large partie de ce chapitre est donc consacré à une étude technologique poussée des appareils s'inscrivant dans la technique de métrologie choisie à l'issue du premier chapitre.

Dans un second temps, nous consacrerons les trois derniers chapitres à la Tache 2. Le troisième chapitre débutera par une introduction complète à la problématique de la stéréovision ; c'est à dire, à la problématique de restitution de l'espace tridimensionnel à partir de projections. Il continuera ensuite par une approche détaillée de la base sous-jacente de toute reconstruction par stéréovision, à savoir l'appariement de points d'intérêts.

Les résultats présentés dans le chapitre trois nous conduiront à explorer de façon plus approfondie les techniques d'appariement se basant sur un motif projeté, ce qui occupera l'entièreté du quatrième chapitre. De par les connaissances de l'équipe de recherche GAMMA3 et de par les tests effectués, nous avons développé un algorithme novateur d'appariement basé sur des motifs projetés de type maillage. Le cas de la triangulation y a particulièrement été étudié. C'est dans ce chapitre qu'auront lieu la majeure partie des avancées proposées lors de cette thèse.

Enfin, le dernier chapitre introduira la reconstruction tridimensionnelle à partir des appariements. Ce chapitre expose en grande partie les idées n'ayant pas pu être traitées dans cette thèse par manque de temps. Il soulignera aussi de nouveaux intérêts majeurs au choix effectués lors du chapitre précédent.

Chapitre I

État de l'art sur les méthodes de reconstruction

Afin de pouvoir au mieux concevoir la cabine d'acquisition surfacique nécessaire au projet Bramms, une étude des méthodes d'acquisition actuelles a été réalisée.

Dans ce chapitre, différentes méthodes de métrologie tridimensionnelle sont donc présentées et brièvement expliquées. Ces méthodes seront ensuite comparées entre-elles en termes d'utilisation et d'efficacité. Pour chaque cas, les limites de leur emploi dans le cadre de la cabine d'acquisition seront exposées.

Par soucis de simplification, ces méthodes ont été ici regroupées en 3 catégories :

- les **Méthodes par contact**, qui utilisent un système mécanique afin d'acquérir la surface.
- les **Méthodes basées sur des propriétés physiques**, qui utilisent les propriétés physiques d'une onde ou d'un laser afin d'acquérir une surface.
- les **Méthodes utilisant des capteurs CCD**, qui utilisent les propriétés optiques et géométriques des caméras afin d'acquérir une surface.

Nous ne présenterons pas ici de méthodes venant du domaine médical. La plupart de ces méthodes fonctionnent par tomographie. Elles sont très utiles pour déterminer l'intérieur d'un objet. Ce n'est toutefois pas ce que l'on recherche ici. De plus, il n'est pas question de mettre en danger la santé des personnes entrant dans la cabine d'acquisition. Ainsi la technique du rayon X se révèle trop dangereuse car il n'y a aucun gain en termes de santé. L'imagerie par résonance magnétique est extrêmement longue en temps d'acquisition et il est donc impossible de l'employer aussi pour cette raison. Enfin, l'échographie est similaire aux mesures basées sur des propriétés physiques. Il ne nous est donc pas paru pertinent de développer ces méthodes dans le présent manuscrit.

1 Méthodes par contact

Les méthodes de cette famille sondent le sujet par contact physique. Ce type de scanner est utilisé pour sa précision. Cependant, son usage ne peut s'appliquer à tous les cas à cause de son principe même. En effet, à cause du contact physique nécessaire, cette famille de méthodes a tendance à détruire ou altérer les objets fragiles. Son utilisation sur des objets uniques ou historiques est donc à proscrire. De plus, cette méthode a le défaut d'être relativement lente (comparativement aux autres). En effet, le palpeur ne peut faire des relevés qu'à faible fréquence de l'ordre de 0,1kHz.

Un premier exemple d'utilisation de ce type de scanner se trouve dans l'industrie mécanique, par exemple, pour des machines à mesurer tridimensionnelles. Dans certains cas, le palpeur est placé sur un bras motorisé. Dans d'autres, c'est l'objet lui-même qui est déplacé. Le procédé reste très long bien que l'automate remplace le travail humain.

Autre exemple, dans l'animation cinématographique, des modèles sculptés dans la terre glaise sont numérisés en trois dimensions à l'aide de ce type de méthode. Le palpeur peut dans ce cas être directement déplacé par l'utilisateur.



Figure I.1 — Système MicroScrib à gauche et système Patriot à droite.

Bien que précise, cette méthode a le défaut d'être trop lente. De plus, le contact physique entre l'appareil d'acquisition et le sujet n'était ici pas envisageable pour le type de sujet visé.

2 Méthodes basées sur des propriétés physiques

Les différentes utilisations du laser à des fins de métrologie vont être présentées dans cette section. L'imagerie radar sera aussi abordée, car reposant sur une approche similaire reposant sur des ondes radio.

2.1 Le temps de vol

Le scanner tridimensionnel LIDAR¹ est un appareil qui utilise un faisceau laser pour sonder le sujet. Au cœur de ce type de scanner se trouve un télémètre laser permettant de calculer la distance avec la surface de l'objet étudié, en comptant le temps nécessaire au trajet aller-retour de l'impulsion du faisceau laser réfléchi. Le temps de retour permet de déterminer la distance parcourue par la lumière et on a donc, si t est le temps de retour et c la vitesse de la lumière : $d(\text{Objet}, \text{scanner}) = (c \cdot t)/2$. L'exactitude des mesures dépendent donc directement de la précision de la mesure de t .

Cependant cette méthode ne permet que de détecter un point à la fois. Pour reconstruire la totalité d'un objet il faut alors changer l'orientation du télémètre laser afin d'en balayer la surface point par point. L'orientation est changée soit par rotation du télémètre lui-même, soit plus couramment par un système de miroirs rotatifs qui peuvent changer plus rapidement et plus précisément de direction grâce à leur légèreté.

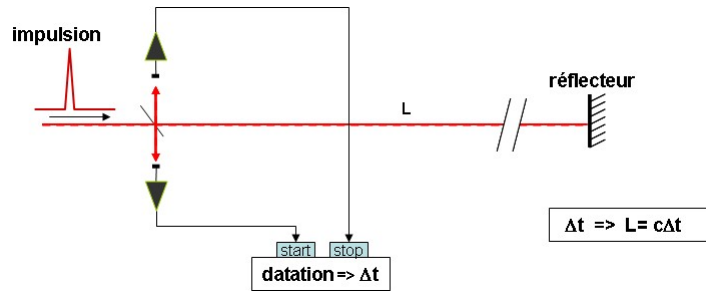


Figure I.2 – Schéma du principe par temps de vol.

Les capteurs par temps de vol sont généralement destinés à évaluer de grandes distances (généralement supérieur à 10m) et ont été développés initialement pour la détection automatisée d'obstacles par des véhicules [LJ77, Bes88].

Les scanners tridimensionnels par temps de vol peuvent mesurer la distance de 10000 à 100000 points par seconde, avec cependant une imprécision supérieure au centimètre. Il est possible aujourd'hui de trouver des systèmes semblables bien plus précis puisqu'allant jusqu'au millimètre mais leur rapidité en est directement impactée (acquisition de l'ordre de 800 points par seconde).

De plus, la précision des scanners par temps de vol peut-être perdue quand une impulsion frappe le bord d'un objet, l'information renvoyée provenant alors de deux endroits différents pour une seule émission, sa localisation sera erronée. Lors de l'utilisation d'un scanner à haute

1. *LIght Detection And Ranging*

résolution sur un objet, la probabilité pour que le faisceau frappe un bord est de fait plus élevée et un bruit sera visible sur les bords acquis.

Cette méthode relativement robuste propose une acquisition rapide mais peu précise, ou précise mais peu rapide. Lors de l'acquisition de la surface d'un objet vivant, la rapidité est primordiale pour pallier le flou de bougé. Une précision assez élevée est néanmoins requise afin d'obtenir le niveau de détail voulu pour la reconstruction. Bien que cette méthode soit très utile pour le calcul de distance entre objets éloignés, ce n'est pas notre cas. Considérant ce fait, nous pouvons nous focaliser sur des méthodes plus adaptés à la courte portée.

2.2 Le décalage de phase

Plus rapide et plus précise que la méthode du temps de vol, la mesure de décalage de phase permet elle aussi de mesurer des distances. Le laser émis par le scanner est ici encore réfléchi par l'objet. En comparant la phase de l'onde sinusoïdale connue du rayon émis, et la phase du rayon retour, il est possible de déterminer le décalage de phase et donc la distance de la cible au scanner.

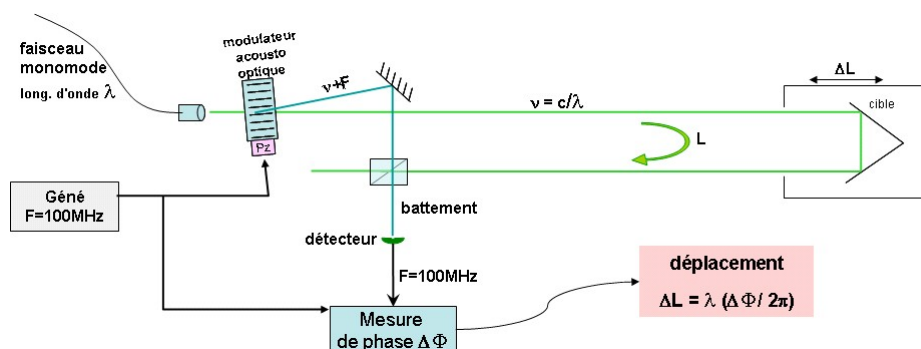


Figure I.3 — Schéma du principe par décalage de phase.

Les scanners à décalage de phase sont généralement plus rapides et plus précis - il est possible de mesurer la distance au millimètre près - que les scanners laser tridimensionnels à temps de vol, mais ont une portée plus réduite.

Plus précise et rapide, cette méthode semble bien plus prometteuse. Elle possède cependant l'inconvénient majeur de ne s'appliquer qu'à des objets réfléchissants orientés dans la direction du scanner. Dans notre cas, son application en est donc fortement compromise. De plus, il n'est toujours pas assuré que le flou de bougé ne sera pas présent malgré sa vitesse puisque l'acquisition se fait toujours point par point.

2.3 L'holographie conoscopique

Les recherches de Dennis Gabor sur l'amélioration de la résolution des microscopes électroniques en 1948 [Gab48] ont débouchés sur le début de l'holographie. Emmitt Leith et Youris Upatnieks ont par la suite développés l'holographie "hors-axe" grâce à l'apparition du laser dans les années 60 [LU62]. L'holographie interférométrique a été appliquée au contrôle non-destructif de matériaux [Col71, Ves79]. Cette méthode permet d'acquérir des déplacements de surfaces, des changements d'épaisseurs ou d'indices de réfraction d'objets transparents, des changements d'amplitudes locales et de phases relatives, ou des mouvements vibratoires.

Dans un système conoscopique, un rayon laser est projeté sur une surface, puis la réflexion au travers du même faisceau passe à travers un cristal biréfringent et est envoyée sur un capteur CDD (ou CMOS). La fréquence des motifs de diffraction peut être analysée, et permet de déterminer la distance entre le scanner et la surface. L'atout principal de l'holographie conoscopique est qu'un unique faisceau (aller-retour) est nécessaire pour effectuer la mesure, permettant de mesurer par exemple la profondeur d'un trou finement percé.

L'holographie conoscopique permet d'acquérir plus d'information que la simple position spatiale de chaque point. L'information de mouvement pourrait notamment être utilisée pour pallier les mouvements du sujet. Elle conserve cependant les mêmes inconvénients que ses pairs utilisant le laser. C'est à dire qu'elle ne permet l'acquisition que d'un point à la fois.

2.4 L'imagerie radar

L'imagerie radar est surtout utilisée pour obtenir des informations de position et de déplacement d'objets situés à une grande distance. Essentiellement utilisée dans le domaine militaire, l'étude du sol (Fig. I.4) ou la météorologie, on notera tout de même son emploi dans le secteur automobile pour les radars de recul et la sécurité routière.

Le système radar envoie une onde radio (ou sonore dans le cas d'un sonar en milieu aquatique) et récupère sa réverbération sur un objet. En analysant le signal réfléchi, il est possible de localiser et d'identifier l'objet. Il est aussi possible de calculer sa vitesse grâce à l'effet Doppler. En balayant une zone, on peut ainsi obtenir la surface de cette zone.

Parmi les systèmes de radar imageurs, on notera le radar à synthèse d'ouverture [Cut90] (RSO, ou SAR^2 en anglais) qui effectue un traitement des données acquises afin d'améliorer la résolution en azimuth qui est le principal défaut de la technologie radar utilisée à des fins météorologiques. On notera également de récents travaux visant à diminuer le temps d'acquisition

2. *Synthetic Aperture Radar*

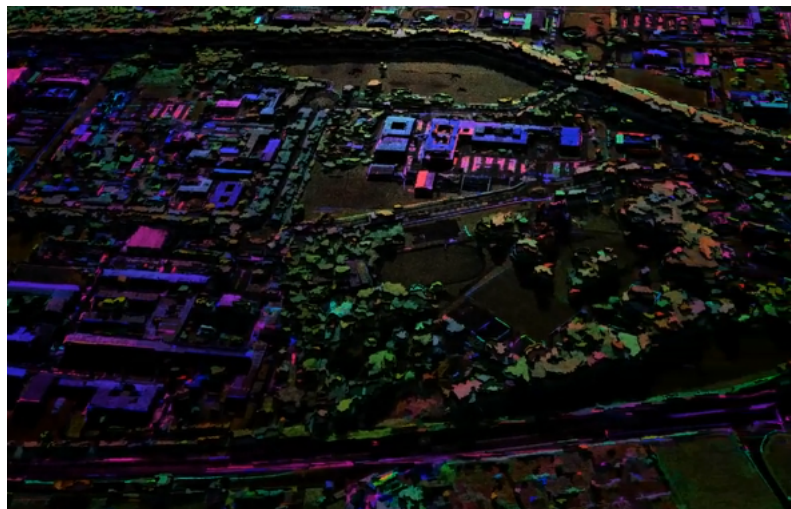


Figure I.4 – Acquisition tridimensionnelle d’une partie de la ville de Toulouse par l’équipement aéroporté de l’Onera (acquisition par une méthode hybride RSO et polarimétrie), 2011.

en utilisant des données éparées [AEM11].

Ce type d’imagerie est plutôt adapté à la prise de mesure sur longue distance. Des versions existent pour des distances plus courtes. On augmente alors la fréquence du signal. Encore une fois, cette technologie acquière les points un à un, ce qui est problématique en termes de précision dans le cadre d’une acquisition surfacique sur sujet mobile.

3 Méthodes utilisant des capteurs CCD

Dans cette section sont présentées diverses approches utilisant les capteurs CCD. Les capteurs CCD sont utilisés afin d’acquérir des informations optiques. L’utilisation de capteurs CCD se fait généralement conjointement à l’étude géométrique de la scène afin d’obtenir une mesure. Contrairement à la famille précédente, il est ici possible dans certains cas d’acquérir toute la scène d’une seule acquisition.

3.1 La lumière modulée

Les scanners tridimensionnels à lumière modulée éclairent le sujet à l’aide d’une lumière changeante. Habituellement, la source de lumière a un cycle dont l’amplitude décrit un motif sinusoïdal. Une caméra détecte la lumière réfléchie, mesure l’importance de sa variation et détermine la distance que la lumière a parcourue. La lumière modulée permet également au

scanner d'ignorer toute source de lumière autre que celle de son laser, de telle sorte qu'il n'y a pas d'interférence. Pour de plus amples détails, le lecteur est renvoyé à [OLK⁺04].

Cette méthode s'apparente par sa conception aux scanners par décalage de phases. Contrairement à ceux-ci cependant, l'utilisation du laser n'est pas obligatoire et l'on peut acquérir toute la scène en une acquisition. Cette méthode ne tolère toutefois pas de grandes disparités (en termes de profondeurs) dans la reconstruction, en raison du motif sinusoïdal utilisé. Elle est de plus peu précise en cas d'objet constitué de différentes matières. En effet, la réflectance étant dépendante de la matière, il est impossible de numériser correctement sans a priori sur la matière observée.

3.2 Les silhouettes

Ces types de scanners tridimensionnels utilisent les contours créés à partir d'une séquence de photos prises autour d'un objet en trois dimensions, devant un arrière-plan contrasté. Après avoir détaché les silhouettes de leur fond, on les assemblent afin de former une enveloppe tridimensionnelle de l'objet [Sze93].

Cette méthode bien que très rapide reste très approximative, en effet les concavités de l'objet ne peuvent être détectées. De plus il est nécessaire d'acquérir la silhouette sous plusieurs angles et il est donc difficile d'effectuer les acquisitions en parallèle en travaillant avec de la lumière projetée.

3.3 La triangulation

La triangulation reposant sur la loi des triangles est sans doute la plus ancienne méthode de mesure tridimensionnelle [Bes88, Spe93]. La configuration de base d'un capteur de ce type est montrée figure I.5 et comprend une source lumineuse, un détecteur et un analyseur.

Le scanner laser par triangulation est un scanner actif qui utilise également la lumière laser pour sonder son environnement. Le détecteur peut être vu comme un appareil photographique. Le système optique est réglé de telle manière que lorsque la surface à analyser est située au milieu du champ de mesure (I.5), la lumière projetée qui est réfléchiée par l'objet sera focalisée au milieu du détecteur. Connaissant la distance a entre la source lumineuse et le détecteur, et l'inclinaison d'angle θ entre le détecteur et la source lumineuse, il est possible de retrouver la transformation opérée par un point spatial (x, y, z) lors de son acquisition dans le plan image. Si on note (u, v) les coordonnées dans le plan image dont la position de référence est le centre

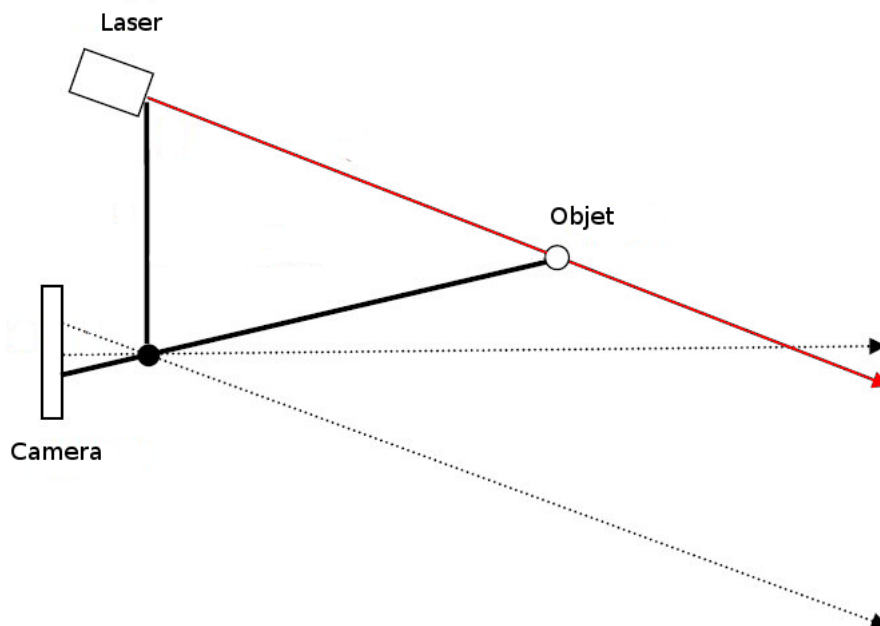


Figure I.5 – Scanner tridimensionnel par triangulation.

du capteur et f la focale du détecteur, on a :

$$(x, y, z) = \frac{a}{f \cot \theta - u}(u, v, f) \quad (\text{I.1})$$

Ce type de systèmes fournit des résultats précis jusqu'à $1\mu m$, son inconvénient principal étant la position angulaire fixe qui induit des angles morts. Une très légère variation de θ peu aussi engendrer des erreurs importantes.

Il existe des versions portables de ce dispositif dans lesquelles les positions sont enregistrées par rapport à un système de coordonnées interne et le scanner étant lui-même en mouvement sa position doit alors être mesurée. La position peut alors être déterminée par le scanner en utilisant des points de repères caractéristiques sur la surface en cours de numérisation ou en utilisant une méthode de repérage externe.

Plus médiatisé actuellement, la *Kinect* de *Microsoft* utilise cette méthode avec une longueur d'onde en dessous du spectre visible. Elle est équipée d'un projecteur et d'une caméra infrarouge qui balaye la scène à très haute fréquence. Sa précision est de l'ordre du centimètre en profondeur et varie quadratiquement avec la distance entre l'objet et l'appareil [Kho11a]. Originellement développée pour les jeux vidéo sur *Xbox*, sa version PC est de plus en plus utilisée dans le monde de la recherche pour des applications temps-réel. On notera son concurrent direct, la *Xtion Pro Live* par *Asus*.

Cette méthode est à la base de toutes les techniques utilisant un appareil d'acquisition et un appareil de projection. Encore une fois les points du nuage tridimensionnel sont acquis un à un. Il est possible de l'améliorer en structurant la lumière envoyée sur l'objet. On ne se contente plus alors d'envoyer un unique point à la fois mais tout un motif ce qui est développé dans la partie suivante.

3.4 La lumière structurée

Les scanners tridimensionnels à lumière structurée projettent un motif lumineux sur le sujet et en observent la déformation à l'aide d'une ou plusieurs caméras. Le motif peut être à une ou deux dimensions. Classiquement les motifs utilisés sont des lignes, des grilles ou des matrices de points.

Le générateur du motif peut se déplacer selon un axe de l'objet à reconstruire, de manière à balayer complètement la surface dans le cas de petits motifs. Les caméras sont soit fixes par rapport à l'objet, soit fixes par rapport au générateur. Dans ce dernier cas, on a alors généralement une seule caméra, et un montage sensiblement équivalent aux scanners par triangulation. La mesure tridimensionnelle est obtenue en analysant l'image des courbes formées par l'intersection du motif et de l'objet.

Dans le cas d'un motif linéaire, le générateur balaye le champ de vision afin d'enregistrer, une bande à la fois, les informations relatives aux distances.

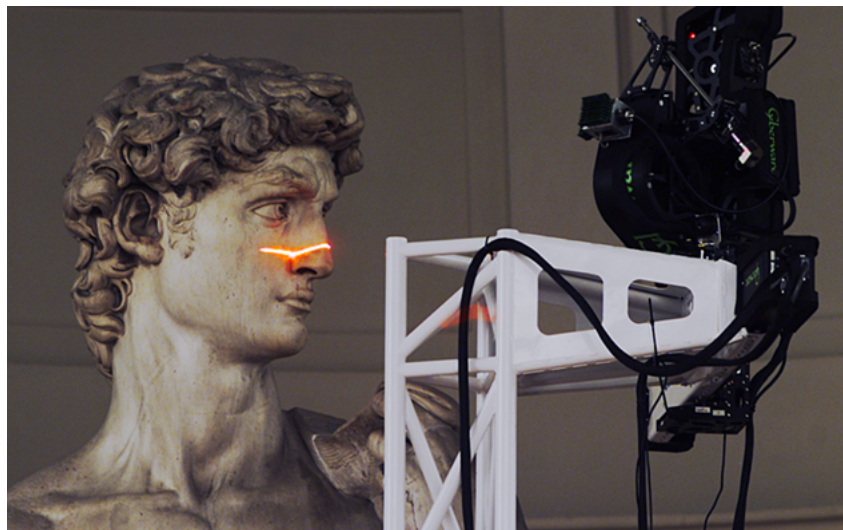


Figure I.6 — Le projet Digital Michelangelo.

Dans le cas d'un motif en forme de grille ou de bande, une caméra est utilisée pour enregistrer

les déformations et un programme informatique complexe sert à calculer les distances des points composant ce motif. La complexité est due à l'ambiguïté entre chaque bande et sa voisine. La contrainte d'ordre n'est pas forcément respectée lors de l'acquisition, suite aux phénomènes d'occultation ou de trou dans l'objet qui peuvent dans certains cas extrêmes inverser l'ordre de certaines bandes. Ce problème spécifique a été résolu par une avancée technologique appelé "Triangulation laser multibande" (en : *Multistripe Laser Triangulation*) [JKC85, PHH10]. Cette technologie est par exemple utilisée dans les lasers *InSpeck*.

Le point fort des scanners tridimensionnels à lumière structurée est leur rapidité. Au lieu de scanner un point à la fois, ils scannent tout le champ de vision. Ceci limite, voire élimine, le problème de distorsion liée au mouvement. Des systèmes existant sont capables de scanner des objets en mouvement en temps réel. Récemment, Song Zhang et Peisen Huang ont mis au point un scanner à la volée utilisant une projection d'une frange numérique et une technique à lumière structurée [ZH04]. Ce système est capable de saisir, de reconstruire et de restituer les détails d'objets se déformant dans le temps (comme une expression faciale) à une fréquence de 120 Hz, pour la totalité de la zone d'acquisition.

Cette méthode est très précise lorsque le motif est correctement détecté dans l'image d'acquisition. En utilisant des motifs construits particuliers, il est possible de reconstruire des zones plus larges qu'un unique point à la fois. Avec un motif suffisamment complexe il est alors possible de numériser toute la surface de façon discrète en une seule acquisition. Nous reviendrons sur cela dans le chapitre IV.

3.5 La stéréoscopie

Les scanners sans contact passifs (c'est-à-dire n'étant émetteur d'aucun type de rayonnement) se basent sur la détection de rayonnement ambiant réfléchi. La plupart des scanners de ce type travaillent dans le visible mais il est également possible de travailler dans l'infrarouge. Les méthodes passives peuvent être bon marché, du fait que dans la majorité des cas elles ne nécessitent pas d'appareil d'émission spécifique.

Les systèmes stéréoscopiques utilisent au moins deux appareils imageurs (caméras ou appareils photographiques), légèrement espacées, pointant vers le même objet. En analysant les différences entre les deux vues acquises, il est possible de déterminer une carte de profondeur de la scène. Cette méthode, plus largement développée dans le reste de ce document, repose sur les mêmes principes que la vision humaine.

Ce type de méthode nécessite souvent des algorithmes assez complexes et une bonne connaissance géométrique des caméras. Son principal défaut est qu'il est difficile d'estimer la profondeur

dans une zone homogène ou lorsqu'une partie est occultée dans l'une des deux vues.

La précision de ce type de méthode dépend de l'algorithme utilisé. Il existe de très nombreux algorithmes, tous partant d'un moins une paire d'images. L'acquisition se fait donc en une seule passe, mais les temps de calcul peuvent être conséquents pour obtenir la surface de la scène. Certains algorithmes sont fait pour être rapides avec un calcul très approximatif de la profondeur ("*bloc matching*" par exemple) alors que d'autres sont très précis, mais leur temps de calcul se compte alors en minutes voire en heures surtout lorsque l'on travaille avec plus de deux appareils de capture. De plus, il existe de nombreux problèmes liés aux zones d'occultation et aux zones homogènes de l'image.

Une évaluation de plus de 160 algorithmes de reconstruction par stéréovision peut être trouvé sur le site <http://vision.middlebury.edu/stereo/eval/>. Pour le cas particulier de la reconstruction à l'aide de plus que deux appareils de captures, la problématique est légèrement différente. Même si moins d'algorithmes existent pour cette problématique, le site <http://vision.middlebury.edu/mview/eval/> en évalue tout de même plus de 60.

4 Conclusion

Les méthodes par contact sont très précises mais très longues lors de l'acquisition. Elles sont de plus en plus souvent remplacées par d'autre techniques lorsque cela est possible.

Les télémètres par triangulation ou par temps de vol ont tous deux des forces et des faiblesses qui les rendent appropriés pour des situations différentes. L'avantage d'un télémètre par temps de vol est sa grande portée lui permettant de travailler sur de longues distances de l'ordre de plusieurs kilomètres. Ces scanners sont alors indiqués pour de grandes structures comme des bâtiments ou des reliefs géographiques. Leur faiblesse est leur imprécision. La précision de la mesure est de l'ordre de plusieurs millimètres. À l'opposé, les scanners par triangulation ont une faible portée, de quelques mètres, mais leur précision est relativement bonne, de l'ordre du dixième de micromètre.

Les scanners basés sur la silhouette manquant drastiquement de précision, seul certains scanners à lumière structurée et les scanners stéréoscopiques peuvent acquérir toute la surface d'un objet mobile sans obtenir de flou de bougé important.

Dans le cadre du projet BRAMMS, il a été nécessaire d'acquérir un système de numérisation tridimensionnel afin de pouvoir obtenir la surface tridimensionnelle d'un buste féminin. L'IFTH avait acquis un scanner pour une campagne nationale de mensuration en 2004 mais les résultats obtenus au niveau du buste n'étaient pas satisfaisant pour plusieurs raisons :

- la respiration de la personne causait un phénomène de bruit sur les mesures.
- les femmes portant toutes un soutien-gorge, les résultats obtenus étaient imprécis.

L'une des principales retombées pour cette thèse est donc l'élaboration d'un système d'acquisition permettant d'écourter au maximum la durée d'acquisition, afin de limiter le flou de bouger, tout en permettant de reconstruire la surface avec le plus d'autonomie possible. Ce système d'acquisition se doit de plus respecter l'intimité des personnes lors de son utilisation.

Cette étude sera présentée au prochain chapitre. La solution retenue se base sur de la lumière structurée et de la stéréovision car ce sont les seuls méthodes permettant une acquisition instantanée de toute une scène.

Lors de la fin de la rédaction de ce manuscrit, je me suis rendu compte de l'existence d'une thèse similaire qui a eu lieu en parallèle. Cette thèse de *laurea magistrale* intitulée "A Low-Cost, Efficient And Accurate Human Body Scanner" a été soutenue en 2012 par Christian Piccolo à l'Université *Degli Studi* de Padoue, c'est à dire à peu près en même temps que notre cabine a été construite. Elle décrit la conception d'un système de métrologie tridimensionnelle basée sur une approche par lumière structurée. Ses conclusions quant au type de matériel à utiliser concordent avec ce que nous avons établi dans la conception de la cabine d'acquisition qui sera présenté au prochain chapitre. Cependant, nos avis diffèrent drastiquement quant à la méthode à employer concernant le protocole d'acquisition. En effet, la méthode proposée par Christian Piccolo se passe en plusieurs acquisition d'un motif binaire alors que nous nous sommes efforcé de travailler en une unique acquisition d'image afin de pallier au maximum le biais d'acquisition lié au flou de mouvement.

Chapitre II

Mise en place d'un dispositif d'acquisition expérimental

Ce chapitre présente les travaux liés à l'élaboration d'une cabine d'acquisition surfacique. Dans le cadre du projet Bramms, l'un des buts premiers de cette cabine est de pouvoir numériser des surfaces tridimensionnelles sur des personnes vivantes, et notamment des bustes féminins.

La méthode de scanner utilisée lors de la campagne de mesure nationale de mensuration de l'IFTH en 2004 avait donné des résultats imprécis au niveau du buste. Un nouveau dispositif d'acquisition devait donc être mis en place.

Les contraintes imposées à ce dispositif expérimental étaient les suivantes :

- acquérir la surface de manière instantanée.
- garantir l'intimité de la personne lors de l'acquisition.
- permettre le déclenchement de l'acquisition depuis l'extérieur de la cabine.

Nous nous sommes alors tournés vers un dispositif combinant une méthode multi-vue (prise de photographies simultanément par plusieurs appareils photographique) et une méthode par lumière structurée (projection d'un motif). Ce sont en effet les deux seuls méthodes permettant une acquisition en une fois de toute une scène et réponde donc à la première contrainte.

Dans ce chapitre sera d'abord discuté le choix et la disposition des appareils de la cabine. Ensuite, son fonctionnement sera expliqué plus en détails dans la section "Interaction cabine - ordinateur". Enfin, le dernier point traité dans ce chapitre portera sur l'étalonnage de la cabine.

1 Choix et disposition des appareils de captures

Dans le chapitre I, nous avons exposé un rapide état de l'art des différentes techniques pouvant être utilisée pour acquérir la surface tridimensionnelle d'un objet. Dans le cas non plus d'objets, mais d'êtres vivants, il est nécessaire de prendre en compte le flou de bougé dû à la

respiration et au micro-mouvements involontaires du sujet. C'est pourquoi seules les techniques avec un temps d'acquisition très faible peuvent ici être retenues. Seules deux familles de méthodes permettent un temps d'acquisition adéquat à des sujets vivants : certaines méthodes par lumière structurée et les méthodes stéréoscopiques.

La peau humaine nous livre un autre dilemme. Son uniformité empêche un traitement stéréoscopique classique permettant un rendu de haute précision. En effet, la stéréovision ne permet pas de différencier un pixel issu d'une vue centrée dans une zone homogène, d'un autre pixel d'une vue centrée dans une autre zone homogène. Sur un visage humain, il y a peu de pixels concernés par ce genre de situation à cause des rides, des contours des différents éléments (yeux, nez, bouche, etc.), de la pilosité faciale, grains de beauté, et autres marqueurs singuliers. Le problème est de fait bien plus fréquent sur une large portion de peau comme dans le dos ou sur le buste. Il est donc nécessaire d'y ajouter de l'information. A ce stade, il ne reste donc qu'une seule méthode possible : celle basée sur de la lumière structurée.

La stéréovision pure aurait pu tout de même être employée mais aurait nécessairement entraîné une plus grande imprécision. Or, le but est ici non seulement d'acquérir la surface 3D, mais aussi d'obtenir des résultats compétitifs en termes de précision et de temps d'acquisition.

Afin d'éviter une fois de plus le flou de bougé lors de l'acquisition, il a été nécessaire de trouver un moyen de numériser toute la surface en une seule acquisition rapide. A cette fin nous avons combiné la méthode par lumière structurée avec une méthode multi-vue (stéréovision appliquée à plus de 2 caméras).

Concrètement, le dispositif employé est donc constitué de plusieurs appareils photographiques répartis tout autour de la personne ou de l'objet à numériser afin de pouvoir appliquer une méthode multi-vue. À cela s'ajoute des vidéoprojecteurs envoyant un motif servant à différencier les points dans les zones qui seraient homogène sans cela.

1.1 Appareils de captures d'images

Après un rapide état des lieux, nous nous sommes portés sur la technologie d'appareils photographique *reflex*. Ce type d'appareil est en effet dédié à la photographie, et a donc une meilleure résolution que la plupart des caméras à moindre coût. De plus, étant donné les multiples inconnues liées à l'assemblage de la cabine d'acquisition lors de sa création, et ensuite vis-à-vis de ses utilisations probables, il était nécessaire de pouvoir ajuster notre outil de capture. Avec l'essor du numérique dans la technologie *reflex*, ce type d'appareil photographique est devenu plus abordable. Ce type de technologie offre de plus la possibilité d'être piloté via un ordinateur, ce qui nous servira par la suite.

Les spécifications étaient donc les suivantes :

- Avoir une bonne résolution
- Avoir une palette de sensibilité ISO¹ (ie : sensibilité à la lumière) pouvant acquérir dans un environnement comportant peu de lumière.
- Avoir un champ de vision englobant le champ de reconstruction souhaité (ce qui influe directement sur l'objectif).
- Avoir la possibilité de piloter les appareils via l'ordinateur.
- Avoir une bonne cadence d'acquisition, ou un mode vidéo pour des acquisitions évoluant dans le temps.
- Optimiser la dimension budgétaire du projet.

Marque	Modèle	Monture	ISO max	Taille CCD (mm)	Résolution (px)	Rafale (img/s)
Canon	EOS 1000D	EF/EF-S	1600	22,2 × 14,8	3888 × 2592	3
	EOS 1100D	EF/EF-S	6400	22,2 × 14,7	4272 × 2848	3
	EOS 500D	EF/EF-S	12800	22,3 × 14,9	4752 × 3168	3,4
	EOS 600D	EF/EF-S	12800	22,3 × 14,9	5182 × 4356	3,7
Nikon	D3000	AF-S DX FX	1600	23,6 × 15,8	3872 × 2592	3
	D5000	AF-S DX FX	3200	23,6 × 15,8	3872 × 2592	3
Sony	Alpha SLT33	AF/KM	12800	23,4 × 15,6	4592 × 3056	7

Tableau II.1 – Étude comparative des différents appareils photographiques numériques. Seules les principales caractéristiques sont ici exposées par soucis de clarté. La liste a été réduite aux appareils fournissant des données brutes (RAW) pouvant être supporté par le financement fixé par la région.

Notre choix s'est porté sur l'appareil photographique Canon 600D après étude comparative entre divers modèles, voir Tab II.1. C'est en effet le modèle rentrant le mieux dans les spécifications. Un autre point important en faveur de la marque *Canon* est qu'elle propose une bibliothèque EDSDK² permettant la gestion complète de l'appareil photographique via USB³. Cette librairie nous donnant directement les codes utilisés pour communiquer avec l'appareil photographique numérique, il a été possible de fournir un certain nombre d'ajouts logiciel à la cabine.

24 appareils *Canon* ont ainsi été intégrés à une structure métallique, disposés en deux cercles de 12 appareils chacun. Chaque appareils a donc un écart de 30° avec son voisin, voir Figure II.1. Afin d'assurer une distance de mise au point minimale, le diamètre de la structure métallique a

1. *International Organization for Standardization* est l'organisation qui a donnée son nom à cette mesure
2. EOS (*Electro-Optical System*) *Digital Camera Software Development Kit*
3. *Universal Serial Bus* est un port d'échange de donnée

Mise en place d'un dispositif d'acquisition expérimental

été fixé à 3m afin d'avoir un champ de vision cylindrique d'1m de rayon pour la reconstruction.

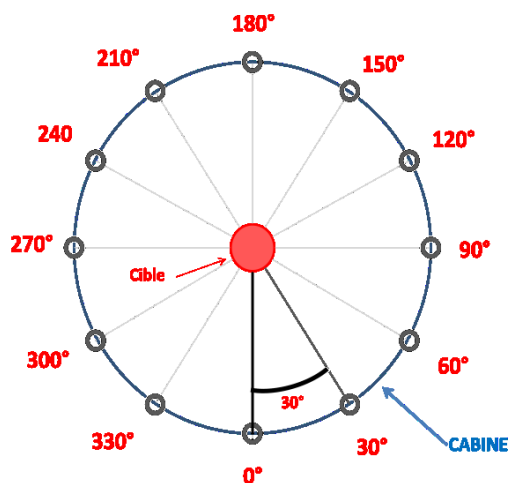


Figure II.1 – Disposition des appareils photographiques numériques. Diamètre de la cabine : 3m. Diamètre de la zone de reconstruction (centrale) : 1m.

Toutes les méthodes de reconstruction utilisant un capteur CCD sont dépendantes de la focale. Afin d'obtenir des données stables, il est nécessaire de n'utiliser que des focales fixes. Le calcul du cercle optique (Figure II.2) à 1m et à 1,5m fournissent des informations discriminantes complémentaires. Afin d'acquérir l'objet situé dans la cabine, il est nécessaire que chaque appareil puisse voir une large portion de l'objet. Il est toutefois inutile de voir une trop large portion. Le choix s'est porté sur le "Canon EF 50mm f/1.6 II" pour ses caractéristique concordant avec le projet (voir Tab.II.2).

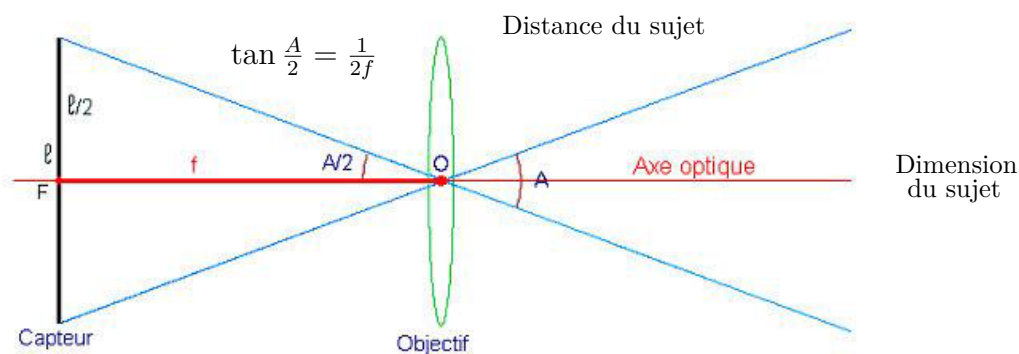


Figure II.2 – Schéma de calcul du cercle optique.

Marque	Modèle	f	Ouverture Mini - Maxi	Angle de champ diag.	Champ à 1m (m)	Champ à 1.5m (m)
Canon	EF-S Macro USM	60	32 – 2,8	24,3°	0,38 × 0,25	0,58 × 0,38
	EF II	50	22 – 1,8	46°	0,47 × 0,30	0,7 × 0,46
	EF Comp. macro	50	32 – 2,5	46°	0,47 × 0,30	0,7 × 0,46
Sigma	DC EX	30	NC – 1,4	45°	0,86 × 0,53	1,29 × 0,79
	EX DG HSM	50	NC – 1,4	46,8°	0,47 × 0,30	0,7 × 0,46
Tamron	SP Di II macro	60	22 – 2	26°	0,38 × 0,25	0,58 × 0,38

Tableau II.2 – Étude comparative des différents objectifs. Seuls les objectifs étudiés s'adaptant sur le Canon 600D avec une focale fixe (nécessaire pour la précision) sont ici exposés par soucis de clarté. Adapté sur le 600D, la focale f est multipliée par 1,6.

1.2 Appareils de projection

Afin de pouvoir travailler avec des motifs structurés, il est nécessaire de les projeter. Voulant rester sur un système bas coût, nous nous sommes tournés vers la vidéo-projection plutôt que sur une technologie laser. Il existe plusieurs types de technologies utilisées par les vidéoprojecteurs. Leur présentation donnera lieu à une discussion sur le choix effectué.

vidéoprojecteurs LCD⁴

Cette technologie est basée sur des écrans à cristaux liquides de petite taille, mais de très haute résolution. La lumière d'une lampe est divisée en trois faisceaux pour passer à travers trois panneaux à cristaux liquides, qui correspondent aux trois couleurs fondamentales. Les trois images sont ensuite recomposées pour n'en faire plus qu'une, voir Figure II.3.

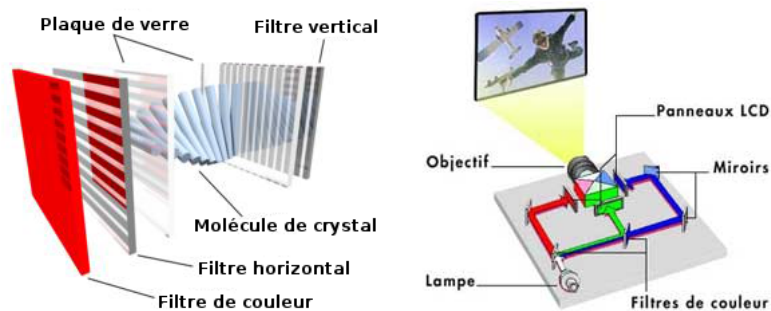


Figure II.3 – Technologie LCD. À gauche, la structure interne d'un panneau LCD. Les trois couleurs sont ensuite assemblées comme dans la figure de droite.

Cette technologie implique un quadrillage lors de la projection et celui-ci peut être vu en

4. *Liquid Crystal Display*

s'approchant de l'écran d'affichage. Cet effet est plus connu sous le nom de “*screendor*”, voir Figure II.4.

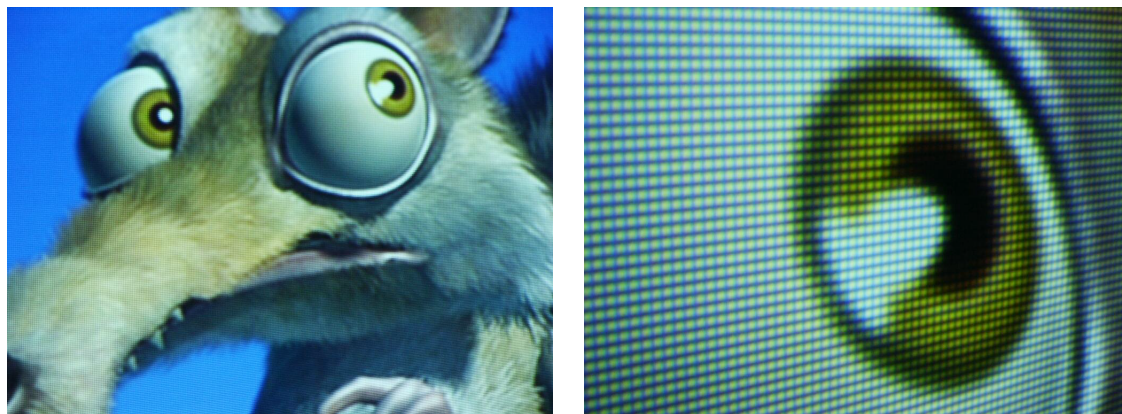


Figure II.4 – Effet *Screendor*.

vidéoprojecteur DLP ⁵

Ces vidéoprojecteurs utilisent une seule puce numérique qui fonctionne grâce à des centaines de milliers de miroirs oscillants microscopiques (DMD ⁶). Pour obtenir la couleur, entre la matrice DLP et la lampe, la lumière passe par un filtre tricolore en rotation, voir Figure II.5.

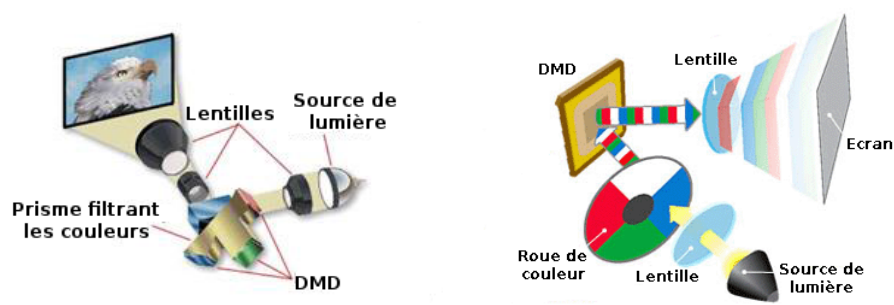


Figure II.5 – Technologie DLP. Les deux schémas représentent deux montages possibles pour ce type de technologie.

Avec cette technique, certaines personnes peuvent percevoir de temps en temps des petits arcs-en-ciel sous forme de flashes pendant la projection. Cet effet “arc-en-ciel” arrive dans le cas d'un dispositif utilisant une roue de couleur. Classiquement, la fréquence de rafraîchissement de l'œil humain nous permet de voir l'image correctement, mais dans le cas d'un dispositif

5. *Digital Light Processing*

6. *Digital Micromirror Device*

de capture ayant un temps d'ouverture plus court, cet effet est facilement détectable, voir Figure II.6.



Figure II.6 – Effet “Arc-en-ciel” dû à la roue de couleurs.

Vidéoprojecteur LCoS⁷/SXRD⁸/D-ILA⁹

Une source lumineuse émet un rayon de lumière qui est ensuite séparé par un prisme en trois faisceaux. Chacun frappe les panneaux LCoS/SXRD réfléchissants correspondant à sa couleur (rouge, vert ou bleu) et se charge de ladite couleur. Enfin, la dernière étape consiste à fusionner ces trois rayons avant leur passage dans la lentille de projection, voir Figure II.7.

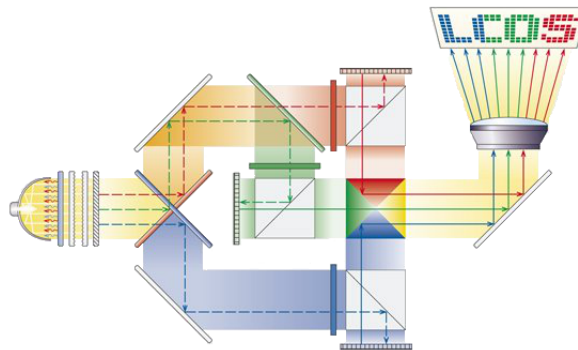


Figure II.7 – Technologie LCoS.

Cette technologie permet de faire totalement disparaître le quadrillage visible sur les images projetées par un vidéoprojecteur LCD ainsi que l'effet “arcs-en-ciel”.

7. *Liquid Crystal on Silicium*

8. *Silicon Crystal [X-tal] Reflective Display*

9. *Digital Direct Drive Image Light Amplifier*

Afin de pallier aux effets indésirables “*screendor*” et “arc-en-ciel”, nous nous sommes tournés vers la technologie LCoS. Celle-ci est pénalisée en termes de rendement optique par la complexité de sa structure, qui entraîne un mauvais rapport de contraste. Seuls deux constructeurs ont persévéré en développant chacun leurs propres solutions : Sony, avec son SXRD, et JVC, avec le D-ILA.

Sony et JVC ont tous deux identifié un problème de planéité du substrat de silicium. Ce problème a été corrigé à l'aide d'un polissage chimique. JVC a de plus constaté que des irrégularités subsistaient autour de l'électrode et du passage du conducteur qui assure le contact électrique entre l'électrode et la partie de pilotage. Ce creux a été comblé par un matériau inorganique. Ajoutez à cela de nouveaux prismes polariseurs, et la technologie D-ILA a vu passer son contraste natif de 2000 :1 au taux exceptionnel de 12000 :1.

Les vidéoprojecteurs conventionnels LCoS sont des appareils dont la technologie est très récente et développée par peu de constructeurs. Par conséquent le prix de ces appareils reste très élevé (de l'ordre de 2000?) par rapport aux projecteurs de type LCD ou DLP. Seuls les mini-projecteurs de cette famille sont d'un coût raisonnable.

Le modèle qui a finalement été retenu pour cette étude est le “PHILIPS PPX 1430”, appartenant à cette famille de vidéoprojecteur. Sa résolution (1280×768), sa distance d'affichage (0,2 à 3m), sa luminosité (30lm) et son contraste (500 :1) sont suffisants pour notre application. De plus il est capable de projeter les informations contenues dans une carte SD, permettant ainsi de limiter la câblerie de la cabine.

Leur disposition dans la cabine a été soumise à plusieurs contraintes : leur répartition sur 360° , le chevauchement des motifs, et leur positionnement afin d'observer le motif projeté de façon symétrique.

La symétrie d'un groupe de capteur est une contrainte importante afin d'uniformiser le plus possible la luminance de la scène (réflexion du motif projeté sur l'objet) dans les deux vues. Partant de cela, le plus simple était de constituer des blocs formés de deux appareils de captures entourant un appareil de projection. Cette disposition demande l'utilisation de 6 vidéoprojecteurs pour 12 colonnes d'appareils photographique. Bien sûr, il aurait aussi été possible de placer les video-projecteurs sur une colonne d'appareils de capture, mais il aurait soit fallu 12 vidéoprojecteurs, avec les soucis de recouvrement que cela implique alors, soit la contrainte de symétrie dans la luminance aurait été rompue. En effet, les appareils de capture n'ayant pas de vidéoprojecteurs dans leur colonne auraient capté moins de luminance dans le cas d'objet cylindrique, ce qui est la forme moyenne la plus courante des objets à numériser.

Par la suite, la méthode utilisée bien que basée sur la luminance (voir chapitre III) ne pourra pas tenir compte de cette différence de l'asymétrie existante.

N'ayant aucun à priori sur la forme de l'objet à numériser, il est impossible de définir une largeur maximale pour les motifs. Lors de la formation des images sur un capteur CCD, les photons s'empilent dans trois type de cellule, leur quantité faisant varier l'intensité lumineuse. Les trois cellules possibles sont constituées de la même façon, avec un filtre permettant de laisser passer les photons correspondant à une couleur de référence. Ces trois couleurs sont le rouge, le vert et le bleu. Il semble donc naturel, pour dissocier les motifs projetés, de se baser sur ce même principe. Les motifs seront donc projetés dans une des trois couleurs de référence, les deux projecteurs voisins se devant de projeter dans une autre couleur, voir Figure II.8.

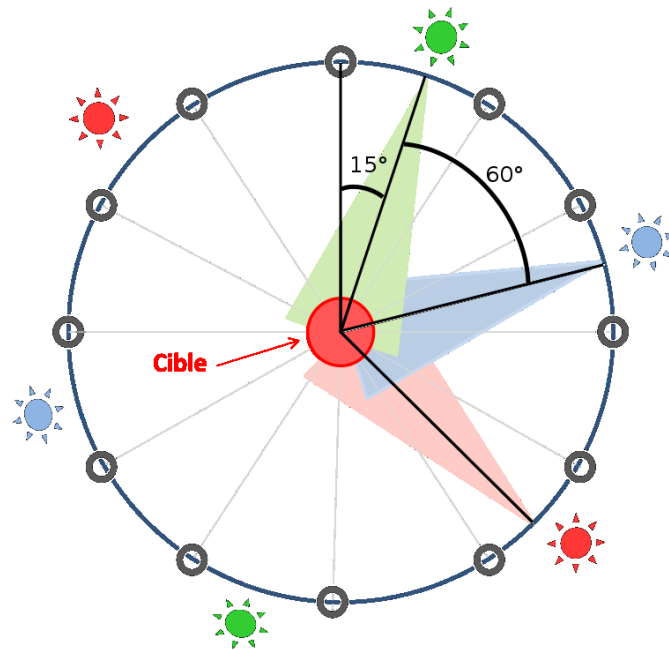


Figure II.8 – Disposition des vidéoprojecteurs.

1.3 Assemblage

Afin de pouvoir réutiliser la cabine pour d'autres applications, elle a été pensée de façon modulable dès sa conception, en utilisant les composants universels du système modulaire de PROFILES de Bosch Rexroth, voir Figure II.9. Ces composants ont divers intérêts tels que la facilité et la rapidité de montage, tout en offrant une structure solide et facilement démontable.

Afin de pouvoir aisément faire rentrer les objets constitutifs de la scène à numériser, l'un

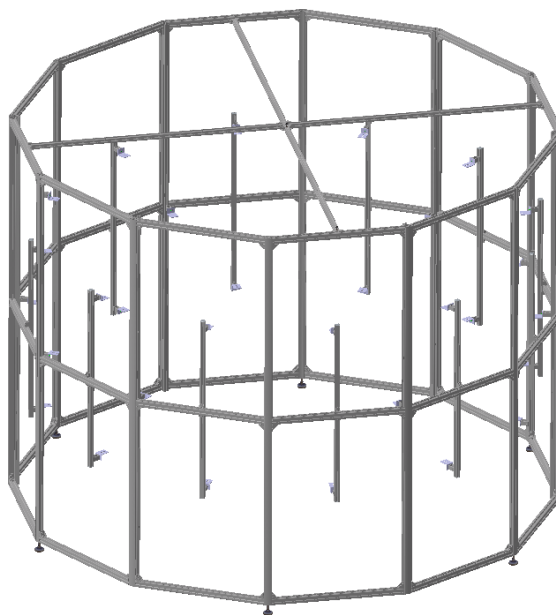


Figure II.9 – Structure de la cabine.

des pans du cylindre à douze faces de la cabine est amovible, et permet un basculement selon un axe vertical vers l'intérieur. Il est à noter que les deux appareils d'acquisitions situés sur cette "porte" sont les moins précis. En effet, lors de l'acquisition de la surface, les paramètres géométriques de la disposition des capteurs doivent être très précisément connus. Or même avec un système de verrouillage du pan mobile, les appareils de la porte sont soumis à des vibrations répétées lors de son ouverture et sa fermeture ce qui entraîne une dégradation progressive et inéluctable de la géométrie. Ce problème peut se pallier en faisant régulièrement un étalonnage (ci-après nous utiliserons l'anglicisme *calibration*, plus couramment utilisé dans le domaine de l'imagerie) des appareils d'acquisition. Notons ici que la *calibration* peut être limitée à ces deux appareils puisque la position des autres appareils est supposée fixe et connue.

Les interactions avec l'extérieur de la cabine, essentiellement lumineuses, ont été limitées à l'aide d'un tissu noir épais entourant la cabine. La pose d'un plafond en contreplaqué a permis de finir l'isolement optique du système, tout en ayant l'avantage d'offrir un support stable à l'ensemble de la câblerie. De plus, des prises de courant ont été aménagées au sein de la structure afin de pouvoir brancher les appareils de capture et de projection directement sur secteur. Ainsi, seuls quelques câbles relient les prises internes de la cabine au réseau électrique, limitant les manipulations à la mise en route de la cabine.

La liaison entre la cabine et l'ordinateur est assurée via USB. Il a donc été nécessaire de tester différents câblages entre les appareils constituant la cabine. Deux problèmes ont été rencontrés

essentiellement. Le premier est le grand nombre d'appareils à connecter à l'ordinateur ; le second, la longueur des câbles USB.

La bonne transmission de l'information est reliée à la longueur des câbles. Il est donc important de s'assurer que ceux-ci ne sont pas trop longs. De plus, afin d'être précis, il est nécessaire d'avoir la même longueur de câble entre l'ordinateur et les différents appareils de la cabine. Afin de résoudre ces deux points, une topologie en étoile a été réalisée. L'ordinateur ne pouvant pas se trouver physiquement au centre de tous les appareils, les câbles USB qui partent de celui-ci raccordent d'abord le centre de la cabine, avant d'être repartis entre les différents appareils.

Cette topologie en étoile nous permet aussi de limiter le nombre de ports USB nécessaires. En effet, il a été déterminé que le déclenchement par USB via la bibliothèque EDSDK ne permettait pas le déclenchement simultané des différents appareils de capture (voir Sec. 2.2). Nous avons donc pris le parti de limiter le nombre de port nécessaire à 2 en utilisant des concentrateurs USB (plus communément appelés "*hub*"). Deux concentrateurs ont donc été fixés sur le toit de la cabine afin d'assurer la distribution d'information en étoile.

La câblerie utilisée pour le déclenchement — gérée séparément du paramétrage, voir section 2.2 — utilise une topologie similaire. Les concentrateurs sont cette fois remplacés par une interface de contrôle.

Enfin, le toit de la cabine est équipé d'une source lumineuse afin de permettre l'installation de la scène. Un système de ventilation y a aussi été intégré afin de pallier la chaleur dégagée par l'ensemble du dispositif. Ces deux ajouts sont pilotés par deux interrupteurs situés près de la porte de la cabine.

D'un point de vue extérieur, trois câbles USB sont branchés à l'ordinateur (deux pour le paramétrage et un pour le déclenchement). L'alimentation de la cabine se fait par le branchement de trois prises de courant alimentant pour l'une la lumière, les ventilateurs, les concentrateurs et le dispositif de déclenchement, et pour les deux autres les appareils d'acquisition et de projection.

2 Interaction cabine - ordinateur

La bibliothèque EDSDK2.1 permet le pilotage entier des appareils photographiques numériques canon 600D par liaison USB. Dans cette partie, le système de déclenchement, ainsi que l'interface développée entre la cabine et l'ordinateur vont être développées.

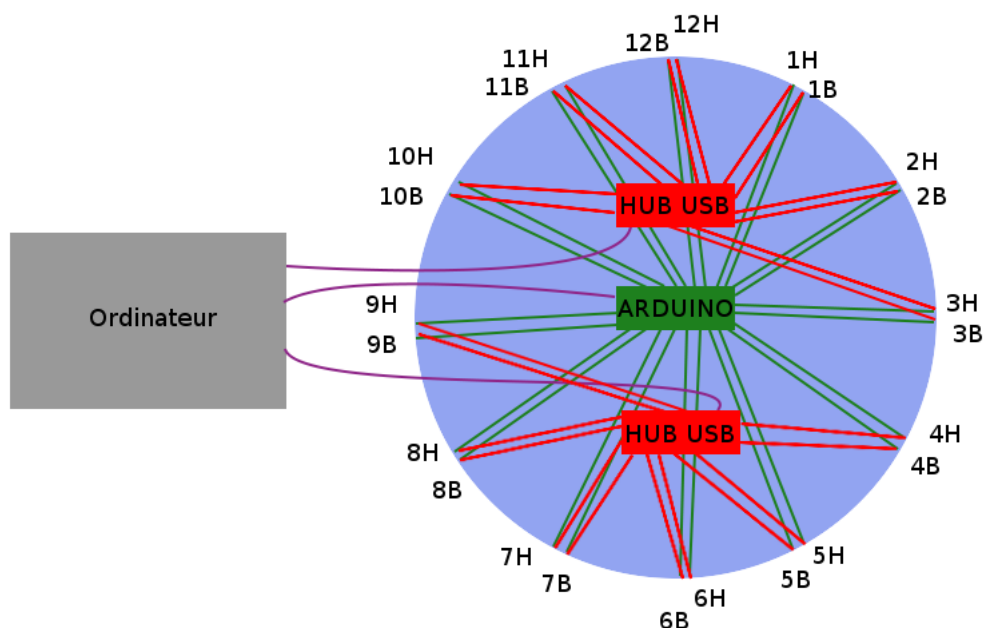


Figure II.10 – Cablage de la cabine.

2.1 Interface

La bibliothèque EDS SDK permet l'accès à l'appareil photographique et à ses fonctions. L'appareil dispose via ce biais des mêmes options de configurations qu'en mode manuel. Il est donc possible de le configurer entièrement depuis l'ordinateur afin de répondre aux critères requis pour une bonne acquisition. En pratique, les paramètres d'ouverture, d'ISO et de vitesse d'obturation changent peu à cause de l'environnement confiné. Cependant, il est possible qu'il faille affiner ces paramètres pour une expérience particulière.

Afin d'interagir avec les appareils d'acquisitions, l'interface inspecte à son lancement les appareils connectés via USB à l'ordinateur. Des interfaces sur ces appareils sont stockées dans un tableau de données nous permettant d'y avoir accès séparément.

L'un des points importants étant que tous les appareils doivent être configurés de la même façon, l'interface qui a été développée pour le pilotage de la cabine via l'ordinateur se devait de respecter cette contrainte. Pour se faire, chaque modification des paramètres d'acquisition entrée via l'interface (Figure II.12) est reportée sur chaque appareil photographique.

Le mode EVF¹⁰, qui permet normalement d'afficher sur l'écran LCD d'un appareil photographique ce qui est dans le champ optique de l'appareil, peut être détourné pour être affiché directement à l'écran. Même s'il est malheureusement impossible de le faire pour plus d'un

10. *Electronic View Finder*

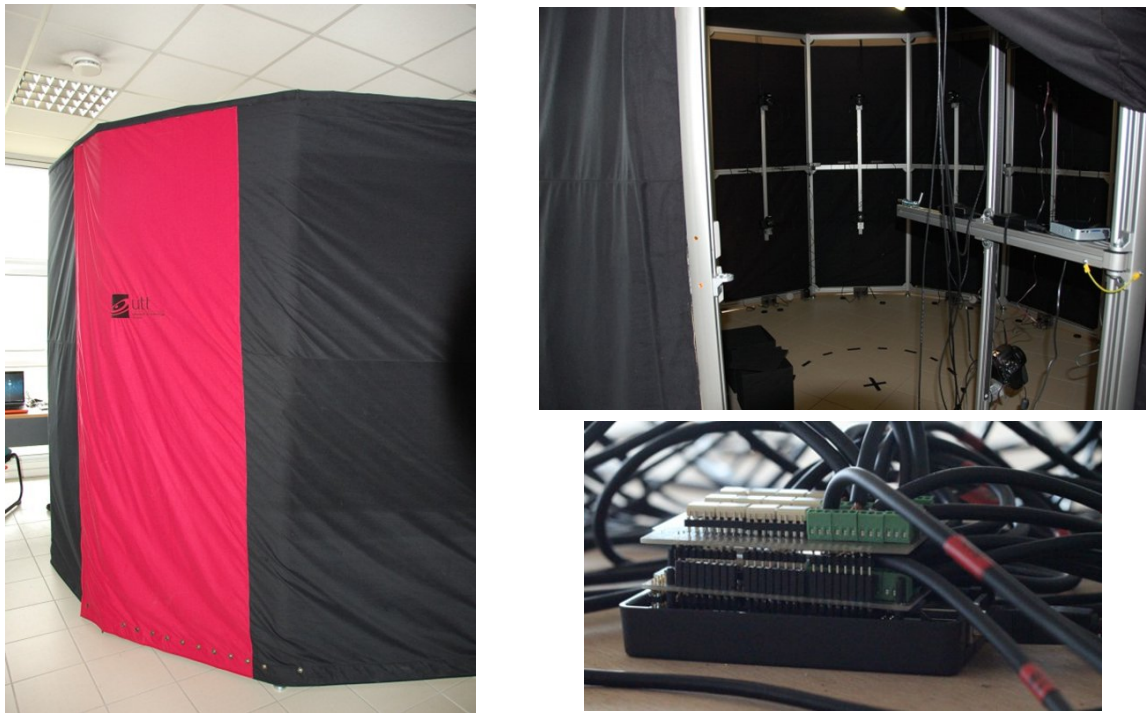


Figure II.11 – Photographies de la cabine d'acquisition. On peut voir en bas à droite le montage sur la plaque Arduino pour la gestion du déclenchement.

appareil à la fois, il est possible de s'en servir sur un appareil (par défaut, celui sur la porte en haut), voir Figure II.12. Il est ainsi possible de contrôler ce qui se passe à l'intérieur de la cabine. Ceci peut, entre autre, être utilisé afin de ne pas déclencher d'acquisition inopportune si les conditions de prise ne sont pas réunies.

A terme, l'interface devrait aussi assurer le rapatriement des données qui est pour l'instant assuré par une autre routine. Il est techniquement possible de récupérer une miniature de la dernière photographie sur chaque appareil bien que cela soit couteux en temps. Cette fonctionnalité a donc été abandonnée, car elle empêche l'acquisition rapide de plusieurs scènes successives.

2.2 Système de déclenchement

Le déclenchement de l'acquisition photographique peut être entièrement réalisé via USB grâce à la bibliothèque EDSDK. Toutefois, divers tests ont montré que le déclenchement simultané par USB, directement via la bibliothèque EDSDK, était impossible. En effet, la bibliothèque réserve les ports USB un à un afin d'y envoyer l'information. Cette information étant

Mise en place d'un dispositif d'acquisition expérimental

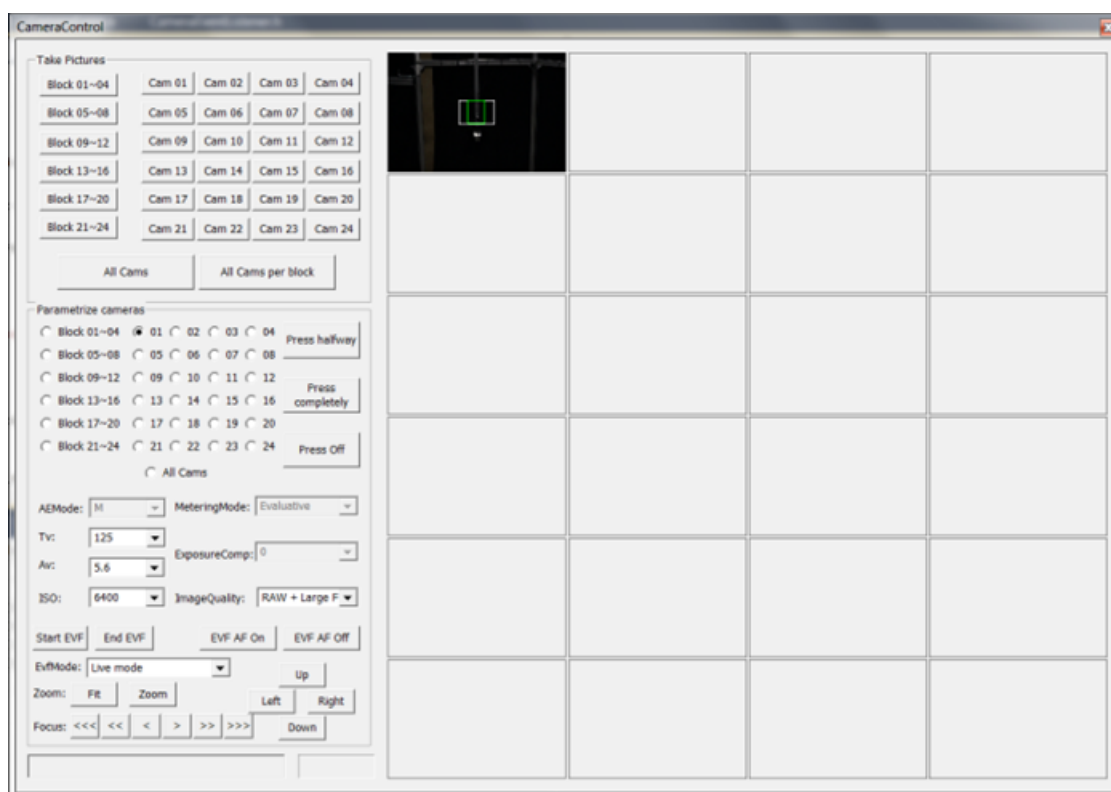


Figure II.12 – Interface. Mode EVF détourné et affiché.

de type asynchrone, et nécessitant une réponse, il n'était pas possible d'envoyer sur deux ports en même temps le même mot binaire (et donc la même commande).

Les appareils 600D sont équipés d'un port pour périphérique de déclenchement. Après renseignement, il s'est avéré que cette commande était en réalité un dispositif de fermeture de circuit. Il est donc possible de la simuler à l'aide de deux interrupteurs comme le montre la figure II.13. Lorsque l'on ferme le circuit correspondant au déclenchement, l'appareil démarre la phase d'autofocus puis passe à l'acquisition une fois le focus adéquat déterminé.

Cette commande permet de plus de gérer séparément le déclenchement de l'autofocus de chaque appareil séparément. Il est aussi possible de bloquer le focus d'un appareil. En effet, afin de pouvoir acquérir les photographies simultanément, il est nécessaire que la phase d'acquisition démarre au même moment sur tous les appareils. Or la durée nécessaire pour effectuer l'autofocus varie selon l'appareil, l'illumination et la scène.

La méthode retenue consiste donc à effectuer une phase de focus ayant un temps assez long pour que tous les appareils l'aient finie, et ensuite d'effectuer le déclenchement. Pour cela, il suffit d'effectuer une phase de détection du focus sans relâcher la connexion. Une fois le focus adéquat

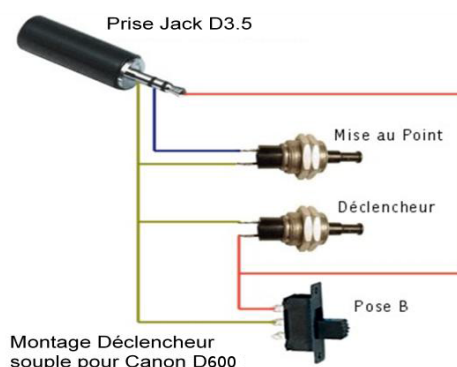


Figure II.13 – Circuit de déclenchement pour un appareil photographique de type 600D.

acquis sur chaque appareil, il est nécessaire de fermer le circuit correspondant au déclenchement de l'acquisition sur tous les appareils. Le circuit correspondant à l'autofocus étant toujours fermé, celui-ci ne se déclenchera pas et l'appareil acquerra directement la photographie.

Afin de pouvoir automatiser ce procédé et permettre le déclenchement depuis l'ordinateur, il a été nécessaire d'établir un circuit permettant de gérer le déclenchement du focus et le déclenchement de l'acquisition sur chaque appareil. Ceci a été réalisé grâce à l'utilisation d'une interface de contrôle. A cause des limitations dues au grand nombre d'appareils à piloter — chaque appareil se devant d'avoir ses propres circuits pour éviter les interactions inter-appareils photographiques — l'interface de contrôle se devait de pouvoir gérer au moins 48 sorties (2 par appareil). La solution finalement retenue utilise une carte Arduino Mega, voir Figure II.14. Cette carte possède en outre l'avantage de pouvoir être gérée comme un montage embarqué possédant un flux de communication USB avec un ordinateur.

2.3 Flux de récupération des données

L'analyse des flux USB est une étape incontournable dans l'optimisation du transfert de données. En effet, lors d'une acquisition de la cabine, 24 photographies sont acquises et celles-ci doivent être transférées à l'ordinateur pour le traitement. De plus, afin d'effectuer une reconstruction précise, il est nécessaire d'acquérir des photographies d'assez haute résolution. Avec les appareils 600D, une photographie haute résolution dans un format non compressé — afin de ne perdre aucune donnée — prend environ 25Mo. Cela fait donc 24 photographies de ce poids à transférer sur l'ordinateur depuis 24 sources différentes et cela en un minimum de temps.

Grâce aux cartes de stockage, il est possible d'acquérir plusieurs scènes avant de les transférer. Sur une campagne d'acquisition d'une quarantaine de scènes, le transfert des photographies peut donc prendre deux heures lors d'un traitement manuel. Le but est donc d'automatiser ce

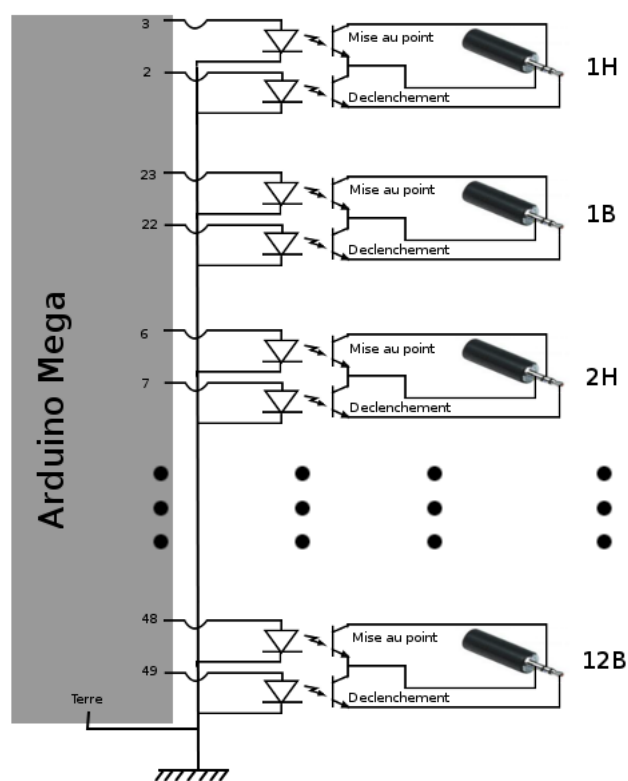


Figure II.14 — Circuit de déclenchement simultané utilisant une interface Arduino. Le brochage sur l'arduino est donné en annexe [B.3](#)

procédé en profitant du fait que chaque appareil photographique est connecté via USB à l'ordinateur, et de l'accélérer au maximum.

Afin d'automatiser le transfert des informations, il est nécessaire de pouvoir identifier chaque appareil photographiques. Cette opération, qui peut paraître simple, a demandé de nombreux tests du fait de la gestion interne des fichiers par l'appareil 600D. Une solution a cependant été trouvée et est exposée dans la notice d'utilisation de la cabine (voir Annexe [B.2.3](#)). Les appareils sont donc tous identifiés à l'ouverture de l'interface.

La dernière étape consiste donc à transférer les photographies dans une arborescence correspondant à notre cabine (voir Annexe [B.2.3](#)). Puis on procède à l'effacement sur les cartes mémoires des appareils photographiques.

Une fois les photographies transférées, tout est en place pour assurer le bon déroulement de la partie algorithmique de la cabine. Ce transfert peut toutefois être optimisé. Outre le fait de préférer de l'USB3 à l'USB2 et l'USB2 à l'USB1, il faut aussi vérifier que les deux ports USB utilisés — nous parlons ici des deux câbles USB reliés aux appareils photographique et il n'y a aucun prérequis pour celui relié à l'interface de déclenchement — correspondent à des

ports USB matériels différents (à ne pas confondre avec la notion de port physique, on parle ici de sorties gérées séparément sur la cartes mères, et non de sorties PC différentes). En effet, une sortie USB sur carte mère — port matériel — peut donner naissance à plusieurs sorties extérieures de connexion — port physique. Si les ports matériels sont différents, il seront gérés par des files d'exécution (plus communément appelé *thread*) différents ce qui permettra de diviser le temps de transfert des données par deux dès lors que le processeur gère plusieurs *thread*, ce qui est le cas de quasiment tout ordinateur de nos jours.

3 *Calibration* du système

Afin d'obtenir avec précision la géométrie exacte qui relie chaque capteurs — paramètres extrinsèques (position et orientation relatives) — ainsi que la géométrie interne de chaque capteur — paramètres intrinsèques (focal, coordonnées du centre optique sur le capteur, facteur d'orthogonalité) — il est nécessaire de procéder à une étape de *calibration* de la cabine d'acquisition. La précision de ces paramètres influe directement sur la précision de la reconstruction tridimensionnelle. C'est donc une étape capitale dans le procédé.

Il n'est toutefois pas nécessaire d'effectuer cette étape avant chaque acquisition. Il faut en effet l'effectuer après chaque changement géométrique. C'est à dire, comme dans notre cas nous utilisons des focales fixes, un changement de position ou d'orientation des appareils.

Le principe de l'étalonnage des caméras est connu depuis des décennies en photogrammétrie [STH⁺80]. Une première famille de méthodes utilise un objet de référence, ou mire, de dimensions parfaitement connues. La seconde, au contraire, s'affranchit de cette contrainte ; on parle alors d'auto-étalonnage ou *autocalibration*.

Étalonnage avec un objet de référence ou mire

Les premières méthodes d'étalonnage nécessitaient l'utilisation d'une mire tridimensionnelle. Cette mire doit être de grande précision, ce qui la rend couteuse et son emploi demande généralement beaucoup d'opérations. En effet, la diversité des applications nécessite généralement de recourir à plusieurs mires adaptées à chaque champ de vue, voir Figure II.15.

La précision de l'estimation des paramètres dépend non seulement de la précision de la mire dans sa conception, mais aussi du soin apporté aux mesures faites dans l'image. Une première amélioration est proposée sous la forme d'une optimisation multi-images par ajustement de

Mise en place d'un dispositif d'acquisition expérimental

faisceaux [LVD98], qui permet de ré-estimer conjointement les paramètres et la structure de la mire à partir de valeurs initiales assurant la convergence.



Figure II.15 – A gauche, mire 3D utilisée dans les travaux de thèse de Bertrand Vandeportaele (2006) ; acquisition par un capteur à large champ. A droite, une autre mire 3D utilisée lors d'une acquisition avec du matériel David Laser Scanner.

Une seconde catégorie de méthodes propose un étalonnage utilisant des mires planes [Tsa86, SM99, Zha00], voir Figure II.16. Beaucoup plus faciles à réaliser, elles représentent aujourd'hui un standard en vision par ordinateur. Plusieurs outils sont disponibles sur Internet, comme par exemple [Bou10]. Ils permettent de réaliser une *calibration* à partir d'une série d'images prises sous des points de vue différents. L'étalonnage multi-images conduit à des résultats plus stables.



Figure II.16 – Mire plane de type damier et acquisition des points d'intersection via *OpenCV*.

C'est encore aujourd'hui un sujet d'étude actif qui voit régulièrement de nouvelles approches de construction de mires. Quelques chercheurs se sont attachés à simplifier encore les mires jusqu'à des objets linéaires [Zha04, HSH05], voir Figure II.17. D'autres se sont intéressés aux problèmes liés au fait de ne parfois pas pouvoir voir la mire dans sa globalité [OFN12], voir Figure II.18.



Figure II.17 – Mire mono-dimensionnelle utilisée dans les travaux de thèse de Benjamin Albouy (2006) en complément d'une *autocalibration*.

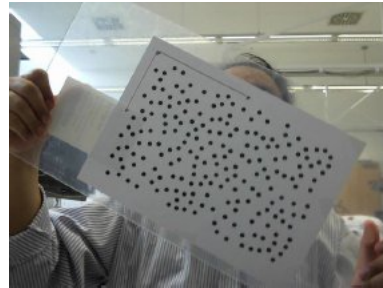


Figure II.18 – Mire plane aléatoire pour résoudre les problèmes d'occultation. [OFN12]

Auto-calibration

Utiliser une mire peut vite s'avérer contraignant hors d'un environnement de laboratoire. Une solution à ce problème a été initialement apportée pour des paramètres fixes [FJM92, Har92b]. Ils proposent une *autocalibration* du capteur d'image qui repose sur les équations de Kruppa [Kru13] déjà connues en photogrammétrie. L'étalonnage s'effectue directement à partir d'un ensemble de correspondances entre les images sans requérir la présence d'une mire. Ces résultats ont ensuite été étendus aux cas d'un ou plusieurs paramètres variables, notamment à celui de la focale [PKVG99].

Recherche des paramètres extrinsèques

Afin de déterminer la position relative des capteurs les uns par rapports aux autres, il est nécessaire d'avoir des points de correspondances dans chaque couple de vue. Lors d'une *calibration* à l'aide d'une mire, ces points de correspondance sont triviaux (coins sur un damier, etc.). La solution au problème devient moins évidente lorsque l'on utilise de l'*autocalibration*. Il faut alors pouvoir déterminer des couples de points se correspondant.

Classiquement, huit points de correspondances [LH87, Har97] sont utilisés afin de déterminer la translation et la rotation entre deux appareils. Le nombre de couples réellement utilisé est cependant généralement bien plus grand, afin de pouvoir établir une estimation précise, à l'aide d'un algorithme RANSAC¹¹ [FB81] ou BLUE¹² [Hen75] — généralement avec l'ajout de petites

11. *RANdom SAmple Consensus*

12. *Best Linear Unbiased Estimator*

hypothèses réduisant le cas aux moindres carrés classique. De nombreux travaux ont eu lieu afin de réduire le nombre minimal de points nécessaires. On trouve aujourd'hui des algorithmes performants fonctionnant à l'aide de cinq points [LH06] — le cas minimal considérant la focale comme une constante — et d'autre fonctionnant avec six points [SSNK05] — estimant au passage la focale. Ces algorithmes permettent une plus grande précision avec le même nombre total de couples car il y a plus de possibilités dans les algorithmes d'estimation.

Choix effectué pour la *calibration*

Dans le cadre de la cabine, possédant un grand nombre d'appareils d'acquisition, la *calibration* peut vite être fastidieuse. L'*autocalibration* est toutefois impossible. En effet, la cabine étant construite avec de nombreuses symétries dans son agencement, il n'est pas possible de retrouver avec certitude les positions et orientations relatives des appareils photographiques sans avoir au minimum une scène à l'intérieur.

De plus, Sturm a montré que lorsque la focale est fixe entre les deux prises de vue, alors pour un mouvement orbital il est impossible d'employer l'*autocalibration* [Stu02]. La configuration orbitale est respectée lorsque les axes optiques se coupent et que les positions des prises de vue sont équidistantes de cette intersection. Or la disposition des appareils photographiques dans la cabine d'acquisition correspond justement à un mouvement orbital avec une focale fixe. Il est donc logique que nous nous appuyés sur une mire de référence et cela pas uniquement par soucis de simplicité.

Nous avons choisi d'utiliser une mire plane de type damier. Son utilisation dans un ensemble de vues permet de déterminer la position relatives des appareils les uns par rapport aux autres ainsi que d'estimer ses paramètres internes. Afin d'obtenir les paramètres, la “*Camera Calibration Toolbox*” [Bou10] a été utilisée.

Pour résumer son fonctionnement, l'acquisition des paramètres se fait en plusieurs parties. Premièrement, une étape de détection des coins de la mire dans les images à lieu. Cette étape extrait chaque intersection dans le quadrillage du damier et demande comme pré-requis que la mire soit totalement visible dans les deux vues. Deuxièmement, les couples sont formés entre chaque vue en s'appuyant sur le motif du damier. Troisièmement, un algorithme BLUE est utilisé pour optimiser le calcul des paramètres extrinsèque et intrinsèque du système. Ces paramètres sont encodés sous formes de matrices et sont détaillés dans le chapitre suivant (section 1).

4 Conclusion

Dans le cadre du projet BRAMMS, une cabine d'acquisition de surface tridimensionnelle devait être élaborée. Les critères de sa conception étaient basés sur une phase d'acquisition rapide et une bonne précision de la reconstruction en termes de positionnement spatial des points reconstruit. L'acquisition se devait aussi de s'effectuer simultanément sur chaque appareil de capture. La cabine d'acquisition devait pouvoir supporter d'autres projets à l'avenir et donc rester modulable.

À partir de l'état de l'art des techniques actuellement utilisées, exposé dans le premier chapitre, et en respectant un certain budget, la conception et la réalisation de cette cabine a pu être effectuée de façon optimum. La technologie utilisée repose sur des appareils photographiques disposés en cercle autour de l'objet, ainsi que sur des vidéoprojecteurs.

Son interface permet la gestion des paramètres de chaque appareil photographique. Elle permet aussi d'en utiliser un pour voir le contenu de la cabine en déviant le flux EVF, afin de ne pas solliciter inutilement le système de déclenchement.

Celui-ci ne gère actuellement que le déclenchement simultané de tous les appareils de la cabine d'acquisition. Il est cependant très facilement modifiable grâce à la technologie Arduino utilisée. Les supports modulaires des appareils photographiques permettent aussi de faire varier la disposition de la cabine pour l'adapter facilement à d'autre projet.

La récupération des acquisitions photographiques a été automatisée. La partie concernant l'exploitation de celles-ci est traitée dans les prochains chapitres.

Des premiers tests ont eu lieu en mai 2012 à l'occasion d'une campagne de mesure réalisée par l'IFTH dans le cadre de la seconde thèse du projet Bramms, menée en parallèle de celle-ci. La cabine est aujourd'hui entièrement opérationnelle.

Chapitre III

Corrélation point à point

A la base des méthodes basées sur la stéréovision, il y a la mise en correspondance de chaque pixel des deux images. Cette approche est similaire au processus physico-cognitif qui permet à l'homme de discerner la profondeur.

Lorsque nous regardons un objet réel, l'image captée par l'œil droit est légèrement différente de celle perçue par l'œil gauche. Ceci est dû au décalage entre nos yeux. Ce décalage nous permet de percevoir le relief, même si nous n'avons aucun à priori sur la scène (Figure III.1). Cette particularité physique n'est bien sûr pas la seule technique mise en œuvre par le cerveau. On peut par exemple percevoir la profondeur sur une photographie — plane — car notre cerveau fait appel à d'autres processus cognitifs utilisant la grande base de connaissance qu'il possède. Des aprioris sur la taille des objets, sur les formes et sur les nuances d'éclairage sont entre autres utilisés pour affiner ce processus de détermination du relief et de la distance des objets nous entourant.

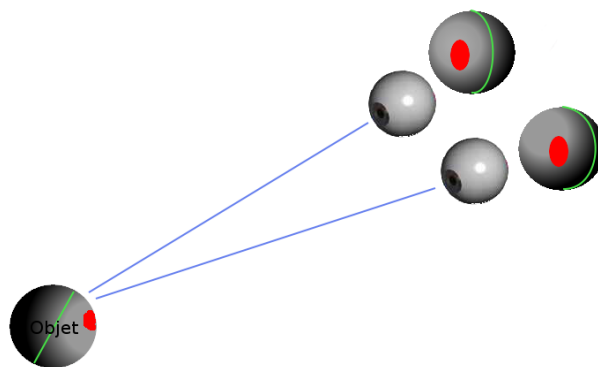


Figure III.1 — Stéréovision binoculaire. Ce que voit chaque œil est représenté derrière celui-ci.

Dans le cas de méthodes basées sur l'utilisation de lumière structurée, le schéma est semblable à la stéréoscopie (Figure III.1) mais l'un des deux yeux est remplacé par un projecteur. Il est cependant bien moins évident d'obtenir les paramètres de projection intrinsèques et extrinsèques

d'un tel appareil. Or puisque dans la création de la cabine nous utilisons plusieurs appareils de captures, le projecteur apportera de l'information sans nécessiter de d'étalonnage. Seule le cas de la stéréovision sera donc considéré dans le cadre de cette thèse. Ce chapitre et le suivant sont donc entièrement consacrés à la mise en correspondance des données acquises. C'est en effet la base nécessaire permettant de restituer la profondeur. Le dernier chapitre abordera lui le processus de triangulation spatiale à partir des correspondances déterminées.

Ce chapitre présente les bases théoriques de la stéréovision par ordinateur. Dans un premier temps, le modèle géométrique de formation des images sera exposé. Puis les relations géométriques entre deux images seront détaillées. Enfin, la suite du chapitre sera dédiée à la mise en correspondance de points entre deux images.

1 Modèle géométrique de formation des images

Un appareil photographique réalise la projection de la scène 3D sur le plan du capteur d'image (Figure III.2). L'objectif de l'appareil photographique focalise les rayons lumineux sur le capteur — une matrice CCD (ou CMOS) pour les appareils numériques, ou un film d'émulsion photographique pour les argentiques. Pour simplifier, un objectif peut être vu comme une lentille convergente, définie par sa distance focale f . Les rayons lumineux parallèles convergent tous en un point situé sur le plan focal, situé à une distance f de la lentille et perpendiculaire à l'axe optique (Figure III.3). Les objets situés à l'infini se forment sur le plan focal. Les autres ont une image qui se forme en arrière de celui-ci. Le fait de ramener l'image de l'objet dans le plan focal s'appelle “la mise au point” ou encore le “focus”. Avec une seule lentille, il faudrait pouvoir changer sa courbure ce qui n'est pas possible. L'objectif d'un appareil photographique est donc composé de plusieurs lentilles afin de pouvoir déplacer le plan focal, en les éloignant ou les rapprochant.

Une autre caractéristique des lentilles et des objectifs est que l'image formée dans le plan focal est inversée. En effet, la lentille fait passer tous les rayons lumineux par son centre optique, avant que l'image ne puisse se former dans son plan focal. Historiquement, on retrouve ce phénomène avec les premiers appareils de captures, qui étaient à l'époque dépourvus de lentille ou d'objectif. Ils étaient constitués d'une chambre noire percée d'une minuscule trou, comparable à une tête d'épingle (Figure III.4). Ce trou, appelé sténopé, a donné son nom au modèle sténopé ou encore “pinhole”. En vision par ordinateur, ce modèle très simple est suffisant pour caractériser la formation des images. En pratique, la précision des images y reste limitée par le phénomène de diffraction.

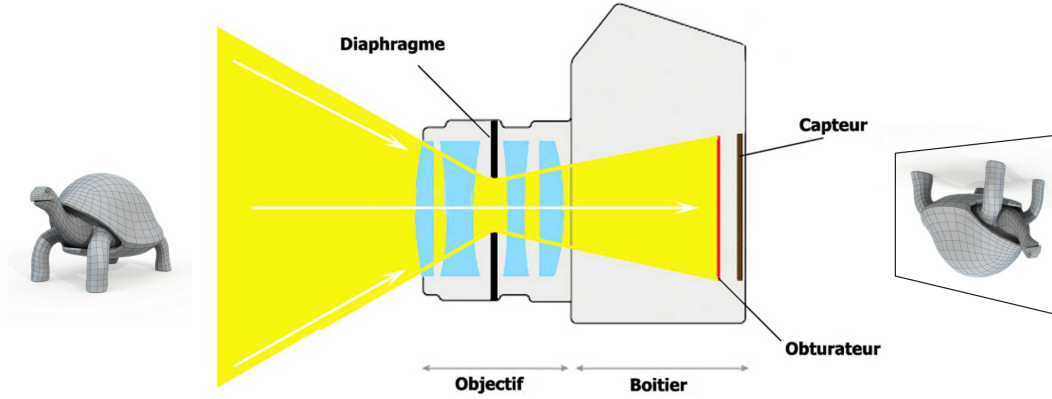


Figure III.2 – Formation d’une image au sein d’un appareil photographique. L’acquisition obtenue par le capteur CCD (ou CMOS) a subi une symétrie centrale.

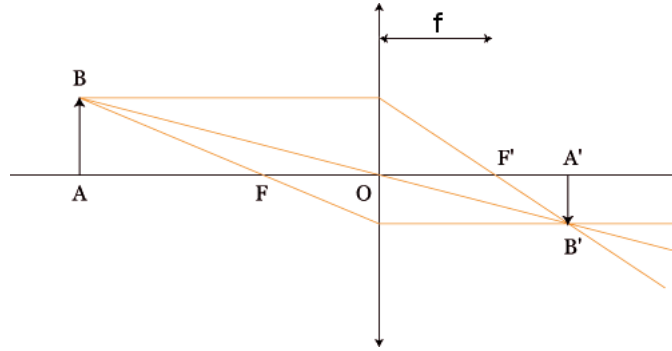


Figure III.3 – Formation d’image au travers d’une lentille mince.

Il s’agit maintenant de trouver les relations mathématiques entre un point de l’espace (sur la scène) et son image sur le capteur.

On désigne par $(O_c, \vec{x}, \vec{y}, \vec{z})$ le repère lié au capteur d’image. L’origine du repère est placée sur le centre optique O_c , l’axe O_z le long de l’axe optique, les z positifs allant vers l’objet. Les axes O_x et O_y correspondent respectivement aux directions horizontale et verticale de l’image et forment un trièdre direct avec l’axe O_z . L’image se forme donc sur le plan perpendiculaire à l’axe optique définie par $z = -f$. Le dispositif est alors celui représenté par la Figure III.5.

L’image produite à cet endroit est inversée car située derrière le centre optique. La même image, non-inversée, est située sur le plan parallèle défini par $z = f$. On désigne par (h, \vec{x}, \vec{y}) le repère utilisé dans le plan image. Ce plan peut être vu comme une restriction de l’espace de projection où vient se former l’image non inversée. h est alors le centre de l’image.

Afin de pouvoir écrire les relations sous formes matricielles, il est nécessaire de faire appel

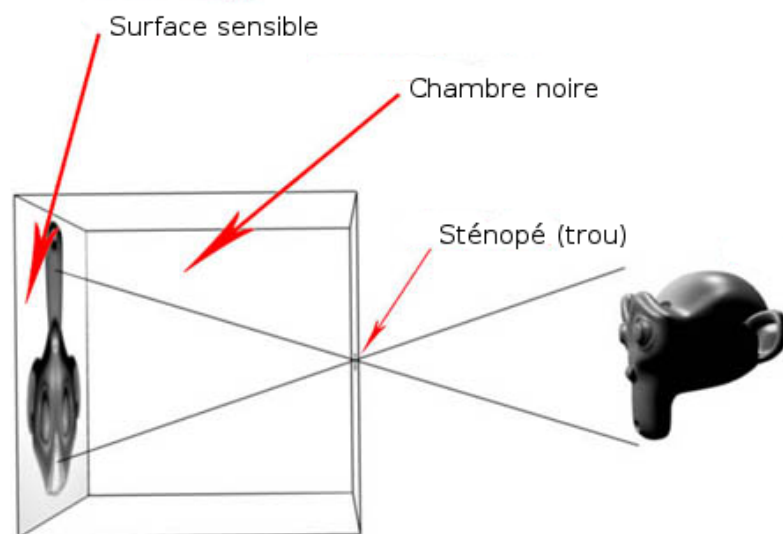


Figure III.4 – Sténopé

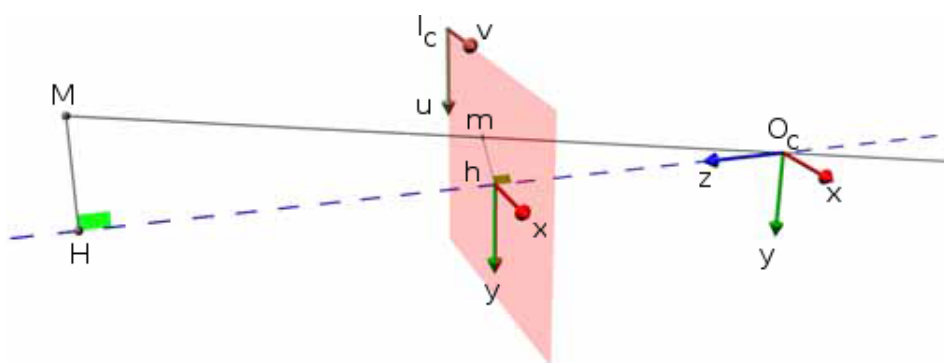


Figure III.5 – Repère sténopé

aux coordonnées homogènes. En coordonnées homogènes, un vecteur de dimension N s'écrit sur $N + 1$ composantes. En notant, \hat{v} la notation en coordonnées homogènes de v , on peut définir les coordonnées homogènes par l'équation III.1. Le passage de l'un à l'autre se faisant en choisissant un a arbitraire (généralement 1) ou en divisant par a . C'est pourquoi par la suite, le signe \propto signifiera « égal à un facteur près non nul ».

$$v = (x, y, z) \iff (\hat{v} = (ax, ay, az, a), \forall a \neq 0) \quad (\text{III.1})$$

Le point h d'intersection entre le plan image et l'axe optique est appelé point principal. Définissons enfin le point de la scène M de coordonnées $(x, y, z)^T$, son image m et sa projection

orthogonale H sur l'axe optique. $O_c h m$ et $O_c H M$ étant semblable, les coordonnées du point m sont :

$$\hat{m}_{(O_c, \vec{x}, \vec{y}, \vec{z})} \propto \begin{pmatrix} f \frac{x}{z} \\ f \frac{y}{z} \\ f \\ 1 \end{pmatrix} \quad (\text{III.2})$$

Les coordonnées de m dans le plan image muni du repère (h, \vec{x}, \vec{y}) sont directement déduites de l'équation III.2. La projection selon $O_{\bar{z}}$ retire la troisième composante, cependant ce repère n'est pas idéal. En effet, la notation usuelle en coordonnée image stipule que les coordonnées de m devraient être exprimées en pixels en prenant comme repère le coin supérieur gauche I_c dans le repère image (I_c, \vec{u}, \vec{v}) . On a dans cette convention \vec{v} qui est dirigé vers le bas de l'image. \vec{u} a donc la même direction et le même sens que x mais a pour longueur celle d'un pixel. \vec{v} quant à lui à la même direction que y mais est de sens opposé et à pour longueur celle d'un pixel. Pour obtenir la bonne représentation, il suffit de réaliser un changement de repère :

$$\hat{m}_{(I_c, \vec{u}, \vec{v})} \propto \begin{pmatrix} f x k_u + u_c \\ f y k_v + v_c \\ z \end{pmatrix} \quad (\text{III.3})$$

où (u_c, v_c) sont les coordonnées en pixels du point h dans le repère (I_c, \vec{u}, \vec{v}) et k_u et k_v sont respectivement les facteurs de conversion horizontal et vertical. Ils permettent par exemple de passer des coordonnées exprimées en millimètres à des coordonnées en pixels. Dans ce cas, l'unité de ces valeurs serait le `pixel.mm-1`.

L'équation III.3 peut aussi s'écrire en notation matricielle :

$$\hat{m}_{(I_c, \vec{u}, \vec{v})} \propto \begin{pmatrix} f k_u & 0 & u_c & 0 \\ 0 & f k_v & v_c & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \hat{M}_{(O_c, \vec{x}, \vec{y}, \vec{z})} \quad (\text{III.4})$$

La matrice des paramètres intrinsèques du capteur d'image est la sous-matrice constituée des 3 premières colonnes de la matrice de l'équation III.4. Dans la littérature, elle est parfois rencontrée avec un paramètre supplémentaire qui est l'angle entre \vec{u} et \vec{v} . En effet, même si les fabricants de capteurs essaient d'avoir un angle le plus proche possible de 90° , ce n'est pas toujours garanti. En notant ϕ l'angle entre \vec{u} et \vec{v} , la matrice des paramètres intrinsèques est

Corrélation point à point

donc définie ainsi :

$$K_{int} = \begin{pmatrix} fk_u & -fk_u \cot(\phi) & u_c \\ 0 & \frac{fk_v}{\cot(\phi)} & v_c \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{III.5})$$

Si nous ne sommes pas dans le repère $(O_c, \vec{x}, \vec{y}, \vec{z})$, l'équation III.4 n'est plus valide. Une isométrie doit être faite pour changer de repère, c'est-à-dire être la combinaison d'une translation et d'une rotation. La première étape consiste à effectuer une rotation R de la scène pour que les directions des axes coïncides, puis on translate par t la scène de sorte que le capteur d'image C vienne sur l'origine O_c . Mathématiquement, et toujours en coordonnées homogènes, cette transformation s'écrit :

$$\left(\begin{array}{c|c} Id_{3 \times 3} & t \\ \hline 0 & 1 \end{array} \right) \left(\begin{array}{c|c} R & 0 \\ \hline 0 & 1 \end{array} \right) = \left(\begin{array}{c|c} R & t \\ \hline 0 & 1 \end{array} \right) \quad (\text{III.6})$$

En prenant en compte le changement de repère de l'équation III.6 et la projection effectuée par l'équation III.4, l'équation générale du calcul des coordonnées du point m calculées à partir de celles M peut être définie. En remarquant la colonne de 0 à droite de la matrice dans l'équation III.4, nous pouvons l'écrire en supprimant la dernière ligne de l'équation III.6 :

$$\hat{m}_{(I_c, \vec{u}, \vec{v})} \propto K_{int}(R|t) \hat{M}_{\text{repère monde}} \quad (\text{III.7})$$

Ce qui définit la matrice de projection P du capteur d'image :

$$P \propto K_{int}(R|t) \quad (\text{III.8})$$

La matrice K_{int} est constituée des paramètres intrinsèques. Bien qu'ils soient au nombre de 6, leur écriture dans K_{int} nous laisse 5 variables à connaître, correspondant aux 5 éléments significatifs constituant cette matrice. $(R|t)$ est construite à partir des paramètres extrinsèques du capteur d'image. Bien que la matrice des paramètres extrinsèques soit de taille 3×4 , les paramètres extrinsèques sont au nombre de 6 car la rotation est complètement déterminée par seulement 3 paramètres (les angles d'Euler). Le modèle sténopé est donc entièrement caractérisé par 11 paramètres (5 intrinsèques et 6 extrinsèques). La détermination de ceux-ci se fait lors du processus de *calibration* vue au chapitre précédent.

En conséquence, les projections de la scène dans différents plans nous permettent d'obtenir

La géométrie épipolaire peut aider à la mise en correspondance lorsqu'elle est connue. En effet, un point m_g dans I_g n'a pas de profondeur. Il peut être la projection d'un point M n'importe où sur la droite $(C_g m_g)$. Quel que soit la position de M sur cette droite, le plan épipolaire Π généré est le même. Donc le correspondant de m_g sur I_d se trouve sur l'intersection entre Π et I_d , c'est à dire l_d .

Cette relation entre point et droite peut aussi être définie analytiquement.

2.2 Relations analytiques

Il existe différentes méthodes pour déterminer la relation analytique entre un point m_g et sa droite épipolaire associée l_d [HZ04]. Nous détaillerons ici les travaux de Xu et Zhang paru en 1996 [XZ96].

Soit P_g et P_d les matrices de projection associées aux capteurs d'image. La résolution algébrique de l'équation $P_g \hat{M} \propto \hat{m}_g$ permet d'obtenir l'ensemble des points formant la droite $(C_g m_g)$:

$$\hat{M}(\lambda) \propto P_g^+ \hat{m}_g + \lambda \hat{C}_g \quad (\text{III.9})$$

où $P^+ = P^T(PP^T)^{-1}$ est la matrice pseudo inverse de P et λ un scalaire.

Deux points arbitrairement choisis de cette droite et projetés dans le plan de I_d permettent de déterminer l_d . En prenant respectivement les valeurs pour $\lambda = 0$ et pour $\lambda = \infty$, on retrouve les points particuliers $P_g^+ m_g$ et C_g . Leurs images sont donc respectivement $P_d P_g^+ m_g$ et $P_d \hat{C}_g \propto \hat{e}_g$. La droite épipolaire l_d passe par ces deux points et est définie par le produit vectoriel $l_d \propto \hat{e}_d \wedge (P_d P_g^+ \hat{m}_g)$. Ce produit vectoriel peut s'écrire de façon matricielle :

$$l_d \propto [e_d]_{\wedge} P_d P_g^+ \hat{m}_g \propto F \hat{m}_g \quad (\text{III.10})$$

où la notation $[x]_{\wedge}$ est l'expression matricielle du produit scalaire :

$$\begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}$$

Cette équation permet de définir la matrice $F = [e_d]_{\wedge} P_d P_g^+$, appelée matrice fondamentale. Elle représente une application d'un espace projectif à 2 dimensions dans un espace à 1 dimension. Son rang est donc de 2. En notant que le produit $m_d^T l_d$ est nul — le point m_d appartient à la droite l_d — on obtient la relation suivante entre deux points mis en correspondance :

$$\hat{m}_d^T F \hat{m}_g = 0 \quad (\text{III.11})$$

Dans ce raisonnement nous avons choisi de partir du capteur de gauche vers celui de droite. La géométrie étant symétrique, le même raisonnement peut s'appliquer réciproquement de droite à gauche. Dans ce cas, la matrice fondamentale obtenue aurait été la transposée de celle de l'équation III.11. Ainsi, de la même façon que $l_d \propto F\hat{m}_g$, on a $l_g \propto F^T\hat{m}_d$.

De par sa définition, la matrice fondamentale intègre les paramètres intrinsèques et extrinsèques des deux capteurs d'image qu'elle lie. Il est possible de l'estimer directement via l'équation III.11 et un ensemble de point de correspondance. C'est généralement de cette manière que se fait la *calibration* (voir chapitre II section 3 pour les méthodes de calibration).

3 Méthodes courantes d'appariement de point d'intérêt

La figure III.6 montre qu'un point M de la scène, visible dans les deux images se projette dans I_g en m_g et dans I_d en m_d . A la différence des pixels qui évoluent dans un domaine discret, ces points évoluent sur un domaine continu. On verra par la suite que cette étape de discrétisation engendre nombre de problèmes. On dit que les points m_g et m_d sont homologues, c'est-à-dire qu'ils se correspondent, car ils représentent le même point de la scène. Pour la suite, le symbole " \doteq " sera utilisé pour symboliser cette relation entre points homologues, tandis que le symbole " \neq " symbolisera naturellement leur non-correspondance. En somme, la mise en correspondance consiste à retrouver les couples de points homologues.

Les méthodes de mise en correspondance de pixel se basent sur l'hypothèse raisonnable suivante :

H1 : *Si deux pixels se correspondent alors leurs niveaux de gris se ressemblent.*

Cependant, cette hypothèse n'étant pas assez discriminante. Elle est donc très souvent complétée de la manière suivante :

H2 : *Si deux pixels se correspondent alors les niveaux de gris de leurs voisinages respectifs se ressemblent.*

Des pixels se ressemblant du point de vue de leur voisinage seront qualifiés de similaire. La non ressemblance dans leur voisinage entrainera ce que l'on appelle la dissimilarité.

Ces hypothèses, bien que simples, ont démontrées à de nombreuses reprises leur efficacité. Généralement, le but est d'obtenir une carte de profondeur (dense ou éparse) qui sert de support à d'autres algorithmes. En plus des applications mentionnées lors du chapitre I, nous pouvons mentionner l'étude du comportement mécanique des matériaux [OCG⁺07], le suivi de personnes (vidéo-surveillance, prévention de chute) [RAR⁺11], la préservation du patrimoine (numérisation) [HKLP09] et la reconstruction de scène urbaine [WCRZ11].

Les méthodes de mise en correspondance sont généralement décrites sous la forme d'un problème énergétique à minimiser. Ainsi formulé, le problème prend en compte des coûts locaux liés à la fonction utilisée et des coûts par zone ce qui le rend assez complexe. En effet, il existe de nombreuses méthodes de résolution différentes et il est facile d'en trouver une qui ne collera pas au problème ainsi formulé.

Scharstein et Szeliski [SS02] distinguent dans leur état de l'art quatre éléments caractérisant les méthodes de mise en correspondance stéréoscopique de pixels :

- le coût local : coût d'une correspondance,
- la zone d'agrégation : l'ensemble des pixels intervenant dans le calcul du coût local,
- la méthode d'optimisation,
- l'affinement des résultats.

Brown et al. [BBH03] différencient dans leur état de l'art les méthodes globales des méthodes locales. Les méthodes globales essayent de procéder à une minimisation de l'énergie en tenant compte de tous les couples à la fois, au contraire des méthodes locales qui se concentrent sur un couple à la fois.

3.1 Notions générales

La mise en correspondance stéréoscopique de pixels s'appuie sur différents éléments. L'évaluation de chacun de ces points permet donc de définir la démarche d'appariement la plus adéquate à chaque situation.

▷ **Les primitives à appairer et leurs attributs :**

- *Les pixels de l'image considérée* : Est-ce que l'on cherche à appairer tous les pixels de l'image, ou seulement une partie — et comment sont alors déterminés ces points d'intérêts ? Encodent-ils une information vectorielle — comme c'est le cas des images en couleur ou lorsqu'il y a des canaux d'information supplémentaire (canal alpha pour la transparence par exemple) — ou une information scalaire — généralement les images en niveaux de gris ?
- *Les primitives structurées* : Cherche-t-on à appairer entre-eux des segments, des contours ou des régions dans l'image ?

- ▷ **Le pré-traitement des images** : Est-il nécessaire d'effectuer un filtrage afin de faire ressortir les informations pertinentes de l'image avant l'appariement ? Si oui, de quel type ? On peut par exemple essayer d'éliminer d'éventuels bruits [KLCL05]. On peut aussi vouloir détecter des pixels remarquables, voir section 3.4.2. On peut encore vouloir effectuer une segmentation des images qui sera utilisée lors de la mise en correspondance.

- ▷ **Le coût de mise en correspondance et la zone d'agrégation :**
 - *Le coût local* : Utilise-t-on à chaque appariement une unique primitive ou alors un ensemble de primitives contenues dans un voisinage ? Autrement dit, quelle est la taille de la zone d'agrégation ?
 - *Le coût de voisinage* : Quel score aurait-on en appariant une primitive avec, non pas elle-même, mais sa voisine ? Est-ce différent de façon notable ? Dès lors, comment différencier le mauvais appariement du bon ? Généralement, cette problématique entraîne une réflexion sur une méthode d'optimisation globale.
- ▷ **La zone de recherche, la zone d'agrégation et le support :** Pour le calcul d'une correspondance, il faut généralement plus que la simple primitive lors d'un calcul de coût pour déterminer l'appariement ; on utilise classiquement aussi son voisinage. Cette zone s'appelle la zone d'agrégation. Le support quant à lui est l'ensemble des primitives candidate pour un appariement dans la zone de recherche. Dans le cas de l'appariement de primitives de type pixel, le support est donc l'ensemble des pixels de la zones de recherche plus les pixels sur le bord à l'extérieur de celle-ci qui sont nécessaire car inclus dans certaine zone d'agrégation, voir Figure III.7. A-t-on besoin de chercher un correspondant dans toute l'image ? Comment mieux adapter la zone de recherche pour éviter les calculs inutiles ? Quelle taille et quelle forme choisir pour la zone d'agrégation ?

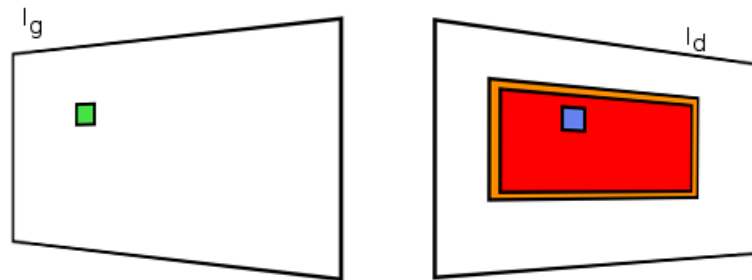


Figure III.7 – Zones de recherche (en rouge), d'agrégation (agrégation locale en vert, agrégation liée à la primitive testée en bleue) et support (en orange qui correspond à l'ensemble des agrégations liées aux primitives candidates).

- ▷ **La méthode d'optimisation :** On garde souvent le coût minimal pour une méthode locale. Mais comment choisir la méthode permettant l'obtention d'une optimisation globale ?
- ▷ **Les passages multiples :** Cherche-t-on à combiner plusieurs méthodes différentes (une à chaque passe) ? Essaye-t-on d'optimiser avec une approche pyramidale multi-résolution — en partant de la résolution la plus grossière vers la résolution la plus fine ?
- ▷ **L'affinement des résultats :** Peut-on corriger des erreurs d'appariement *a posteriori* ?

Corrélation point à point

Lors d'une mise en correspondance, chacun de ces éléments intervient dans une étape de l'Algorithme 1. Toutes les étapes ne sont pas significatives dans chaque algorithme. Dans ces cas-là, soit le calcul n'est pas nécessaire, soit la boucle correspondante est limitée à un unique passage.

Algorithme 1 Algorithme général de mise en correspondance

Donnée: Les deux images acquises par les deux capteurs

Résultat: Ensemble de couples de primitives appariées

Pour chaque passage **faire**

 Construire la pyramide de résolution de l'image {Uniquement si multi-résolution}

Pour chaque niveau de la pyramide **faire**

 Pré-traiter les images {Uniquement si nécessaire}

 Partitionner l'ensemble des primitives dont on cherche les correspondants en support

Pour chaque support **faire**

Pour chaque primitive du support **faire**

 Calculer les coût locaux en utilisant la zone d'agrégation du coût local

 Calculer les coûts de voisinage en utilisant la zone d'agrégation

Fin pour

 Calculer le coût global

 Déterminer les correspondances sur le support selon la méthode d'optimisation

Fin pour

 Affiner les résultats {Uniquement si traitement à posteriori}

 Propager les résultats {Uniquement si multi-résolution}

Fin pour

Fin pour

Retourner Ensemble des couples de primitives appariées

Ces algorithmes essayent de déterminer les associations appariant correctement le plus de couples, afin de répondre au problème de la corrélation. Ce vaste problème peut être reformulé sous forme énergétique. En effet, le but étant d'obtenir le plus grand nombre possible de couples appariés, le problème peut se résumer à minimiser l'énergie dépendant de la fonction de disparité

d :

$$E_{\text{global}}(d) = \sum_{P_g^{i,j} \in \mathbb{S}} \left((1 - \lambda) \overbrace{\sum_{\substack{P_g^{i',j'} \in \mathbb{A}(P_g^{i,j}) \\ P_d^{u',v'} \in \mathbb{A}(P_g^{i,j} + d(P_g^{i,j}))}} E_{\text{dissimilarité}}(P_g^{i',j'}, P_d^{u',v'})}^{E_{\text{local}} \text{ (ie : coût local)}} + \lambda \sum_{c=0}^{N_c-1} \underbrace{\sum_{P_g^{i',j'} \in \mathbb{AC}^c(P_g^{i,j})} E_{\text{lissage}}^c(P_g^{i,j}, P_g^{i',j'})}_{E_{\text{voisinage}}^c \text{ (ie : coût du voisinage pour la contrainte c)}} \right) \quad (\text{III.12})$$

Dans cette équation, $E_{\text{dissimilarité}}$ évalue à quel point deux pixels ne se ressemblent pas. E_{lissage}^c permet de comparer les disparités associées aux deux primitives considérées et est calculé sur le domaine $\mathbb{AC}^c(P)$ qui est la zone d'agrégation associée à la contrainte c pour P , le nombre total de contrainte étant noté N_c . La plupart du temps, $\mathbb{AC}^c(P)$ correspond à un voisinage et n'est pas forcément identique à $\mathbb{A}(P)$ qui est la zone d'agrégation locale. E_{lissage}^c est généralement calculé en utilisant la disparité associée à chaque pixel en tenant compte des occultations. \mathbb{S} est le support.

Dans le cas d'une méthode purement locale, l'équation III.12 se simplifie. Les supports \mathbb{S} sont tous réduits à des singletons. La zone d'agrégation $\mathbb{A}(P)$ peut être une fenêtre carrée centrée en P , ou plus rarement une fenêtre adaptative (c'est à dire de dimension variable). Les méthodes locales n'ont pas de coût de voisinage par définition. La résolution de l'énergie nous mène donc à :

$$\forall P_g^{i,j} \in \mathbb{Z}_{\text{considérée}}, \quad P_g^{i,j} \doteq P_d^{u,v} = \underset{P_d^{u',v'} \in \mathbb{Z}_{\text{recherche}}(P_g^{i,j})}{\text{argmin}} E_{\text{local}}(P_g^{i,j}, P_d^{u',v'}) \quad (\text{III.13})$$

Dans le cas d'une méthode purement globale, la zone d'agrégation \mathbb{A} est réduite à un pixel. L'équation III.12 devient donc :

$$E_{\text{global}}(d) = \sum_{P_g^{i,j} \in \mathbb{S}} \left((1 - \lambda) E_{\text{dissimilarité}}(P_g^{i,j}, P_d^{u,v}) + \lambda \sum_{c=0}^{N_c-1} \sum_{P_g^{i',j'} \in \mathbb{AC}^c(P_g^{i,j})} E_{\text{lissage}}^c(P_g^{i,j}, P_g^{i',j'}) \right) \quad (\text{III.14})$$

Les coûts de dissimilarité et les coûts de lissage sont définis par la mesure utilisée et par les contraintes utilisées. Dans un premier temps, la notion de disparité va être introduite. Puis

les différentes contraintes seront exposées. Finalement, nous exposerons différentes mesures de corrélation.

3.2 Disparité et fonction de disparité

Toutes les explications et les formules peuvent être utilisées symétriquement — de l'image de gauche vers l'image de droite ou de l'image de droite vers l'image de gauche. C'est pourquoi dans la suite, l'image de départ sera arbitrairement l'image de gauche I_g et l'image d'arrivée sera celle de droite I_d .

Disparité

Une manière de représenter le résultat d'une mise en correspondance consiste à associer à chaque pixel de I_g noté $P_g^{i,j} = (i, j)^T$ un vecteur appelé disparité. Il est défini par ses composantes $(u - i, v - j)^T$ où $(u, v)^T$ sont les coordonnées dans I_d du point correspondant au pixel $P_g^{i,j}$.

Fonction de disparité

Le problème de la mise en correspondance peut être assimilé à la recherche d'une fonction de disparité d qui attribue une disparité à chaque pixel $P_g^{i,j}$. Cela donne, dans le cas général :

$$\begin{aligned} d : \quad \mathbb{N}^2 &\longrightarrow \mathbb{R}^2 \\ P_g^{i,j} &\longmapsto d(P_g^{i,j}) = (u - i, v - j)^T \end{aligned} \tag{III.15}$$

L'espace d'arrivée est l'ensemble des couples de réel, car il est possible d'effectuer un appariement à une échelle plus basse que celle du pixel. Cependant la plupart des méthodes n'étant pas sous-pixels, les disparités auront des valeurs entières.

3.3 Contraintes

Une contrainte est une propriété découlant d'une hypothèse faite sur :

- ▷ **La géométrie du capteur** : Ce sont les contraintes géométriques induites par la géométrie du capteur. La principale contrainte de ce type est la contrainte épipolaire.
- ▷ **La géométrie de la scène** : Il y en a trois : l'unicité, l'ordre et la symétrie. Elle sont parfois appelée contraintes de compatibilité [Jon97]. Elles s'appliquent aux primitives.

- ▷ **La réflectance de la surface des objets constituant la scène :** Ceci est lié aux contraintes de dissimilarité. Cette contrainte constitue l'élément essentiel du coût de mise en correspondance.

Leur utilisation peut avoir plusieurs buts. Il est possible de les utiliser pour limiter le nombre de correspondants potentiels pour chaque primitives (en restreignant la zone de recherche par exemple). Elles permettent aussi de trancher lorsque plusieurs candidats sont potentiellement acceptables pour l'appariement. Enfin, il est possible de les utiliser pour vérifier la cohérence de l'ensemble des mises en correspondances (et supprimer les faux appariements).

Une contrainte est dite unaire si elle ne prend qu'une correspondance. Elle est dite binaire si elle utilise les informations de deux correspondances. La liste des contraintes ci-après n'est pas exhaustive, et seule les plus courantes sont ici exposées.

3.3.1 Contrainte épipolaire et rectification épipolaire

Contrainte épipolaire

Cette contrainte géométrique et unaire peut être utilisée pour faciliter la mise en correspondance en réduisant la zone de recherche. La géométrie épipolaire est expliquée plus tôt dans ce chapitre, voir section 2.

La contrainte épipolaire permet d'obtenir une relation entre les coordonnées de I_g et les coordonnées de I_d et donc des primitives contenues dans les deux images. Cette transformation est représentée par la matrice fondamentale, voir section 2.2. Ainsi l'espace de recherche de la primitive correspondante est réduit à une recherche monodimensionnelle, diminuant ainsi les temps de calcul et les risques d'erreur. Dans certains cas, on préférera utiliser une bande étroite centrée sur l'épipolaire correspondante afin de pallier les petits défauts de géométrie possible lors de la calibration du système.

Rectification épipolaire

La rectification épipolaire consiste à effectuer une transformation de l'image en altérant sa géométrie. Le but étant de se retrouver dans une configuration parallèle. Dans cette configuration, les droites épipolaires sont toutes parallèles et horizontales dans les deux images. Ainsi deux pixels qui se correspondent seront sur la même ligne (ie : $P_g^{i,j} \doteq P_d^{i,k}$, avec \doteq la relation symbolisant la correspondance). La disparité est donc ramenée à un vecteur horizontal ce qui

Corrélation point à point

simplifie grandement les calculs lors de l'implémentation :

$$\begin{aligned} d : \mathbb{N}^2 &\longrightarrow \mathbb{R}^2 \\ P_g^{i,j} &\longmapsto d(P_g^{i,j}) = (0, v - j)^T \end{aligned} \quad (\text{III.16})$$

Cette transformation s'effectue sur les deux images mais l'estimation du paramètre de cette transformation n'est pas aisée. Suivant l'information dont on dispose, on utilisera les paramètres de calibrage des deux images [Har99] pour estimer les matrices de projection perspective, ou alors on cherchera à estimer la matrice fondamentale à partir de points de correspondance lorsque les informations complètes de calibrage ne sont pas connues [Zha98].



Figure III.8 – Rectification épipolaire. Les images rectifiées sont en bas. On notera que les nouveaux coins des images résultats peuvent être hors champ ou à l'intérieur de l'image.

Les deux nouvelles images générées par cette transformation (voir Fig. III.8) sont calculées par interpolation. Le contenu de l'image subit donc une dégradation. Cette méthode est toutefois souvent employée pour un gain de temps évident lorsque la contrainte épipolaire est employée. En effet, il n'est plus nécessaire de calculer la droite épipolaire lors de la recherche du correspondant de chaque primitive, cette droite étant immédiatement connue (voir Equation III.16).

3.3.2 Contrainte d'unicité

Cette contrainte binaire largement employée en stéréovision est définie par :

$$\text{Si } P_g^{i,j} + d(P_g^{i,j}) \doteq P_d^{i,v} \text{ alors } \forall k \neq j, P_g^{i,k} + d(P_g^{i,k}) \not\doteq P_d^{i,v} \quad (\text{III.17})$$

Deux pixels différents ne peuvent avoir le même correspondant. Il faut ici faire attention car en fait, si cette contrainte semble cohérente, son emploi dans l'espace discret des pixels est généralement faux. En effet, lorsque le plan de la scène est plus incliné dans l'une des deux vues, les correspondances vont forcément être sous-pixel car l'espace d'arrivée est plus petit que l'espace de départ, voir Fig. III.9.

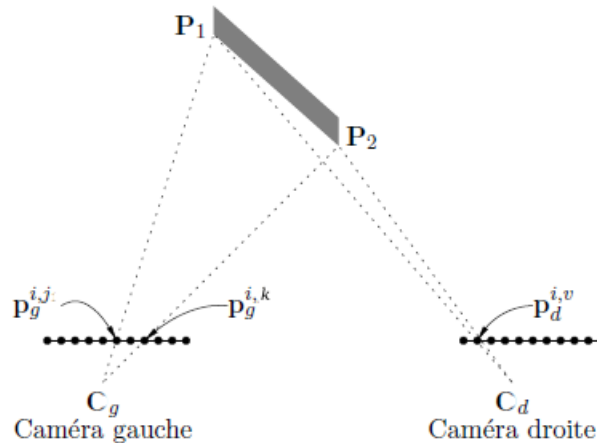


Figure III.9 – Contrainte d'unicité non respectée.

3.3.3 Contrainte d'ordre

Cette contrainte binaire est parfois utilisée en stéréovision en tant qu'outil de contrôle. Elle est définie ainsi :

$$\text{Si } P_g^{i,j} + d(P_g^{i,j}) \doteq P_d^{i,v} \text{ et } P_g^{i,k} + d(P_g^{i,k}) \doteq P_d^{i,w} \text{ alors } (j - k)(v - w) \geq 0 \quad (\text{III.18})$$

L'ordre des pixels de l'image de gauche le long de la droite épipolaire doit être le même que celui de leurs correspondants. Cette contrainte généralement vraie peut être violée en cas de scène composée par des objets disjoints, voir Fig. III.10.

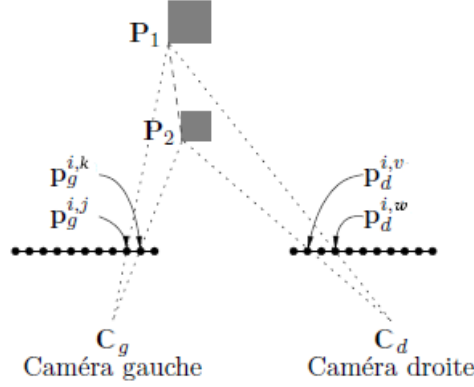


Figure III.10 – Contrainte d'ordre non respectée.

3.3.4 Contrainte de symétrie et contrainte de consistance forte

La contrainte de symétrie est une contrainte binaire. Elle est employée comme outil de contrôle *a posteriori*. Elle se définit ainsi :

$$\text{Si } P_g^{i,j} + d(P_g^{i,j}) \doteq P_d^{i,v} \quad \text{alors} \quad P_d^{i,v} + d(P_d^{i,v}) \doteq P_g^{i,j} \quad (\text{III.19})$$

Elle correspond au fait que le système est symétrique. Autrement dit, le correspondant d'un point de I_g dans I_d doit avoir pour correspondant dans I_g le point de départ, voir Fig. ???. Elle permet donc *a posteriori* de retirer les appariements qui ne respectent pas cette contrainte. L'emploi de cette contrainte nécessite d'effectuer une nouvelle fois l'appariement en échangeant l'image de départ et l'image d'arrivée, ce qui implique un temps de calcul deux fois plus long.

Cette contrainte est équivalente à la contrainte de consistance forte [KS03] qui se définit de la même façon que sur un ensemble d'appariement. Elle consiste à vérifier la contrainte d'unicité sur les primitives de la droite vers la gauche et *vice versa*. Cette seconde contrainte étant très forte, il est parfois nécessaire de la rendre plus tolérante, encore une fois à cause de la discrétisation de l'image. Cette contrainte de symétrie faible [CMP96] est définie ainsi :

$$\text{Si } P_g^{i,j} + d(P_g^{i,j}) \doteq P_d^{i,v} \quad \text{et} \quad P_d^{i,v} + d(P_d^{i,v}) \doteq P_g^{i,k} \quad \text{alors} \quad |k - j| < \text{Seuil} \quad (\text{III.20})$$

3.3.5 Contrainte de consistance faible

La contrainte de consistance faible est une variante de la contrainte de consistance forte. Elle s'appuie sur la contrainte d'unicité et la contrainte d'ordre. Elle est donc binaire et se

définie ainsi :

$$\begin{aligned} \text{Si } P_g^{i,j} + d(P_g^{i,j}) &\doteq P_d^{i,v} \text{ alors} \\ \forall k > j, P_g^{i,k} + d(P_g^{i,k}) &\not\doteq P_d^{i,v} \text{ et } \forall w < v, P_d^{i,w} + d(P_d^{i,w}) &\not\doteq P_g^{i,j} \end{aligned} \quad (\text{III.21})$$

Cela signifie que lorsqu'il y a une correspondance, la contrainte de symétrie doit être vérifiée pour toutes les primitives à droite du point de I_g et pour toutes les primitives à gauche du point de I_d correspondant. Cette contrainte a pour but de limiter l'effet des contraintes d'unicité et d'ordre qui rejettent beaucoup d'appariement, voir Fig. III.11.

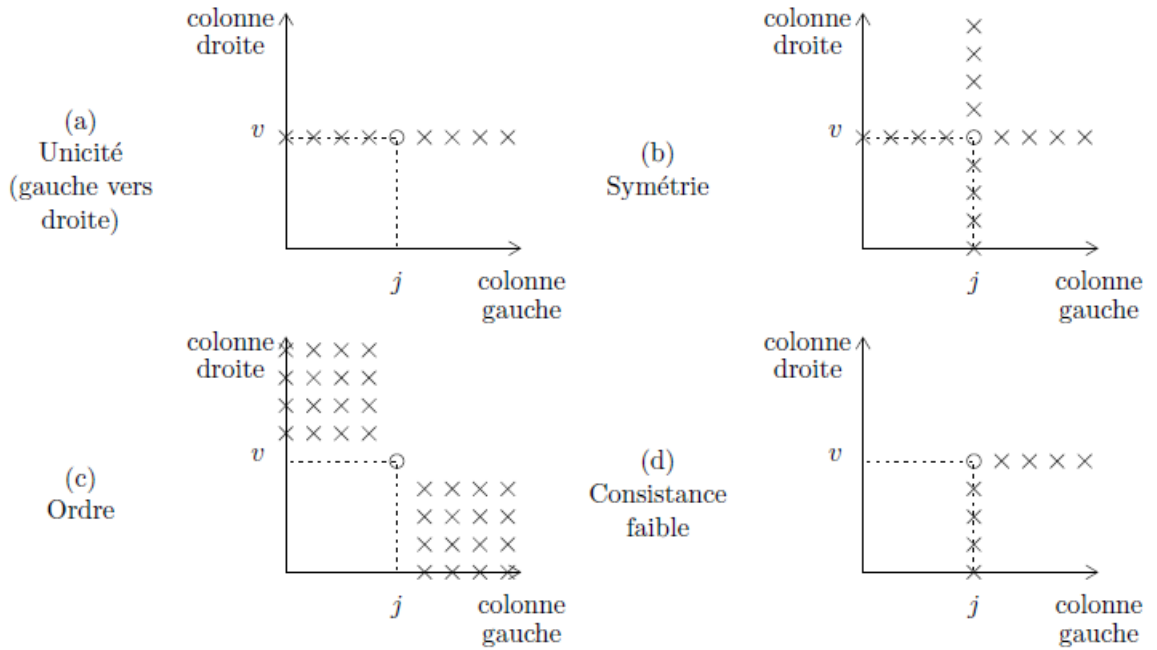


Figure III.11 – Synthèse de toutes les contraintes géométrique.

3.3.6 Contrainte sur le gradient de disparité

Cette contrainte binaire très utilisée s'appuie sur l'hypothèse que la plupart des objets d'une scène ont une surface majoritairement continue [HM95]. La continuité de la surface entraîne le fait que le gradient ne doit pas dépasser un seuil donné. Le gradient de disparité peut être approximé comme étant la différence des disparités, divisée par la séparation cyclopéenne (différence entre la distance entre deux pixels de I_g et la distance entre leurs correspondants

Corrélation point à point

respectifs dans I_d). Il est noté $\dot{\nabla}d$. Cette contrainte est donc définie par :

$$\begin{aligned} \text{Si } P_g^{i,j} + d(P_g^{i,j}) \doteq P_d^{i,v} \text{ et } P_g^{i,k} + d(P_g^{i,k}) \doteq P_d^{i,w} \text{ alors} \\ \dot{\nabla}d(P_g^{i,j}, P_d^{i,v}, P_g^{i,k}, P_d^{i,w}) = \frac{\|(P_g^{i,j} - P_g^{i,k}) - (P_d^{i,v} - P_d^{i,w})\|}{\|\frac{1}{2}((P_g^{i,j} - P_g^{i,k}) - (P_d^{i,v} - P_d^{i,w}))\|} < \text{Seuil} \end{aligned} \quad (\text{III.22})$$

3.3.7 Contrainte de rang

Cette contrainte unaire [BB01] s'appuie sur la comparaison de niveau de gris du pixel central avec celui des pixels de la zone d'agrégation. Elle peut se définir ainsi :

$$\begin{aligned} \text{Si } P_g^{i,j} + d(P_g^{i,j}) \doteq P_d^{i,v} \text{ alors} \\ \forall P_g^{i',j'} \in \mathbb{A}(P_g^{i,j}), \forall P_d^{i',v'} \in \mathbb{A}(P_d^{i,v}), j' - j = v' - v \Rightarrow (I_g^{i',j'} - I_g^{i,j})(I_d^{i',v'} - I_d^{i,v}) > 0 \end{aligned} \quad (\text{III.23})$$

L'ensemble $\mathbb{A}(P)$ correspond à celui des primitives dans la zone d'agrégation de la contrainte de la primitive P . Cette contrainte est déduite du fait que la géométrie de la surface ne dépend pas du point d'observation. Cette assertion implique que la radiance de la scène autour d'un point donné est semblable quel que soit son point d'observation. Cela se traduit par le fait que si deux pixels se correspondent, la variation des niveaux de gris dans leur voisinage doit être semblable, d'où l'expression mathématique donnée ci-dessus.

3.3.8 Contrainte de continuité figurale

Cette contrainte [HM95] est binaire. En notant respectivement C_g et C_d un contour dans I_g et I_d et “#” le cardinal d'un ensemble, on peut définir cette contrainte ainsi :

$$\begin{aligned} \text{Si } \# \{P_g^{i,j} \in C_g \mid P_g^{i,j} + d(P_g^{i,j}) \in C_d\} \geq \frac{1}{2} \# \{P_g^{i,j} \in C_g\} \text{ alors} \\ \forall (i', j') \in \mathbb{N}^2, P_g^{i',j'} \in C_g \Rightarrow (P_g^{i',j'} + d(P_g^{i',j'})) \in C_d \end{aligned} \quad (\text{III.24})$$

Si une primitive P_g appartient à un contour C_g dont la majorité des primitives incluses dans ce contour ont été appariées avec des primitives appartenant à un contour C_d , alors le correspondant de P_g est dans C_d . Cette contrainte a été utilisée uniquement à but correctif *a posteriori*. Bien que logique, elle demande de connaître les contours dans les deux vues ce qui fait appel à un autre processus comportant une marge d'erreur.

3.3.9 Conclusion sur les contraintes

Les contraintes présentées précédemment sont les plus couramment utilisées. Lors de nos tests, nous avons utilisé la contrainte épipolaire, la contrainte d'unicité et la contrainte d'ordre

qui sont des contraintes géométriques. Nous avons aussi utilisé la contrainte sur le gradient de disparité qui permet de déterminer directement si le voisinage de deux points se ressemble suffisamment pour qu'ils soient homologues.

L'utilisation de contraintes a effectivement fait chuter le nombre de mauvais appariements. Toutefois, ces résultats n'ont pas été suffisants du fait de la géométrie choisie pour la cabine. Plutôt que de rajouter des contraintes qui auraient sans doute amélioré nos résultats — mais non significativement — nous nous sommes d'abord tournés vers les méthodes accélératrices, qui permettent outre l'accélération du processus de retirer avant comparaison un grand nombre de mauvaises possibilités d'appariement.

3.4 Méthodes accélératrices

Outre la rectification épipolaire qui permet au biais d'une transformation de simplifier considérablement les calculs, d'autres méthodes existent, l'idée générale étant de minimiser la zone de recherche. Pour cela il y a deux façon de procéder. La première consiste à contraindre la zone de recherche. La seconde, à la discrétiser en ne gardant que des points remarquables.

3.4.1 Réduction de la zone de recherche

La contrainte épipolaire nous permet de limiter la zone de recherche à une droite ou une bande très fine. Il est possible de limiter d'avantage cette zone si l'on connaît une enveloppe de la scène, voir Fig. III.12.

En calculant le point le plus à gauche et le point le plus à droite sur l'épipolaire représentant la projection de l'enveloppe, la zone de recherche est limitée à un segment ou un rectangle aplati.

Il est de plus possible d'utiliser la contrainte d'ordre pour restreindre d'autant plus la zone de recherche, si les objets de la scène ne sont pas disjoints. En effet, si on connaît un appariement sur la même épipolaire, il est inutile d'essayer de le croiser dans la recherche. Cette méthode expose toutefois au risque dû à l'incertitude de l'appariement qui nous permet de restreindre la zone de recherche.

De plus, ce type de restriction doit être codé dans un but d'optimisation temporel en plus d'une optimisation en terme de bon appariement. A cette fin, l'enveloppe doit être une forme géométrique suffisamment simple à projeter dans les matrices de projection correspondant aux deux caméras afin d'obtenir leurs bornes.

Comme expliqué au chapitre I section 3.2, un scanner à silhouette peut permettre d'acquérir la projection d'une enveloppe vue d'une direction. Il est possible de combiner une première phase

permettant d'acquérir l'enveloppe vue des deux appareils afin de simplifier les calculs dans une seconde phase de stéréovision.

Malheureusement, il n'est pas possible d'utiliser ceci dans le cadre de la cabine pour son application dans le cadre du projet Bramms. L'hypothèse de travail qui a été faite afin de restreindre la zone est que la scène se situe au centre de la cabine dans un cylindre de rayon variable (par défaut : un mètre). Les bornes sont ainsi calculées rapidement lors de la création de la zone de recherche pour chaque primitives de I_g .

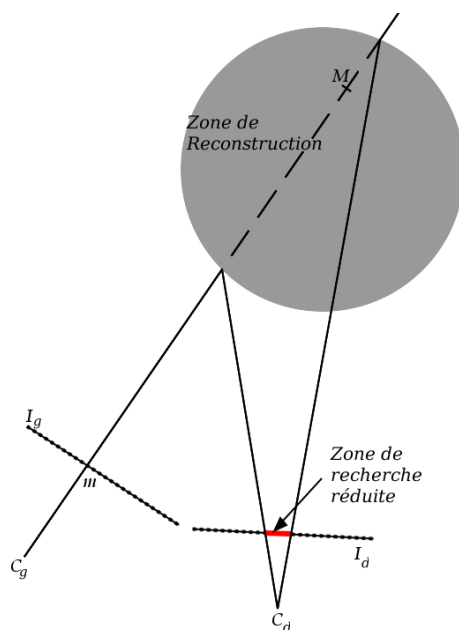


Figure III.12 – Réduction de la zone de recherche en connaissant l'enveloppe.

3.4.2 Caractérisation locale des points remarquables

On distingue deux types de points remarquables : les points de contours et les points d'intérêt. Ces points permettent de diminuer drastiquement le nombre de test à effectuer pour trouver le bon appariement. Combinés à une réduction de la zone de recherche, les candidats sont d'autant moins nombreux.

Toutefois, même s'ils sont relativement précis, il ne permettent pas une reconstruction dense de la surface lorsqu'ils sont employés dans le cadre d'une reconstruction par stéréovision. Ils peuvent toutefois servir à contraindre d'avantage la surface, les appariements sur ce type de point ayant globalement de meilleures chances d'être de bons appariements.

De plus, du fait de leur situation, ces points remarquables contiennent de l'information supplémentaire qui peut être exploitée. Ces données appelées descripteurs augmentent d'autant plus la discrimination.

Points d'intérêt

Les points d'intérêts possèdent des caractéristiques particulières les rendant plus discriminants lors de la mise en correspondance. On peut distinguer deux sous-familles :

- ▷ **Les détecteurs basés sur des propriétés spécifiques en présence de contours :** Le détecteur le plus populaire est le détecteur d'Harris et Stephens [HS88b] qui est utilisé dans de nombreux algorithmes de reconstruction [LQ05, PCZL09]. D'autres méthodes [MC04, PNF⁺08] préfèrent utiliser la variante de Shi et Tomasi [ST94b]. D'autres méthodes [IG98, KRS04] utilisent le détecteur SUSAN [SB97]. Ces trois algorithmes détectent la présence de coins dans l'image. On peut définir la notion de coin comme celle d'angle important dans un contour.
- ▷ **Les détecteurs basés sur les régions :** Matej [ML96] fut l'un des premiers à utiliser ce type de descripteur pour faire de la reconstruction 3D en utilisant des "blobs" [KPT77]. Le détecteur le plus populaire est le SIFT [Low04]. Il est notamment utilisé dans les algorithmes de reconstruction de Brown [BL05] et Peng [PCZL09]. On trouve aussi une variante du SIFT appelée SURF [BTVG06] qui contre une perte en discrimination apporte un gain de temps considérable. Cette variante est aussi utilisée pour faire de la reconstruction [BETVG08]. Parmi les autres détecteurs, il faut citer DAISY [TLF08] utilisé lui aussi pour faire de la reconstruction [Eng10]. Ces détecteurs sont tous basés sur la même idée pour déterminer leur position. On cherche les extrema locaux de la réponse de l'image à un filtre de type ondelette. C'est le descripteur associé et le type d'ondelette qui change à chaque fois.

Plus de détecteurs et descripteurs peuvent être trouvés dans [GBTD⁺09] et ses références.

Points de contours

Les points de contours sont généralement détectés par le détecteur de Canny [Can86a]. C'est le cas dans les algorithmes de reconstruction de Xu [XZ96], Ishikawa [IG98], Petrakis [PDM02] et Xiao [XS03]. Belli [BCPF00] préfère utiliser l'opérateur de Shen et Castan [SC92] dans son algorithme. D'autres [PdC03] préfèrent utiliser le Laplacien de Gaussienne ou d'autres méthodes basées sur les dérivées premières des niveaux de gris.

Ce type de détecteur tend à être de moins en moins utilisé du fait qu'il est parfois difficile d'extraire les contours d'une image, et que l'on a aucune information sur les zones n'en

comportant pas. On préfère actuellement les approches basées sur des détecteurs de points remarquables basés sur une approche région. Il n'est pas rare de trouver des méthodes mélangeant les différents détecteurs, ce qui ne peut qu'améliorer la détection en ajoutant de l'information dans les zones où elle manque.

Descripteurs de points remarquables

De par leurs situations, les points remarquables sont porteurs d'informations. Il est possible de l'encoder de plusieurs manières. Le descripteur est le vecteur d'information lié à un point remarquable qui encode les informations que l'on désire exploiter.

Parmi les données encodées sur le descripteur, on peut trouver :

- un histogramme local du voisinage [CJ04]
- un histogramme local du gradient dans le voisinage. C'est le cas pour la majeure partie des points remarquables détectés sur une approche région (SIFT [Low04], SURF [BTVG06], DAISY [TLF08], etc.)
- une caractérisation des droites formant un angle [SP03]
- un invariant local [SM⁺95]

3.4.3 Conclusion sur les méthodes accélératrices

Dans le cadre de nos travaux, nous avons implémenté une réduction de la zone de recherche basée sur la contrainte épipolaire et en prenant comme hypothèse que la scène à reconstruire se situe dans un cylindre au centre de la cabine. Cela nous a permis de restreindre les zones de recherche à des segments d'épaisseur fine.

Nous avons aussi utilisé des descripteurs d'image adaptés aux motifs afin de contraindre d'autant plus la zone de recherche, voir chapitre IV et section 4.

Ces méthodes offrent un réel gain. En contrepartie, il est nécessaire de calculer la zone de recherche restreinte pour chaque colonne (dans notre système) et/ou la rectification. Cela peut ne pas toujours être rentable. On peut notamment remarquer que le coût d'une rectification épipolaire pour une grande image demande plus de calcul lors d'un appariement éparé que le calcul direct. De plus, il est difficile d'être sûr que le point considéré a une image dans I_g et une autre dans I_d et qu'elles soient bien situées.

Mais tout ceci n'est utile que si l'on peut définir une mesure pour déterminer si deux points sont bel et bien homologues.

3.5 Mesures de corrélation

Il existe un très grand nombre de mesures de corrélation. On notera notamment les travaux de Chambon [Cha05, chap3] qui sont très aboutis dans ce domaine.

Le terme “mesure de corrélation” peut désigner aussi bien le degré de similarité que le degré de dissimilarité entre deux pixels. Le score de corrélation est la valeur de cette mesure. Il est évalué à partir de la zone d’agrégation $\mathbb{A}(P_g)$ qui correspond au voisinage de P_g et à partir de $\mathbb{A}(P_d)$. Généralement, une fenêtre centrée sur le pixel est utilisée pour définir ce voisinage. La taille de cette fenêtre devient donc un paramètre à prendre en compte dans l’algorithme.

La présence de bruit, le changement de luminosité, les déformations dues à la projection et les occultations contredisent l’hypothèse H1 (et donc H2). Pour le bruit, un pré-traitement pour débruiter l’image est nécessaire. Le changement de luminosité peut se résoudre en utilisant des mesures centrées et normalisées. Il n’existe cependant pas de remède miracle pour les occultations et les déformations dues à la projection ; même si plusieurs méthodes existantes essayent de pallier ce problème.

La chaîne d’acquisition photographique prend en compte de nombreux éléments et sa modélisation est difficile. La fonction d’étalement du point PSF¹ joue un rôle dans l’acquisition d’une image [Goo05]. En réalité, les projections ne sont pas parfaites et on obtient la projection de taches lumineuses et non directement des points projetés sur le capteur. Ceci est la principale cause d’apparition de bruit gaussien à l’acquisition. Sans faire de généralité, l’image obtenue peut être vue comme la convolution de la projection parfaite de la scène par la fonction PSF. En utilisant P la matrice de projection définie à l’équation III.8 et en symbolisant la scène par S , une image est formée ainsi :

$$I = (PS) * \text{PSF}(C) + \text{IMP}(C) \quad (\text{III.25})$$

Toutefois, le capteur peut aussi avoir des défauts comme les pixels chauds et les pixels morts. Ceci revient à ajouter plus de bruit impulsif IMP sur l’image I . Même si ce n’est pas un phénomène rare et qu’on trouve généralement quelques pixels défectueux dans chaque capteur CCD (ou CMOS), ils se limitent à un très petit nombre. On se contente donc généralement de débruiter uniquement le bruit gaussien. De plus les pixels chauds et morts sont généralement liés à l’appareil de capture C . Il suffit de les détecter une fois pour connaître les rares points problématiques lors de traitements sur une image acquise par ces capteurs.

Pour pallier le problème de déformation due à la projection, il est possible d’adapter locale-

1. Point Spread Function

ment la fenêtre de corrélation à la courbure et l'orientation locale de la surface considérée. Pour faire cette opération, il est nécessaire d'estimer les paramètres de la transformation en chaque pixel [Gar01, chap. 2 et 3]. Cette opération demande donc bien plus de calcul.

Pour pallier le problème des occultations, il est nécessaire de faire appel à des mesures naturellement plus robustes aux occultations. Ces mesures intègrent ce problème dans leur schéma de calcul du score de corrélation (voir les travaux de Chambon [Cha05, chap3]).

La suite de cette section présente quelques mesures parmi les plus utilisées et parmi les plus efficaces. Nous utiliserons la notation ω pour désigner le vecteur contenant toutes les intensités lumineuses (niveau de gris dans le cas mono-dimensionnel) contenue dans la fenêtre de corrélation. Sa dimension dépend directement de la taille de la fenêtre n . Pour simplifier les expressions, le domaine Ω désignera \mathbb{N}^{n^2} si l'on reste dans le cas usuel où les luminances sont codées sur des entiers positifs ou \mathbb{R}^{n^2} sinon.

3.5.1 Corrélation croisée

Les mesures de corrélation croisée sont toutes des mesures de similarité. C'est à dire que deux points homologues ont un score élevé.

La corrélation croisée CC² consiste à faire un simple produit scalaire entre les vecteurs ω_g et ω_d :

$$\begin{aligned} \text{CC} : \quad \Omega \times \Omega &\longrightarrow [-\infty; \infty] \\ (\omega_g, \omega_d) &\longmapsto \omega_g \times \omega_d \end{aligned} \tag{III.26}$$

Elle ne peut être directement exploitée du fait que le score varie suivant l'intensité lumineuse des zones d'agrégation considérées. Il est nécessaire de lui apporter au moins une étape de normalisation. On la désigne alors sous le nom NCC³ et sous les réserve que les vecteurs soient non null, on a :

$$\begin{aligned} \text{NCC} : \quad \Omega^* \times \Omega^* &\longrightarrow [-1; 1] \\ (\omega_g, \omega_d) &\longmapsto \frac{\omega_g \times \omega_d}{\|\omega_g\|_2 \|\omega_d\|_2} \end{aligned} \tag{III.27}$$

Bien qu'alors NCC soit utilisable, elle n'en reste pas moins sensible au changement de luminosité et nécessite un centrage.

La mesure de corrélation croisée centrée normalisée ZNCC⁴ est, en notant $\bar{\omega}$ la moyenne

2. Cross Correlation
3. Normalised Cross Correlation
4. Zero mean Normalized Cross Correlation

des éléments contenus dans ω :

$$\begin{aligned} \text{ZNCC} : \quad \Omega^* \times \Omega^* &\longrightarrow [-1; 1] \\ (\omega_g, \omega_d) &\longmapsto \frac{(\omega_g - \bar{\omega}_g) \times (\omega_d - \bar{\omega}_d)}{\|\omega_g - \bar{\omega}_g\|_2 \|\omega_d - \bar{\omega}_d\|_2} \end{aligned} \quad (\text{III.28})$$

Cette mesure est identique à la corrélation linéaire classiquement utilisée en statistiques. Bien qu'elle soit l'une des plus utilisée, il existe une version plus robuste qui consiste à remplacer la moyenne par la médiane, et la norme L_2 par la norme L_1 . Cette mesure plus robuste ZNCC_R a été introduite par Trujillo et Izquierdo [TI04]. Cette variante demande toutefois de par sa modification un accroissement considérable du temps de calcul nécessaire.

Notons enfin qu'en changeant la normalisation utilisée, on retrouve la mesure de corrélation de Moravec [Mor80].

3.5.2 Statistique

Les mesures statistiques sont toutes des mesures de dissimilarité. C'est à dire que deux points homologues ont un score faible. Il reste toujours possible de prendre le négatif de la mesure pour uniformiser l'implémentation (le meilleur est alors le plus élevé).

Les mesures de type distances sont aussi fréquemment utilisées. Il s'agit d'utiliser une norme L_i sur la différence des zones de corrélation $\omega_g - \omega_d$ avant de mettre le tout à la puissance i . Dans la littérature, on trouve le terme SAD⁵ lorsque la norme L_1 est utilisée et SSD⁶ lorsque la norme L_2 est utilisée.

A ces deux mesures peuvent être appliquées des opérations de centrage et/ou de normalisation comme pour la construction de ZNCC. Il existe aussi des variantes localement centrées — qu'on note généralement avec un L à la place du Z — qui consistent à ramener le niveau de gris de la zone de gauche sur la zone de droite en utilisant le ration des moyennes. Par exemple, pour LSSD :

$$\begin{aligned} \text{LSSD} : \quad \Omega \times \Omega &\longrightarrow [0; I_{max}n^2] \\ (\omega_g, \omega_d) &\longmapsto \left\| \omega_g - \frac{\bar{\omega}_g}{\bar{\omega}_d} \omega_d \right\|_2^2 \end{aligned} \quad (\text{III.29})$$

Des mesures de corrélation basées sur la variances des différences VD — ou VAD lorsque les différence sont absolues — ont été introduite dans [Cox95].

Enfin, notons aussi cette famille basée sur le “kurtosis” K_4 , proposée par Rziza et al. [RAMT01] :

5. *Sum of Absolute Difference*

6. *Sum of Square Difference*

$$\begin{aligned} K_4 : \quad \Omega \times \Omega &\longrightarrow [0; I_{max}^4] \\ (\omega_g, \omega_d) &\longmapsto \left| \overline{((\omega_g - \omega_d)^4)} - 3 \overline{((\omega_g - \omega_d)^2)^2} \right| \end{aligned} \quad (\text{III.30})$$

3.5.3 Dérivée

Cette famille de méthode ne s'appuie plus directement sur les niveaux de gris mais sur les gradients des images. Les formes les plus connues de mesures utilisant la dérivée sont les opérateurs de Seitz [Sei89]. Il s'agit de calculer la norme L_p de la différence de résultats fournis par des opérateurs de dérivée : $\|R_O(\omega_g) - R_O(\omega_d)\|_p$. La notion de dérivée en image n'est pas unique car les images sont des éléments discrets. Avec l'opérateur de Sobel [SF68] (à la place de R_O) nous obtenons les mesures SES_p ⁷. Avec l'opérateur de Kirsh, nous obtenons les mesures SEK_p ⁸. Les mesures de Seitz sont des mesures de dissimilarité.

Il existe de même des mesures s'appuyant sur la corrélation croisées de vecteurs d'information construit avec les résultats d'autres opérateurs de dérivée. La mesure de Nishimura [Nis84] utilise un Laplacien binaire. Nack [Nac75] fait de même avec l'opérateur de Roberts dont le résultat est binarisé. Pratt [Pra78] l'utilise quant à lui sur l'image du Laplacien. NIS, PRATT et NACK sont des mesures de similarité.

Enfin, d'autres mesures utilisent un encodage de la direction du gradient [UK04] ou encore la direction et sa norme [CMP96].

3.5.4 Autres mesures

Il existe quantité d'autres mesures qui ne seront pas présentée ici. Chambon distingue dans son état de l'art des mesures non paramétriques (comme celle du χ^2 ou encore des mesures de rang) et des mesures robustes (utilisant la médiane, des pseudo normes ou encore des estimateurs sur les résidus des moindres carrés).

Nous avons notamment implémenté la mesure MAD ⁹. Cette méthode se base sur la valeur absolue de la déviation par rapport à la valeur médiane des déviations. Plus la valeur de cette mesure est proche de 0, plus les vecteurs d'information ω_g et ω_d tendent à avoir une évolution similaire de leurs composantes. On peut donc en conclure que ces vecteurs d'information ont de grande chance d'être similaire et donc qu'ils correspondent très probablement à des voisinages

7. SEitz Sobel using L_p

8. SEitz Kirsh using L_p

9. *Median Absolute Deviation*

de points homologues.

$$\begin{aligned} \text{MAD} : \quad \Omega \times \Omega &\longrightarrow [0; +\infty[\\ (\omega_g, \omega_d) &\longmapsto \text{med } |(\omega_d - \omega_d) - \text{med } (\omega_d - \omega_d)| \end{aligned} \quad (\text{III.31})$$

4 Évaluation

La chaîne complète de formation des images et l'algorithme complexe de reconstruction ont été développés dans la section précédente. Les tests réalisés afin de déterminer la méthode la plus adaptée à notre problème sont maintenant présentés.

4.1 Protocole expérimental

Afin de tester les différentes mesures utilisables dans la cabine d'acquisition nous nous sommes placés comme cadre l'étude de couples de photographies prises par des appareils éloignés de 30°. C'est à dire lorsque le centre de la scène et les deux capteurs forment un triangle équilatéral de côté 1,5m.

A cause de l'homogénéité de la texture de la peau humaine, les résultats attendus étaient inexploitable. Nous avons donc décidé, comme cela a été dit au chapitre précédent, de projeter un motif afin d'ajouter de l'information, et ainsi casser l'homogénéité. Nous avons choisi à cette fin de projeter un nuage de taches positionnées aléatoirement mais séparables. Chaque tache étant une ellipse de longueur maximale, de longueur minimale, et d'orientation aléatoires. Chaque tache comporte de plus un bruit gaussien d'amplitude différente à sa création. Le but de cette manipulation était d'ajouter une information la moins redondante possible afin d'obtenir des prises de vue sans zones homogènes.

Nous avons dans un premier temps lancé la reconstruction avec plusieurs mesures de corrélation différentes et plusieurs tailles pour la zone d'agrégation (de type fenêtre centrée). Nous avons ajouté successivement la contrainte de gradient de disparité avec différentes valeurs de seuil, puis la contrainte épipolaire au niveau de l'algorithme. Après son exécution, nous avons fait un filtrage *a posteriori* avec la contrainte d'unicité et la contrainte de symétrie. Les mesures ZNCC, ZNSSD, MR6 et MAD ont été testées. Nous avons choisi de nous concentrer sur MR6 et MAD pour leur robustesse face aux occultations qui ressortait des travaux de Chambon (MR6 est un estimateur sur les résidus des moindres carrées).

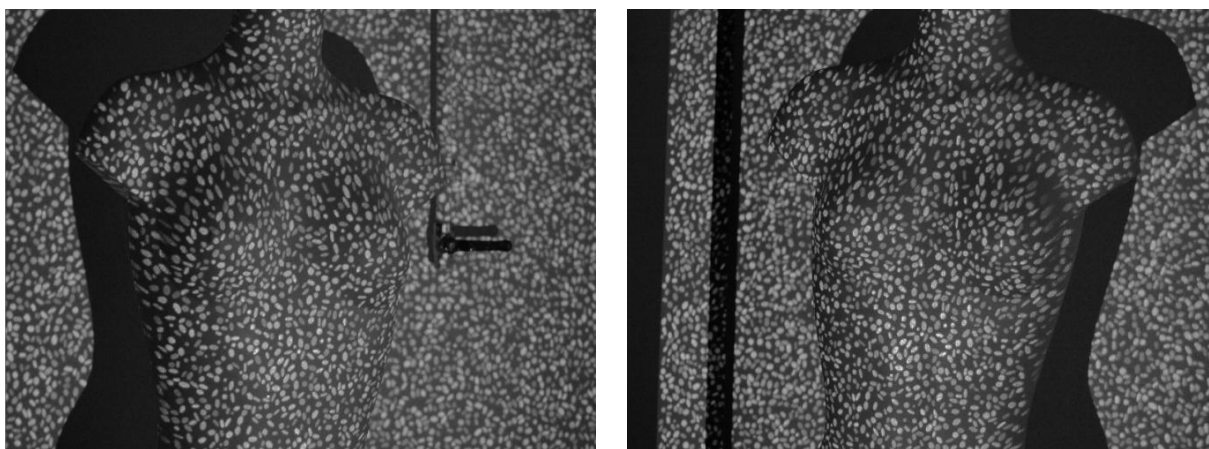


Figure III.13 – Taches.

4.2 Résultats

Contrairement à ce que les travaux de Chambon ont montré, la mesure MAD — très robuste face aux occultations mais moins bonne que les mesures classiques sur des images réelles — se comportait mieux que la mesure ZNCC. Ceci est sans doute dû à une meilleure robustesse face à la déformation induite par la projection dans les deux vues. Un résultats de reconstruction à partir de cette mesure est présenté dans la figure [III.14](#).

Les résultats ici exposés sont montrés sous forme de nuage de points tridimensionnels. Les pixels bien appariés forment des points sur la surface à numériser en utilisant la triangulation spatiale expliquée au chapitre [V](#). Le problème est de déterminer comment filtrer les nombreux mauvais appariements. Le nombre important de mauvais appariements est dû à l'écart important utilisé entre les appareils de captures.

Nous avons appliqué les restrictions de symétrie et d'ordre de même qu'un seuillage sur la réponse de la fonction de corrélation. Si ces méthodes font effectivement baisser le pourcentage de mauvais appariements, elles suppriment aussi quelques appariement valide. Au vu du pourcentage important de mauvais appariement avoisinant les 50%, appliquer ce genre de méthode ne laissent pas assez de points pour garantir une reconstruction tridimensionnelle de qualité.

5 Conclusion

Le but de ce chapitre est d'expliquer la mise en correspondance de points. L'étude de la formation des images permet de mettre en évidence les relations géométriques entre deux images de la même scène prises à des positions différentes. Le principe de la géométrie épipolaire y a

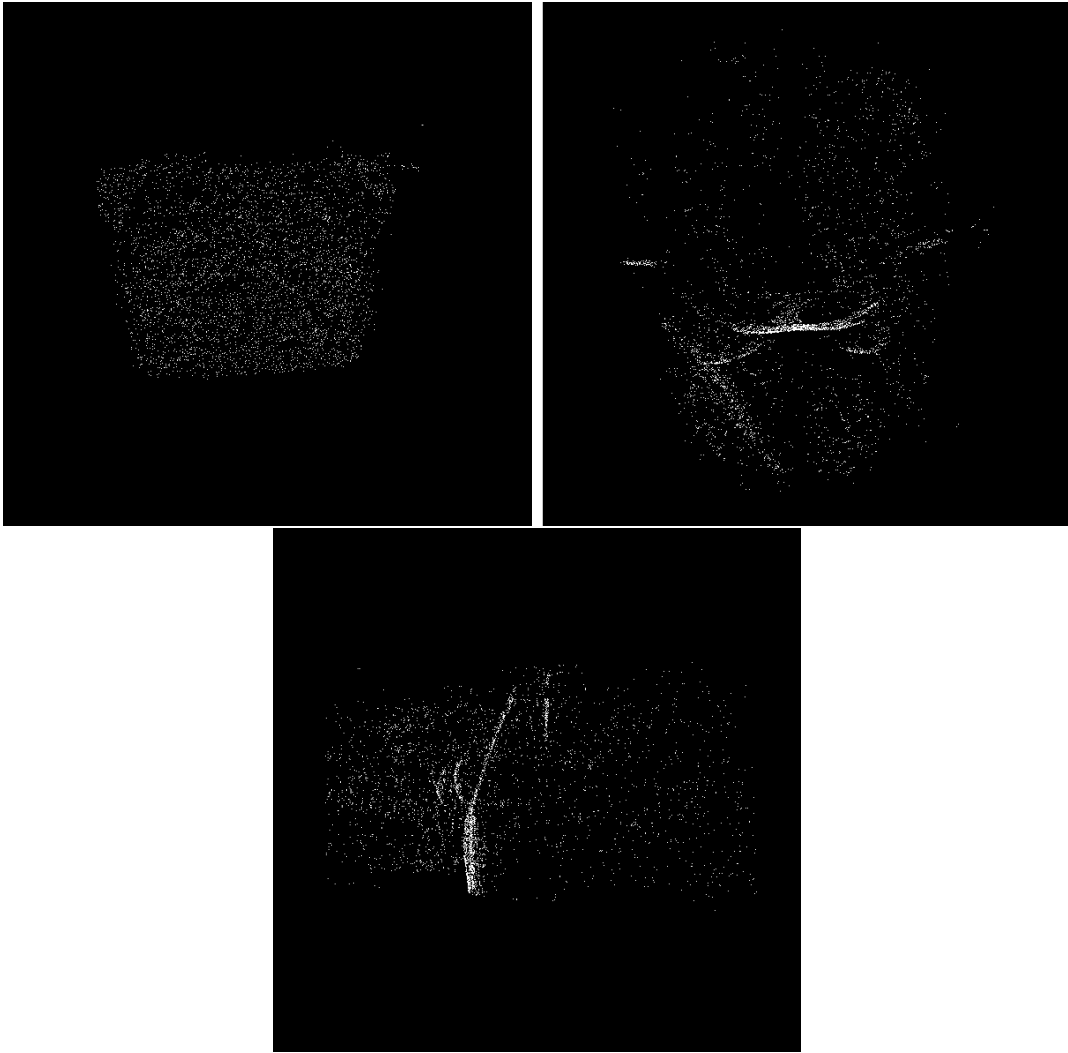


Figure III.14 — Nuage de point acquis lors de la reconstruction d'un buste de mannequin en vue de face, de dessus et de la droite. Les résultats présentés ici sont ceux liés à la méthode MAD qui nous a donné les meilleurs résultats. La zone de recherche a été restreinte suivant la géométrie de la cabine. Il n'y a pas ici de filtrage sur la réponse de la fonction de corrélation.

donc été expliqué, voir section 2.

Le problème de la mise en correspondance a été formulé sous forme algorithmique et énergétique. C'est un algorithme complexe qui prend en compte de nombreuses variantes. Les contraintes s'y appliquant ont été détaillées. Leur utilisation permet de contraindre l'algorithme et donc d'augmenter le nombre de couples de points homologues appariés.

La formation des images au sein d'un appareil photographique qui a été présenté dans la section 1, voir équation III.8 a été approfondie par l'équation III.25 qui explique l'apparition de bruit dans les images. Ce problème s'ajoute à celui évident de la déformation de la scène

lors de la projection dans le plan image. A ces deux problèmes viennent s'ajouter les problèmes liés à l'occultation de certains points de la scène dans une des deux vues et au changement d'illumination qui est lié à l'angle entre la source lumineuse éclairant la scène et le capteur d'image.

Ces problèmes bien qu'ils puissent en partie être simplifiés par les contraintes sur l'algorithme (notamment pour les occultations) interviennent directement au niveau du calcul de la corrélation permettant de déterminer si un couple doit être apparié ou non. Les mesures de corrélation les plus usitées ont été présentées.

Nous avons donc une chaîne complète pour obtenir les couples de points homologues, ou du moins supposés homologues par le score obtenu suivant la mesure de corrélation choisie. Nous avons donc appliqué cette chaîne sur un couple de capteurs d'image dans les mêmes conditions que dans l'utilisation voulue pour la cabine d'acquisition. Pour pallier le problème des larges zones homogènes qui rendent difficile le bon appariement, nous avons ajouté de l'information sur la scène en y projetant un motif aléatoire.

Les résultats obtenus et exposés montrent combien ces mesures bien qu'efficaces dans le cas classique de stéréovision binoculaire — schéma de stéréovision dans lequel les capteurs sont séparés par un déplacement de type translation uniquement, de longueur identique à la distance entre deux yeux — perdent fortement en performance lorsque les capteurs sont plus éloignés avec une rotation entre eux.

Le problème s'est alors posé de tester d'autres mesures plus robustes ou d'essayer une autre voie. Nous avons alors fait le choix de nous tourner vers l'utilisation de motifs structurés particuliers et d'exploiter les informations de ce motif lors de la mise en correspondance. Ceci fera l'objet du chapitre suivant.

Chapitre IV

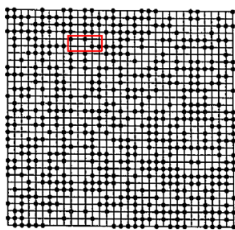
Corrélation appuyée sur un maillage

L'intérêt d'un motif pour aider à la reconstruction n'est pas une approche nouvelle [JKC85] et est généralement utilisée afin d'accélérer le processus de reconstruction pour obtenir une reconstruction temps réel. Ce domaine est toujours très actif dans la recherche aujourd'hui. Les motifs utilisés peuvent être de différentes couleurs [Har92a, HHJC99, ZCS02, ZH04, SOY+09] ou monochrome [SS03, GNG+08].

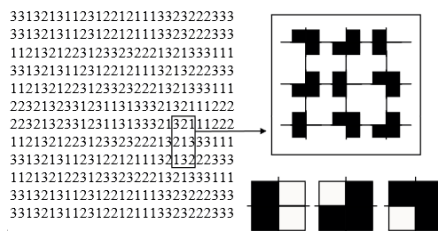
Il y a de plus plusieurs façons d'utiliser la lumière structurée. L'une des premières fut l'usage de Moiré [Har92a] qui peut s'apparenter sous bien des aspects à l'holographie conoscopique. Les motifs à frange sont devenus très populaires par la suite pour leur simplicité de construction et d'extraction [HHJC99, ZH04].

De nombreux autres motifs ont été proposés [RWC+98, RHHL02, HHR01]. Seules certaines équipes de recherche se sont penchées sur l'unicité des motifs à des fins de reconstruction tridimensionnelle [SPB04]. Le premier usage de motif basé sur des graphes a été proposé par Le Moigne et al [LMW88] avec des tableaux binaires pseudo aléatoires. On pourra aussi évoquer les petits motifs utilisés comme mots clefs développés par Griffin et al [GNY92].

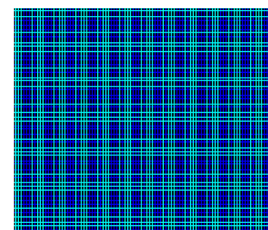
Plusieurs techniques sont basées sur une approche spatio-temporelle de la stéréovision [ZCS03, DRR03] qui est l'appariement spatial lors de la stéréovision dans le domaine temporel [ZCS02, ZH04] sur une séquence d'image. Ceci peut s'apparenter, de façon un peu extrême, à une reconstruction par balayage haute fréquence à l'aide un petit motif tel qu'un point. C'est ce qui



JJ Le Moigne et al. (1988)



Paul M Griffin et al. (1992)



Ryusuke Sagawa et al. (2009)

est utilisé par exemple dans la maintenant bien connue *Kinect* [Kho11b].

Finalement, une étude récente sur les techniques de reconstruction se doit d'être citée [Gen11]. Elle présente de nombreuses approches de type lumière structurée avec des motifs aussi bien monochromes que colorés et aussi bien discrets que continus.

La construction du motif et son extraction dans une scène sont les deux principales utilisations. En réalité, la construction peut se permettre d'être complexe car le motif n'est pas généré régulièrement. Son extraction en revanche se doit d'être particulièrement efficace. Ce que l'on va notamment rechercher lors de l'extraction est donc une identification précise (bonne distinction, bonne localisation) tout en tenant compte de la rapidité du processus.

Ce chapitre débutera donc par l'utilisation de la lumière structurée en tant que support dans le cadre de la cabine d'acquisition. L'utilisation d'un motif en tant que contrainte du support dans l'algorithme d'appariement (Algorithme 1) détaillé au chapitre III y est donc étudié. C'est la prolongation directe du chapitre précédent et nous y emploierons donc le motif elliptique aléatoire qui y apportait de la discrimination.

La solution ensuite proposée dans ce chapitre se base sur des propriétés d'unicité, ce qui entraînent donc une bonne distinction. Son extraction a été implémentée en minimisant le nombre de calcul, ce qui entraîne donc une bonne rapidité du processus d'extraction. De plus, l'extraction a été implémentée suivant le modèle de formation des images développé dans le chapitre III.

La projection d'un motif couvrant l'ensemble de la scène a déjà été proposé plusieurs fois (par exemple [SOY⁺09]). La nouveauté de la méthode que nous proposons et que nous présentons dans ce chapitre réside dans ses propriétés d'unicité. En cas d'unicité, il n'est en effet plus nécessaire d'utiliser la géométrie épipolaire exposé au chapitre précédent. De plus, l'algorithme complexe d'appariement utilisant un grand nombre de contraintes pour augmenter la probabilité d'obtenir de bons appariements se simplifie à une simple recherche. Qui plus est, son utilisation peut toujours fournir une restriction de la zone de recherche dans l'appariement de points.

1 Un motif utilisé comme contrainte

Dans le chapitre précédent, nous avons exposé le fonctionnement des algorithmes de mises en correspondance. Afin de pouvoir obtenir des informations dans les zones homogènes, nous avons projeté un motif composé de tâches elliptiques aléatoires en formes et en textures (différents bruits appliqués à chaque tâche). Les résultats obtenus étaient cependant insuffisant ; une étude approfondie nous a permis d'en comprendre la cause.

1.1 Proposition d'exploitation et implémentation

En cherchant les caractéristiques invariantes à la déformation due à la projection, nous avons fait l'hypothèse que le centre de chaque tâche restait le centre de cette tâche quel que soit le point d'observation.

La recherche des centres est ainsi devenue un prétraitement de l'algorithme de reconstruction avec l'objectif de limiter drastiquement le support.

Les résultats obtenus ne sont cependant pas devenus meilleurs. Afin de comprendre plus en détails le problème nous avons analysé l'éloignement entre le pixel trouvé par les mesures de corrélation et son pixel homologue.

A cette fin nous avons utilisé des appareils éloignés de 30° , disposition identique à celle de la cabine d'acquisition.

1.2 Résultats obtenus

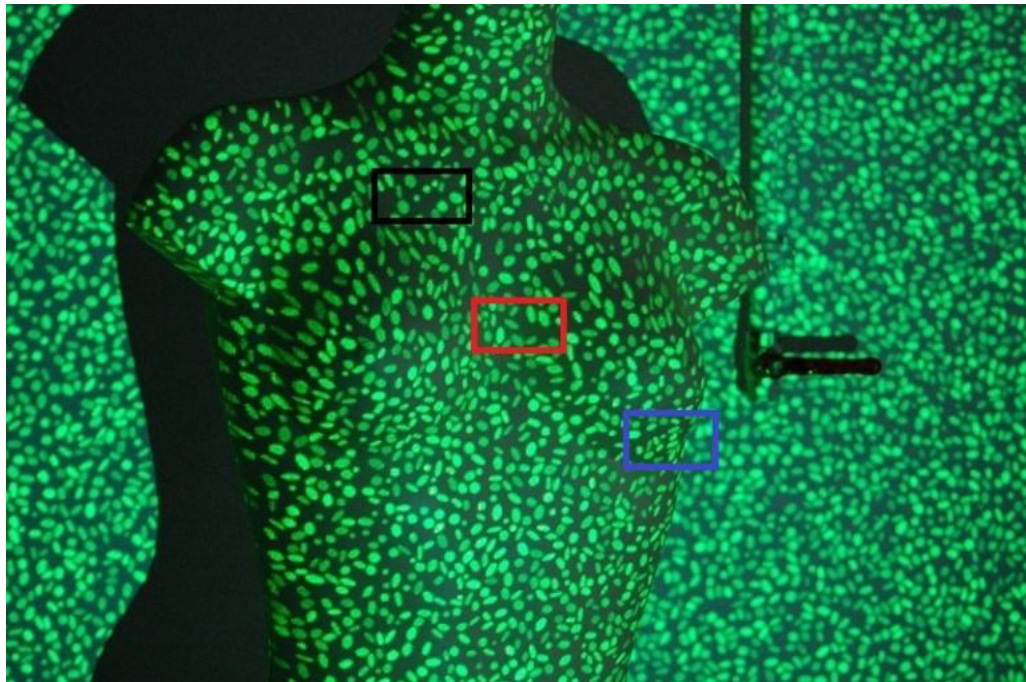


Figure IV.1 – Zones de test pour l'évaluation des mesures de corrélation à 60° . La zone rouge est quasi plane. La zone noire possède des disparités plus prononcées. La zone bleue est partiellement occultée.

Les résultats obtenus peuvent s'expliquer par le fait suivant : projeter sur une surface non-plane entraîne des déformations. Ainsi, il est nécessaire de redéfinir la notion de centre de tâche qui n'est plus forcément la position moyenne des pixels la constituant. Pire encore, la fenêtre de

Corrélation appuyée sur un maillage

Zone	Quasi planaire							Disparate	Occultée
Dist. erreur	0		1	3		10		10	10
Taille fen.	2lpix	4lpix	4lpix	2lpix	4lpix	2lpix	4lpix	4lpix	4lpix
Mesures de corrélation croisée									
NCC	7.2%	6.7%	37.8%	66.6%	78.3%	76.4%	92.6%	64.5%	7.8%
ZNCC	7.8%	6.7%	39.4%	61.3%	80.3%	70.6%	94.6%	67.1%	12.6%
MOR	8.0%	6.6%	41.6%	65.6%	82.2%	75.7%	96.6%	70.9%	13.2%
Mesures statistiques									
SAD	2.2%	5.6%	38.3%	49.8%	81.1%	58.0%	95.1%	84.2%	17.3%
SSD	2.2%	5.6%	38.3%	49.8%	81.1%	58.0%	95.1%	84.2%	17.3%
NSSD	2.2%	5.6%	38.3%	49.8%	81.1%	58.0%	95.1%	84.2%	17.3%
ZSAD	8.4%	6.9%	40.1%	69.6%	83.2%	80.6%	96.7%	82.6%	15.1%
ZSSD	8.4%	6.9%	40.1%	69.6%	83.2%	80.7%	96.7%	82.6%	15.1%
ZNSSD	8.1%	6.5%	41.1%	68.7%	85.0%	79.5%	98.3%	88.9%	15.5%
LSAD	6.3%	6.7%	42.0%	69.2%	85.6%	79.8%	99.1%	82.1%	15.6%
LSSD	6.3%	6.7%	42.0%	69.2%	85.6%	79.9%	99.1%	82.1%	15.6%
VD	7.9%	6.7%	35.7%	66.0%	74.9%	76.3%	89.2%	67.7%	9.6%
VAD	6.6%	6.3%	30.2%	54.9%	66.0%	63.2%	78.4%	44.2%	1.6%
VSD	6.2%	6.7%	34.0%	54.4%	66.0%	64.7%	80.3%	46.9%	2.5%
K ₄	6.0%	7.2%	34.2%	58.4%	68.5%	68.9%	83.6%	54.9%	5.0%
Mesures basée sur la dérivée									
SES ₁	6.7%	3.5%	10.0%	64.2%	58.9%	73.5%	62.0%	52.9%	0.0%
SES ₂	6.7%	3.5%	10.0%	64.2%	58.9%	73.6%	62.0%	52.9%	0.0%
Autres mesures									
MAD	2.6%	6.7%	6.7%	53.7%	76.6%	61.6%	94.3%	89.6%	13.6%

Tableau IV.1 – Distance à l’erreur pour différentes mesures à 60°. Les pourcentages sont les pourcentages de bon appariement dans un système stéréoscopique moyennant une liberté de quelque pixel donnée par la ligne “Dist. erreur”. Les trois zones utilisées sont celle de la figure IV.1

corrélation utilisée lors de l’algorithme d’appariement correspond à un voisinage non-uniforme (ayant plus de voisins dans certaines directions que dans d’autres) et entraîne donc des erreurs de localisation.

Ces résultats nous ont confortés sur le fait que la stéréovision seule ne donne pas de résultats satisfaisants dans le cas d’un dispositif tel que notre cabine d’acquisition comportant des angles importants entre les appareils d’acquisitions.

Nous nous sommes alors orientés vers les méthodes basées sur la lumière structurée, laissant au second plan l’aspect stéréovision pour la mise en correspondance.

2 Intérêt d'un motif de type maillage

L'équipe de recherche GAMMA3¹ dans laquelle s'est déroulée la présente thèse est spécialisée dans les maillages vus comme un support de calcul pour divers problèmes physiques. Nous nous sommes donc naturellement penchés sur l'utilisation de plusieurs motifs de ce type.

L'intérêt d'un motif de type maillage est qu'il permet une couverture totale de l'image avec un unique motif. Ce n'est pas une nouvelle approche [SOY⁺09] mais le motif ici présenté possède de bonne propriété d'unicité. De plus, les maillages possèdent une cohérence spatiale entre ses éléments constitutifs, de par ses relations de voisinage clairement définies. Enfin, à partir d'un maillage, la reconstruction de la surface sera plus rapide qu'à partir d'un nuage de points tridimensionnels. Ceci sera présenté au chapitre V.

Le premier type de maillage testé fût une grille. Outre le fait que cela se rapproche très fortement des travaux de Segawa et al [SOY⁺09], ce type de maillage est très régulier. Cela implique donc de s'appuyer sur la géométrie épipolaire pour effectuer de bon appariement. Le second maillage étudié est la triangulation (au sens pavage d'éléments triangulaires). L'étude de ce maillage a permis d'en déterminer des propriétés très intéressantes concernant l'unicité topologique de sous-ensembles le constituant.

L'unicité ici n'est pas celle du maillage qui est en effet unique car seule pour toute la scène. L'unicité est trouvée dans des propriétés topologiques en chaque endroit du maillage.

De plus, l'utilisation d'un motif coloré est impossible du fait de l'agencement de la cabine. En effet, les motifs se superposeraient. Il a déjà été établi au chapitre II que les vidéo-projecteurs projetteraient un motif monochrome dans une des trois teintes classiquement utilisées dans la perception des couleurs d'un appareil photographique (rouge, vert et bleue).

Si, dans un premier temps, nous nous sommes intéressés à un motif de type grille similaire dans son approche aux travaux de Sagawa et al [SOY⁺09], aucune particularité d'unicité n'a pu y être déterminée qui soit robuste aux changements impliqués par la projection.

Dans un second temps, nous nous sommes intéressés aux propriétés topologiques sous-jacentes dans une triangulation. Cette section décrit en profondeur comment créer de l'unicité dans une triangulation et comment l'utiliser pour effectuer de la reconstruction tridimensionnelle.

1. Génération Automatique de Maillage et Méthodes Avancées

2.1 Structure topologique d'une triangulation

Dans cette partie, une petite réflexion sur chaque élément utilisable pour distinguer un nœud d'un autre va être présentée. Nous discuterons ensuite de leur robustesse face aux déformations.

L'objectif ici est donc de construire un vecteur d'information qui permet de discriminer suffisamment un nœud d'un autre. La construction de ce vecteur se doit d'être robuste vis à vis d'une projection sur une surface non-plane observée depuis un autre point de vue.

2.1.1 Liste des éléments

Les éléments qui peuvent être utilisés pour décrire un nœud sont son degré, l'angle entre les arêtes partant de ce nœud et les triangles incluant ce nœud parmi leurs sommets.

Degré d'un nœud : La notion de degré ici est la même que celle utilisée en maillage ou en théorie des graphes. C'est à dire que le degré d'un nœud η est le nombre de nœuds du maillage connectés par une arête à η (voir Fig. IV.2 pour exemple). Cette information est donc scalaire.

Angles entre les arêtes : Chaque arête consécutive forment un angle. L'ensemble de ces angles autour d'un nœud peut être représenté sous forme de vecteur d'information sous réserve de définir un ordre pour la représentation. L'information est donc ici vectorielle ce qui implique un pouvoir de discrimination plus important.

Triangles (cas dual) : Chaque nœud est utilisé dans la construction de plusieurs triangles. Ces triangles peuvent être placés dans un vecteur sous réserve de définir un ordre pour la représentation. Toutefois, le problème est maintenant de trouver des éléments constitutifs discriminant pour les triangles.

2.1.2 Robustesse des éléments face à la transformation

Degré d'un nœud : Le degré des nœuds est stable par transformation affine et par projection sur une surface non-planaire tant qu'il est possible d'extraire la triangulation dans son voisinage.

Angles entre les arêtes : Les angles ne sont pas robuste face aux transformations affines et donc face aux projections sur des surfaces non-planaires. Il y a donc de fortes chances d'avoir des modifications dans la valeur des angles. À cause de ce manque de robustesse, les angles ne sont pas un bon choix pour établir un descripteur.

Triangles (cas dual) : Tout comme le degré des nœuds, les triangles sont stables si nous les définissons par leurs sommets. Cependant les triangles ont une représentation qui prend plus de place que les nœuds et il y a donc de plus grandes chances d’avoir des erreurs d’extraction lors de la récupération du motif. De plus, cela revient de nouveau à extraire une caractéristique pour les nœuds, ce que nous cherchons. Il vient d’être montré que les angles n’étaient pas un bon choix. Il en va de même pour les arêtes car les distances ne sont pas conservées lors des transformations. Tout ce qui en découle (les surfaces par exemple) sont donc impactées de la même manière. Nous pouvons donc statuer que l’utilisation du cas dual n’est pas adéquate dans ce cas.

Tout cela mène donc naturellement à l’utilisation d’une clef basée sur le degré des nœuds. Cette information étant scalaire, il est nécessaire de trouver un moyen d’améliorer sa puissance de discrimination. Ceci est décrit dans la partie suivante.

2.2 Construction d’une clef

Possédant seulement une information scalaire pour construire une clef d’après nos observations, on ne peut qu’avoir une clef scalaire :

$$K^0(p) = \deg(p) \quad (\text{IV.1})$$

Soit \mathcal{N} l’ensemble des nœuds de la triangulation. En recherchant un nœud correspondant $n \in \mathcal{N}$ au nœud $\eta \in \mathcal{N}$, il est facile de vérifier que $K^0(\eta) = K^0(n)$. Si ce n’est pas vrai, les deux nœuds ne peuvent pas se correspondre. Notez que l’implication inverse n’est pas vraie. Nous parlerons ici de nœuds homologues selon le critère de degré utilisé et nous utiliserons donc les notations utilisées au chapitre III “ \doteq ” et “ \ncong ” :

$$\forall (\eta, n) \in \mathcal{N} \times \mathcal{N}, \left(K^0(\eta) \neq K^0(n) \right) \Rightarrow (\eta \ncong n) \quad (\text{IV.2})$$

Ceci n’est pas suffisant pour discriminer efficacement un nœud même dans le cadre d’un algorithme d’appariement utilisant la contrainte épipolaire, voir chapitre III, et cela à cause de la répartition des degrés des nœuds dans une triangulation. Pour une triangulation aléatoire, la probabilité du degré de chaque nœud est centrée sur 6 à partir d’un certain nombre de nœud dans la triangulation. Il est donc impossible d’y trouver des clefs uniques sauf si la densité de la triangulation est faible (en terme de nombre de nœud). Nous reviendrons sur ce fait dans la section 2.3. Cela peut toutefois être suffisant pour des triangulations construites spécifiquement (répartition non-aléatoire).

2.2.1 Rang d'une clef

Il est possible d'ajouter plus d'information et donc plus de puissance de discrimination en utilisant les relations de voisinage intrinsèquement définies par la triangulation. La clef précédemment définie va être étendue à son voisinage pour le premier rang. Puis, elle sera étendue au voisinage du voisinage, etc. A cette fin, une définition récursive de K est introduite.

Soit $\mathcal{V}(\eta) = [\mathcal{V}_0(\eta), \dots, \mathcal{V}_{\deg(\eta)-1}(\eta)]$ le vecteur encodant tous les nœuds du voisinage de η triés par ordre horaire. Le rang i de la clef notée K^i peut se définir ainsi (voir figure IV.2) :

$$K^i(\eta) = [K^{i-1}(\mathcal{V}_0(\eta)), \dots, K^{i-1}(\mathcal{V}_{\deg(\eta)-1}(\eta))] \quad (\text{IV.3})$$

L'équation IV.2 reste vraie quand le rang change. Cependant, le premier nœud de $\mathcal{V}(\eta)$ est inconnu et dépend de l'ordre horaire choisi. A cause de la projection du motif sur une surface non-plane, il est possible que le premier élément choisi pour une clef ne soit pas le même pour un même nœud après projection. Ce paramètre n'est donc pas fixé et les comparaisons doivent se faire à une rotation près dans $\mathcal{V}(\eta)$.

Il n'existe donc toujours pas de critère disant "ils sont homologues" mais uniquement un critère disant "ils ne sont pas homologues". En fait, il est possible de démontrer probabilistiquement l'ajout de discrimination qui a été fait en prenant une clef de rang supérieur, voir section 2.3.2. Le rang peut donc être choisi pour avoir une plus grande probabilité d'unicité de chaque clef et ceci dépendant seulement du nombre $|\mathcal{N}|$ de nœuds dans la triangulation. De plus, le critère de discrimination est exponentiel en fonction du rang. Ceci est dû au fait que chaque critère inclus les critères de rang inférieur dans sa définition, voir équation IV.6.

Ce critère est récursif de par sa définition. C'est pourquoi il est plus aisé de l'implémenter récursivement. En pratique, le rang n'a pas besoin d'être très élevé, ce qui ne causera aucun problème de pile d'exécution avec une implémentation récursive. De plus, il est aisé de calculer dans un premier temps toutes les clefs de rang 0, puis toutes celle de rang 1, etc.

2.2.2 Utilisation des clefs

De par leur construction, les clefs ne dépendent que de la topologie de la triangulation projetée. C'est pourquoi si la projection est correctement acquise dans l'image obtenue par chaque appareil de capture, deux nœuds n'ont une clef identique que s'ils correspondent au même point tridimensionnel. Les seuls échecs n'arrivent que si la clef possède un doublon.

Dans tous les cas, une extraction du motif projeté doit être faite en amont. Cette extraction n'a pas besoin d'être totale car seule une petite partie de la triangulation est utilisée à chaque

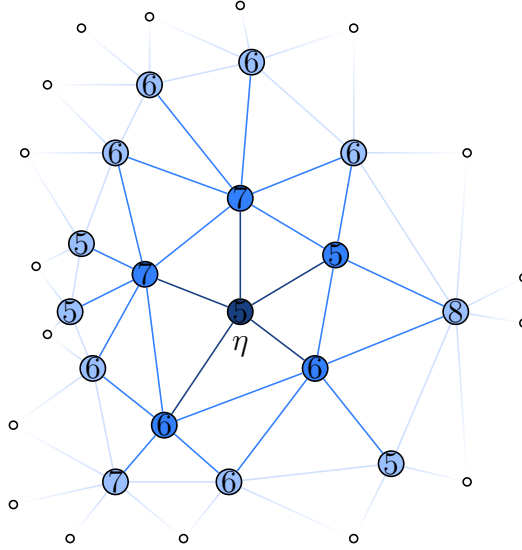


Figure IV.2 – Partie d’une triangulation de Delaunay. Le degré de chaque nœud est inscrit dans celui-ci. $K^0(\eta) = 5$. $K^1(\eta) = [7, 5, 6, 6, 7]$ ou n’importe quelle rotation de ce vecteur. $K^2(\eta) = [[5, 7, 6, 6, 6, 5], [5, 7, 6, 8, 6], [5, 5, 8, 5, 6, 6], [5, 6, 6, 7, 6, 7], [5, 6, 6, 5, 5, 6, 7]]$ ou n’importe quelle rotation incluant n’importe quelle rotation des sous-vecteurs.

calcul de clef. Le but est d’obtenir une bonne localisation des nœuds dans les images obtenues par les appareils photographiques. Ceci a été abordé dans la partie précédente, voir section 4.

La meilleure utilisation que l’on puisse faire de ces clefs est de s’arranger pour qu’elles soient uniques. Si à cause d’une occultation, le rang souhaité n’est pas calculable, il est toujours possible de faire appel à la cohérence spatiale du voisinage et donc de s’appuyer sur une clef voisine qui a été identifiées. Ainsi, connaissant un rang moindre pour ce nœud et un voisin de rang unique, son identification est relativement aisée.

Afin de calculer rapidement les clefs, il est préférable de calculer dans un premier temps toutes les clefs de rang 0. Cela revient à calculer le degré de chaque nœud. Ensuite en utilisant la définition récursive, on calcule les rangs suivants. La comparaison peut alors être effectuée pour les rangs maxima. Les rangs antépénultièmes restant sont alors comparés en tenant compte de leur relation de voisinage avec les rangs maxima. Et cela jusqu’à l’identification des clefs de rang 0 se trouvant au bord des occultations.

Le nuage de points tridimensionnels résultant de cet appariement est alors précis — d’une précision dépendant de la bonne localisation des nœuds lors de l’appariement — mais relativement éparse. Ils peuvent servir de supports pour un raffinement de la triangulation ou être utilisés directement suivant les applications. L’utilisation directe donnerait donc une reconstruction relativement grossière de la surface de la scène constituée de triangles.

2.3 Résultats

Cette méthode a été implémentée afin de créer un motif avec des propriétés d'unicité pour chaque point de la scène afin de pouvoir être utilisée lors d'une unique acquisition de la scène. La carte de profondeur obtenue est cependant éparse à cause de la distance entre les nœuds, voir figure IV.3.

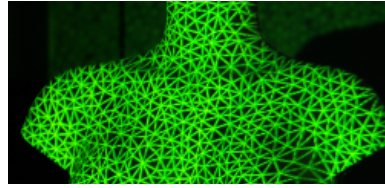


Figure IV.3 – Exemple d'utilisation du motif proposé afin d'obtenir des localisation précise dans une large zone homogène : reconstruction d'une partie de mannequin.

La précision spatiale de chaque point est uniquement liée à la précision de la détection de chaque nœud et à l'estimation des paramètres des appareils de captures d'image. Cette section est donc uniquement centrée sur la propriété d'unicité du motif proposé.

2.3.1 Résultats numériques

Afin de tester l'efficacité de ce motif, plusieurs triangulations aléatoires ont été faites. Afin de ne pas obtenir de triangles trop plats qui nuiraient à la bonne détection de la méthode, seuls des triangulations de Delaunay avec un espacement minimal entre les nœuds ont été utilisées (de l'ordre de 10px pour une triangulation sur une image 10000px × 10000px ; cet espacement est sujet à provoquer de l'instabilité lors de l'extraction et on aura par la suite tendance à l'augmenter). La génération de 100 échantillons a été faite et sur chacun d'eux, les clefs K^0 , K^1 et K^2 ont été calculées. Le nombre de clefs distinctes a aussi été calculé sur chaque échantillon.

Pour 10000 nœuds, il n'y aura qu'environ 13 clefs différentes de rang 0. Il n'y aura aussi que 6040 clefs différentes de rang 1. Pour obtenir une densité de points tridimensionnels suffisante, il semble donc clair qu'il faut à minima aller jusqu'au rang 2. De fait les premiers doublons de clefs de rang 2 apparaissent aux alentours de 1000000 de nœuds, voir table IV.2.

Les observations sur la répartition des degrés nous permettent de statuer que cette dernière est centrée sur 6. Cette assertion est le fruit d'un calcul moyen du nombre de clef de rang 0 de chaque sorte, voir figure IV.4. La probabilité d'avoir un degré de 5 ou 7 est légèrement plus faible que d'avoir 6, mais plus probable que toutes autres possibilités. La courbe peut être approximée par une courbe parabolique dans un espace logarithmique.

$ V $	0-ranked keys	1-ranked keys	2-ranked keys
1000	10.29(1.0%)	889.03(88.9%)	1000.00(100%)
10000	12.88(0.1%)	6039.87(60.4%)	10000.00(100%)
100000	13.62(0.0%)	25293.81(25.3%)	100000.00(100%)
1000000	14.9(0.0%)	55839.01(5.6%)	999999.87(99.9%)

Tableau IV.2 – Pourcentage des différentes clefs calculées par comptage.

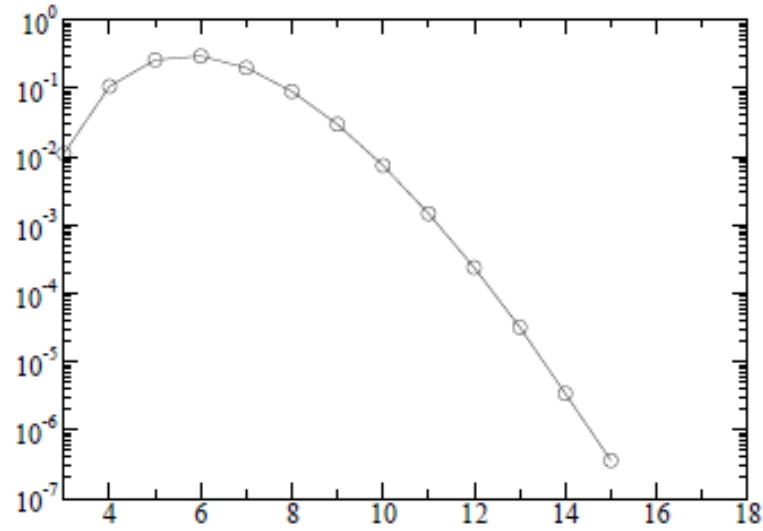


Figure IV.4 – Distribution du degré des nœuds dans une triangulation aléatoire de Delaunay pour 100000 nœuds.

2.3.2 Preuve de la redondance

Afin de construire une preuve de la redondance (ou plutôt du manque de celle-ci) des clefs K^i précédemment décrites, nous proposons d'étudier la probabilité de trouver deux clefs identiques dans une triangulation. Soit \mathbf{P}_i les probabilités suivantes :

$$\mathbf{P}_i = \mathbb{P} \left("K^i(\eta) = K^i(n)", (\eta, n) \in \mathcal{N} \times \mathcal{N} \right) \quad (\text{IV.4})$$

Soit encore \mathbf{E}_0 l'évènement " η et n n'ont pas le même degré". Et soit $\mathbb{D} = \llbracket 0, \deg(\eta) \rrbracket$ pour chaque $(\eta, n) \in \mathcal{N} \times \mathcal{N}$. En notant $+$ l'addition modulo $\deg(\eta)$, on a :

$$\begin{aligned} \mathbf{E}_0: & K^0(\eta) \neq K^0(n) \\ \mathbf{E}_i: & \forall k \in \mathbb{D}, \exists j \in \mathbb{D}, K^{i-1}(\mathcal{V}_j(\eta)) \neq K^{i-1}(\mathcal{V}_{j+k}(n)) \end{aligned} \quad (\text{IV.5})$$

Et donc :

$$\begin{aligned} (K^i(\eta) \neq K^i(n)) &\Leftrightarrow \mathbf{E}_{i-1} \vee (\neg \mathbf{E}_{i-1} \wedge \mathbf{E}_i) \\ &\Rightarrow (\eta \neq n) \end{aligned} \quad (\text{IV.6})$$

Comme la définition de la clef est récursive, K^0 est construite séparément. Il est donc naturel de calculer séparément \mathbf{P}_0 . Pour les autres rangs, ce calcul peut être poussé plus loin en utilisant l'équation IV.6 :

$$\begin{aligned} \forall i \neq 0, \mathbf{P}_i &= \mathbf{P}(\neg(\mathbf{E}_{i-1} \vee (\neg \mathbf{E}_{i-1} \wedge \mathbf{E}_i))) \\ &= \mathbf{P}(\neg \mathbf{E}_{i-1} \wedge \neg \mathbf{E}_i) \end{aligned} \quad (\text{IV.7})$$

Le calcul de \mathbf{P}_0 peut être vue comme un tirage de deux éléments identiques dans un ensemble :

$$\mathbf{P}_0 = \frac{1}{|\mathcal{N}|(|\mathcal{N}| - 1)} \sum_i \mathbf{P}(\deg(\eta) = i) \mathbf{P}(\deg(n) = i | \deg(\eta) = i) \quad (\text{IV.8})$$

Aucun indice ne nous permet d'affirmer que les événements " $\deg(\eta) = i$ " et " $\deg(n) = i$ " sont liés autrement que par le nombre de nœuds de degré i et $|\mathcal{N}|$. De plus, si $|\mathcal{N}|$ est suffisamment grand, la probabilité de tirer un second nœud du même type tend vers la probabilité de tirer le premier. C'est pourquoi pour le reste de cette preuve, nous assumerons que nous travaillons avec un grand nombre de nœuds. C'est pourquoi nous parlerons de probabilité "f.p." pour "fortement probable".

Soit $\pi(d)$ la fonction de distribution de la probabilité qu'un nœud soit de degré d . L'équation IV.8 se simplifie donc ainsi :

$$\begin{aligned} \mathbf{P}_0 &= \frac{1}{\mathcal{N}(\mathcal{N}-1)} \sum_i \mathbf{P}(\deg(\eta) = i) \mathbf{P}(\deg(n) = i) \quad \text{f.p.} \\ &= \frac{1}{|\mathcal{N}|(|\mathcal{N}|-1)} \sum_i (\pi(i)|\mathcal{N}|)^2 \quad \text{f.p.} \end{aligned} \quad (\text{IV.9})$$

D'où :

$$\mathbf{P}_0 \underset{|\mathcal{N}| \rightarrow \infty}{\sim} \sum_i (\pi(i))^2 \quad (\text{IV.10})$$

L'équation IV.10 nous permet d'extraire un majorant — noté \mathbf{P}_i^* — de la probabilité \mathbf{P}_0 . En notant $\Delta_{\deg} = \deg_{\max} - \deg_{\min}$, où \deg_{\max} (respectivement \deg_{\min}) est le degré maximal (respectivement minimal) dans la triangulation, on a :

$$\mathbf{P}_0 < \sum_i (\pi(6))^2 = \Delta_{\deg} (\pi(6))^2 = \mathbf{P}_0^* \quad (\text{IV.11})$$

Maintenant, pour le calcul des \mathbf{P}_i , on obtient à partir de l'équation IV.7 l'équation suivante :

$$\mathbf{P}_i = \mathbf{P}(\neg \mathbf{E}_{i-1}) \mathbf{P}(\neg \mathbf{E}_i | \neg \mathbf{E}_{i-1}) \quad (\text{IV.12})$$

Par définition, il est facile de constater que $\mathbf{P}(\neg \mathbf{E}_i) = \mathbf{P}_i$. Il ne reste donc qu'à calculer l'autre partie :

$$\begin{aligned} \mathbf{P}(\neg \mathbf{E}_i | \neg \mathbf{E}_{i-1}) &= \sum_{k \in \mathbb{D}} \prod_{j \in \mathbb{D}} \mathbf{P} \left(K^{i-1}(\mathcal{V}_j(\eta)) = K^{i-1}(\mathcal{V}_{j+k}(\eta)) \right) \\ &= \sum_{k \in \mathbb{N}} \prod_{j \in \mathbb{N}} \mathbf{P}_{i-1} \end{aligned} \quad (\text{IV.13})$$

La seconde ligne vient du fait que des degrés impossibles entraînent forcément une probabilité nulle. Cela conduit à l'équation suivante :

$$\mathbf{P}_i = \mathbf{P}_{i-1} \sum_{k \in \mathbb{N}} \prod_{j \in \mathbb{N}} \mathbf{P}_{i-1} \quad (\text{IV.14})$$

Cette équation peut être approximée en prenant le centre de la distribution des degrés :

$$\mathbf{P}_i \simeq \mathbf{P}_{i-1} 6 (\mathbf{P}_{i-1})^6 = 6 (\mathbf{P}_{i-1})^7 \quad \text{f.p.} \quad (\text{IV.15})$$

Avec l'aide de la relation IV.15 et de l'équation IV.11, le calcul d'un majorant est possible pour chaque rang. Comme $\pi(6) \simeq 0.2$ et en prenant $\Delta_{\text{deg}} = 15$ ce qui est généralement moindre, il est possible de calculer la table IV.3.

\mathbf{P}_i^*	unicité f.p.
$\mathbf{P}_0 < 0.52$	<i>jamais</i>
$\mathbf{P}_1 < 6.17e - 2$	<i>presque jamais</i>
$\mathbf{P}_2 < 2.04e - 8$	<i>si $\mathcal{N} < 10^6$</i>
$\mathbf{P}_3 < 8.79e - 54$	<i>si $\mathcal{N} < 10^{52}$</i>

Tableau IV.3 – Table des majorants et condition d'unicité f. p.

L'unicité fortement probable calculée correspond bien aux observations obtenues lors de nos tests numériques à la section 2.3.1.

2.4 Conclusion

Il a été prouvé statistiquement qu'il est possible de créer des clefs uniques dans une triangulation. Par ce biais, il est donc possible d'apparier correctement tous les points où la construction d'une telle clef est possible, c'est à dire en tous nœuds d'une triangulation. Pour apparier ces

points, il n'est plus nécessaire de faire une recherche dans l'image mais uniquement dans la liste des clefs extraites. De plus, les relations de voisinage qui existent entre les différents nœuds d'une triangulation peuvent être utilisées pour limiter la recherche à quelques éléments de cette liste.

Si l'utilisation d'un tel motif accélère grandement l'algorithme d'appariement, il est important de noter que puisque les clefs n'existent pas partout, la reconstruction ne peut être qu'éparse. Ceci dit, les résultats du chapitre précédent nous fournissaient aussi un résultat éparse sans aucun contrôle sur la répartition des points bien appariés. Avec ce motif nous pouvons donc contrôler la répartition des points bien appariés ce qui implique forcément une amélioration au niveau de la qualité du maillage surfacique tridimensionnel obtenu.

L'étude de ce type de motif employé pour faire de la reconstruction tridimensionnelle a fait l'objet de publication [\[SGMB13\]](#).

Il reste cependant un problème majeur : comment extraire ce maillage ? Une triangulation de Delaunay, comme utilisée précédemment, peut être définie par ses nœuds. Ainsi définie elle n'est pas forcément unique, car il existe le problème lié à plus de trois points co-circulaires (voir figure IV.5). Elle est donc dans ce cas définie à inversion d'arêtes près.

Bien que la co-cyclicité se produise rarement. Il faut tenir compte du fait que la triangulation de Delaunay a été projetée sur un volume. Les distances entre les nœuds ont donc été altérées lors de la projection. Nous ne pouvons même pas être assurés que la projection acquise est toujours une triangulation respectant le critère de Delaunay. De part ce fait, en partant uniquement des nœuds, nous ne pouvons aisément retrouver les arêtes car le critère de distance minimal n'est plus respecté.

La recherche des nœuds ne peut donc pas être le seul support à son extraction. Il est donc nécessaire de déterminer aussi ses contours.

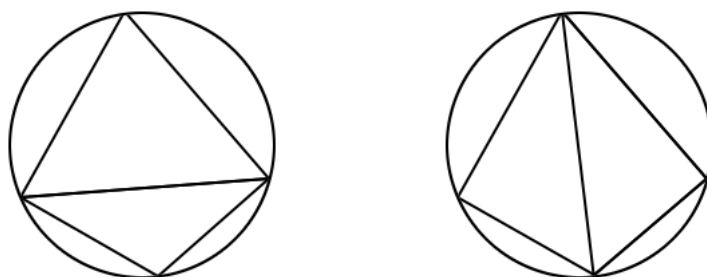


Figure IV.5 – Cas de points répartis de façon cocirculaire empêchant une unicité de la triangulation de Delaunay. Les deux triangulations possibles représentées ici sont donc deux triangulations de Delaunay distinctes du même nuage de points.

La section suivante est consacrée au détecteur de contour élaboré à partir du modèle utilisé pour représenter les images développées dans le chapitre précédent. Cette méthode nous a été nécessaire afin de pouvoir extraire le motif.

3 Détecteur de contour élaboré pour l'extraction d'un motif de type maillage

Nous nous sommes intéressés à regarder l'image comme une surface tridimensionnelle afin d'en extraire des particularités. Nous avons effectivement fait ressortir des propriétés intéressantes discriminant les zones homogènes des zones non homogènes et portant suffisamment d'information pour faire de la discrimination dans ces zones.

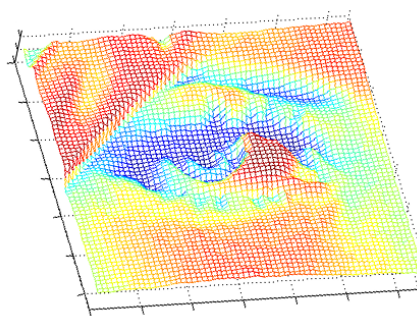
3.1 Image 2D vue comme un assemblage de surface 3D

Il existe de nombreux algorithmes de détection de contours qui peuvent être utilisés afin d'extraire le motif. Toutefois, la modélisation de l'acquisition photographique utilisée nous permet d'en construire un qui permette aussi de décrire le comportement desdits contours.

Il n'est pas question ici de faire un état de l'art sur les méthodes de détections de contours. Le sujet est en effet très bien traité dans de nombreux ouvrages et ce n'est pas le but de la présente thèse. On peut citer entre autre parmi ces états de l'art sur le sujet : [JKS95, Chap.5], [BGM⁺04, Chap.11] et [SR09].



(a) Zoom sur l'œil de Lena.



(b) Zoom sur l'œil de Lena affichée comme une surface.



(c) Zoom sur l'approximation de l'œil de Lena utilisant le modèle de régression de surface proposée.

Figure IV.6 – Illustration du modèle de représentation proposé. L'œil de *Lena* est montré à gauche et son approximation utilisant le modèle de représentation polynomial bidimensionnel est illustré à droite.

Le modèle de la scène $S(x, y)$ utilisé est inspiré de [MS89, CR12]. La scène y est alors supposée composée de N_S objets. La radiance de la scène peut donc être présentée comme une surface continue par partie :

$$S(x, y) = \sum_{i=1}^{N_S} S_i(x, y) \mathbb{1}_{\mathcal{X}_i}(x, y) \quad (\text{IV.16})$$

où $S_i(x, y)$ sont des fonctions continues sur \mathcal{X} , $\mathbb{1}_A$ est la fonction indicatrice de l'ensemble $A \subset \mathbb{R}^2$, $\{\mathcal{X}_i\}$ est une partition finie de \mathcal{X} dans laquelle chaque élément \mathcal{X}_i est connecté avec un intérieur non-vide $\overset{\circ}{\mathcal{X}}_i$ et une frontière $\partial\mathcal{X}_i$. L'ensemble discontinu $S(x, y)$ est noté $\bigcup_{i=1}^{N_S} \partial\mathcal{X}_i$. Soit Γ une courbe paramétrique incluse dans cet ensemble. Cela signifie que l'ensemble discontinu Γ peut être considéré comme un ensemble continu par partie — notons ces parties γ_i — de fonction $\Gamma : T \mapsto \mathcal{X}$, où T est un intervalle compact de \mathbb{R} qui admet la représentation suivante :

$$\Gamma(t) = \sum_{i=1}^{N_C} \gamma_i(t) \mathbb{1}_{T_i}(t) \quad (\text{IV.17})$$

Ici $\{T_i\}$ est une partition finie de T et chaque T_i est un segment non-vide. Plus simplement, cela signifie que l'ensemble des discontinuités de $S(x, y)$ peut être représenté comme la juxtaposition d'un nombre fini de segments.

Afin de simplifier la formulation mathématique, supposons que les ensembles $\{\gamma_i(t), t \in T_i\}$ sont séparables. Le modèle de scène décrit par les équations IV.16 et IV.17 et suffisamment général pour s'adapter à un grand nombre de scènes naturelles (voir [MS89]).

La scène est ensuite acquise par le capteur d'image comme vu au chapitre précédent, équations III.8 et III.25.

Mumford et Sha [MS89] ont souligné certaines propriétés physiques sur l'émission de la lumière par les objets solides :

- P1 :** *Sur chaque domaine \mathcal{X}_i , associé à l'objet N_i , la variance émise varie faiblement.*
- P2 :** *La radiance est discontinue lorsqu'elle franchie les frontières γ_i .*
- P3 :** *Les objets sont de formes régulières dans le sens où chaque courbe $\gamma_{i,j}$ est régulière, ou $\gamma_i \in C^2(T_i)$.*

Ces propriétés concordent avec le modèle proposé. Il peut être prouvé que la radiance de la scène S peut être localement représentées dans un voisinage \mathcal{V}_{x_0, y_0} de $(x_0, y_0) \in \mathcal{X}$, comme montré dans [CZF⁺11, CZR⁺12] avec quelques suppositions faibles :

$$\forall (x, y) \in \mathcal{V}_{x_0, y_0}, S(x, y) = S_c(x, y) + S_d(x, y) \quad (\text{IV.18})$$

où S_c est une fonction continue, $S_d(x, y)$ est une fonction échelon (*Heaviside*) en deux dimensions qui représente la discontinuité de la scène. On a donc en supposant qu'il n'y ait que deux objets visibles dans ce voisinage :

$$S_d(x, y) = d(x, y) \mathbb{1}_{(x, y) \in \mathcal{X}_j} \quad (\text{IV.19})$$

Ici, $d(x, y)$ est l'intensité locale de la discontinuité de la radiance entre les domaines \mathcal{X}_i et \mathcal{X}_j . Ceci se généralise bien sur à plusieurs objets mais l'écriture y est moins évidente.

En appliquant l'équation III.25 à la décomposition proposée, la formulation est la suivante (le produit de convolution étant distributif) :

$$\forall (x, y) \in \mathcal{V}_{x_0, y_0}, I(x, y) = I_c(x, y) + I_d(x, y) \quad (\text{IV.20})$$

où I_c et I_d représentent respectivement la convolution des parties continues et discontinues. Les pixels chauds et morts ajoutant le bruit impulsionnel IMP n'interviennent pas dans cette opération. Cela vient du fait que le capteur a une amplitude définie et les valeurs capturées aux positions des pixels chauds et morts dépassent alors l'amplitude autorisée par le capteur. Nous pouvons donc considérer que la matrice IMP ne comprenne que des éléments nuls (pour les pixels normaux), de valeur $+\infty$ (pour les pixels chauds) ou de valeur $-\infty$ (pour les pixels morts). Il n'est donc pas aberrant de considérer que ce bruit peut donc être réparti sur les deux images. Nous avons donc $I_c = (PS_c) * \text{PSF} + \text{IMP}$ et $I_d = (PS_d) * \text{PSF} + \text{IMP}$.

D'après les propriétés P1–P3 et les équations IV.16–IV.18, la partie continue $I_c(x, y)$ devrait varier faiblement. Qui plus est, le gradient et l'Hessien de I_c devrait avoir une faible magnitude. Au contraire, la partie discontinue de la radiance émise par la scène est une discontinuité rendue floue par le processus d'acquisition (à cause du PSF). C'est pourquoi, le long de la courbe de rupture $\{\gamma_i\}$, le gradient et l'Hessien de I_d devrait avoir une forte magnitude. Il est à noter que S_d (IV.19) est constante en dehors des discontinuités, ce qui implique que I_d peut être considérée constante en dehors des contours et des zones proches (à cause du PSF).

Il est cependant difficile d'exploiter directement cette modélisation du fait de l'impossibilité d'extraire les fonctions représentant la scène. Toutefois, il est possible de les approximer localement. A cette fin, l'image est regardée comme une surface que l'on va chercher à approximer par une surface polynomiale. La régression polynomiale a été choisie afin de plus facilement pouvoir effectuer la décomposition du modèle précédemment exposé.

3.2 Construction d'un détecteur de contour associé

Le détecteur associé à ce modèle — que nous avons appelé RSD^2 dans nos publications — repose sur l'approximation locale de la surface par une régression polynomiale. Toutes les valeurs des pixels dans le voisinage du point où s'effectue le calcul sont donc utilisées. Les coefficients obtenus sont ensuite analysés pour déterminer si le pixel considéré est situé proche d'un contour, et déterminer la configuration de celui-ci le cas échéant.

3.2.1 Implémentation

N^2 equations doivent être définies en utilisant les intensités $I(x_0 + \bar{x}_n, y_0 + \bar{y}_n)$ de chaque pixels de coordonnées relatives (\bar{x}_n, \bar{y}_n) dans le voisinage \mathcal{V}_{x_0, y_0} afin d'effectuer une régression de surface par un polynôme bivarié de degré K . Nous devons donc résoudre le système surcontraint suivant :

$$A_K C_K = B \quad (\text{IV.21})$$

où A_K est une matrice de taille $N^2 \times (K + \frac{K(K+1)}{2})$ encodant les coordonnées relatives pour la régression, B est le vecteur colonne des N^2 composants représentant les intensités des pixels du voisinage lus dans l'ordre lexicographique et $C_K = (c_0 \ c_1 \ \cdots \ c_{K+\frac{K(K+1)}{2}})^T$ représente les coefficients à déterminer de la surface polynomiale. Comme le système IV.21 est surcontraint, il est proposé de le résoudre par la méthode des moindres carrés.

$$\min \|A_K C_K - B\|_2^2 \quad (\text{IV.22})$$

En prenant par exemple une quadrique, le système serait :

$$\min \|A_2 C_2 - B\|_2^2$$

avec

$$A_2 = \begin{pmatrix} 1 & \bar{x}_1 & \bar{y}_1 & \bar{x}_1^2 & \bar{y}_1^2 & \bar{x}_1 \bar{y}_1 \\ 1 & \bar{x}_1 & \bar{y}_2 & \bar{x}_1^2 & \bar{y}_2^2 & \bar{x}_1 \bar{y}_2 \\ \vdots & & & \vdots & & \vdots \\ 1 & \bar{x}_N & \bar{y}_{N-1} & \bar{x}_N^2 & \bar{y}_{N-1}^2 & \bar{x}_N \bar{y}_{N-1} \\ 1 & \bar{x}_N & \bar{y}_N & \bar{x}_N^2 & \bar{y}_N^2 & \bar{x}_N \bar{y}_N \end{pmatrix} \quad (\text{IV.23})$$

$$C_2 = (c_0 \ c_1 \ c_2 \ c_3 \ c_4 \ c_5)^T$$

$$B = \left(I(x_0 + \bar{x}_1, y_0 + \bar{y}_1) \quad \cdots \quad I(x_0 + \bar{x}_N, y_0 + \bar{y}_N) \right)^T$$

La matrice A_K est indépendante du point considéré du fait qu'elle est construite à partir de coordonnées relatives. Donc $A_K^T A_K$ est aussi constante et ne peut être affectée par le bruit. Si ce n'est pas une matrice singulière, l'équation IV.22 possède une unique solution qui peut être trouvée en résolvant le problème équivalent :

$$\min \|A_K^T A_K C_K - A_K^T B\|_2^2 \quad (\text{IV.24})$$

En fait, A_K est plein rang colonne car une base pour les polynômes bivariés peut être retrouvée en ligne. Donc la solution de l'équation IV.22 est unique et $A_K^T A_K$ est inversible. Ce calcul nous mène, grâce au théorème de Gauss-Markov, à l'estimateur BLUE³ [Rao73] des coefficients C_K :

$$C_K = \left(A_K^T A_K \right)^{-1} A_K^T B \quad (\text{IV.25})$$

La matrice $A_K^T A_K$ étant mal conditionnée ($\text{Cond}_2(A_K^T A_K) > 600$ pour $N = 3$ par exemple), des instabilités numériques peuvent être engendrée lors de l'approximation, la valeur du conditionnement représentant l'erreur relative potentielle :

$$\frac{\|\Delta C_K\|}{\|C_K\|} \leq \text{Cond}(A_K^T A_K) \frac{\|A_K^T \Delta B\|}{\|A_K^T B\|}$$

L'inversion utilisant la décomposition QR est alors préférable car plus stable numériquement [Par94, chap1.13]. Une version modifiée de la factorisation QR de Gram-Schmidt [LG96, chap5.2.8] a ici été utilisée : $A_K = Q_K(R_K^T | \emptyset)^T$ avec \emptyset une matrice nulle utilisée pour obtenir le bon nombre de ligne. Notons que R_K^T est une matrice carrée. La solution IV.25) peut être reformulée en utilisant l'orthogonalité de Q :

$$C_K = (R_K^{-1} | \emptyset) Q_K^T B \quad (\text{IV.26})$$

Donc $(R_K^{-1} | \emptyset) Q_K^T$ n'a besoin d'être calculé qu'une unique fois afin de trouver les coefficients de la régression dans les voisinages \mathcal{V}_{x_0, y_0} de chaque pixel (x_0, y_0) de l'image. Ceci revient alors à effectuer une convolution de l'image par le noyau $(R_K^{-1} | \emptyset) Q_K^T$.

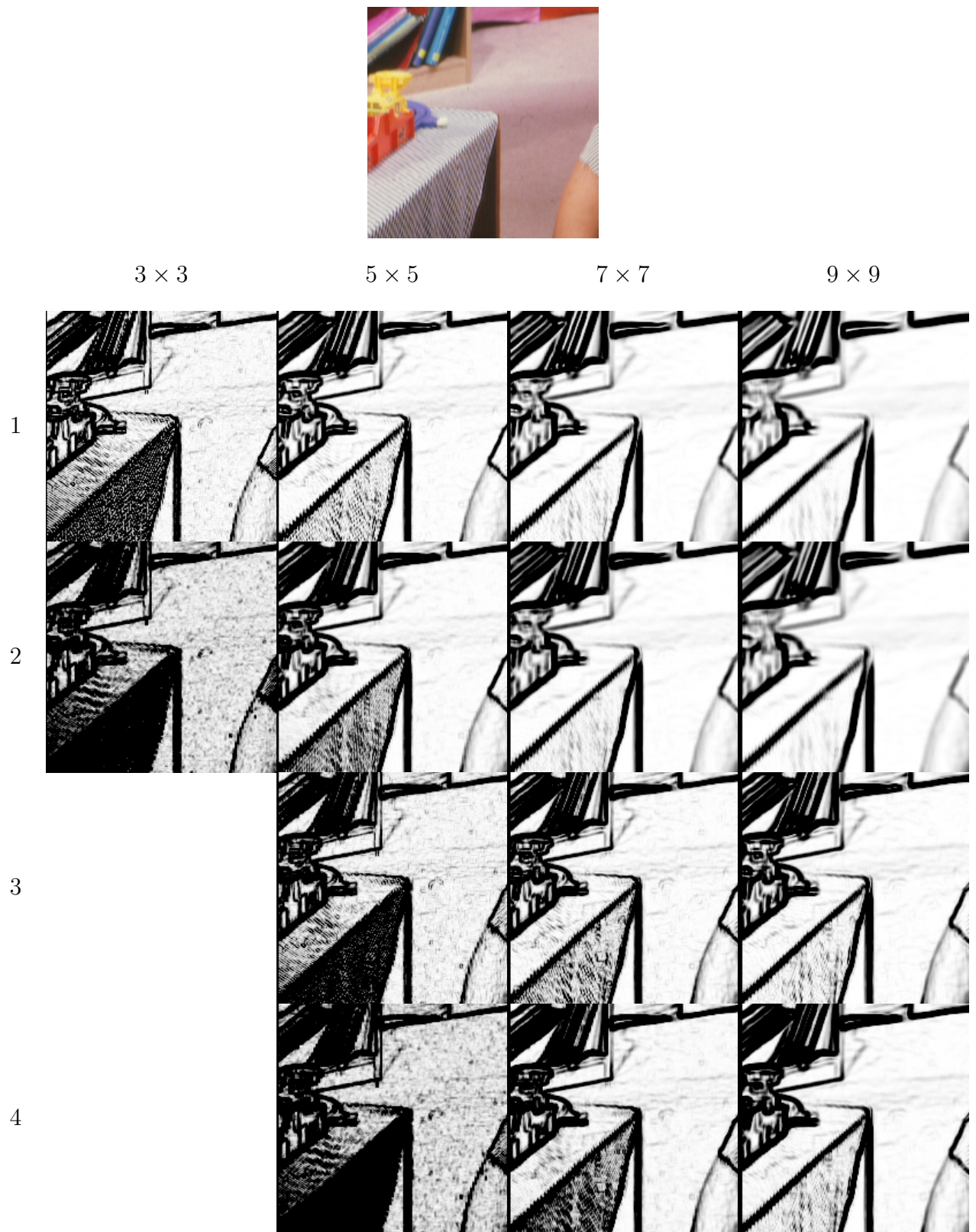


Figure IV.7 – Cartes de contours résultant de la détection sur l'image *Barbara*. Les lignes correspondent à l'ordre de la régression et les colonnes à la taille de voisinage utilisé. Les résultats sont soumis à l'équation [IV.29](#) pour un meilleur affichage.

3.2.2 Détection de la présence d'un contour

Les coefficients de la régression sont un bon indicateur de la présence de contours. En effet, un lien pourrait être fait entre les coefficients de la régression et les coefficients du développement en série de Taylor de la surface si nous connaissions les equations représentant la scène.

Il est à noter que l'orientation et la courbure des contours sont grandement influencés par ces coefficients de Taylor. Nous pouvons par exemple citer les travaux d'Haralick [Har84] dans lesquelles ces coefficients sont utilisés afin d'estimer l'orientation des contours puis de calculer la position du changement de signe de la dérivée seconde. Il est en fait possible d'utiliser directement ces coefficients comme nous allons le voir.

Considérons une scène comportant deux objets solides distincts avec une radiance constante émise par eux. L'image de cette scène représente deux objets solides homogènes en intensité séparés par une courbe. Considérons le cas où cette courbe est l'équation $x = x_0$, voir Figure IV.8. De plus, si l'on retire le bruit lors de l'acquisition, nous pouvons supposer que la fonction PSF est une gaussienne bidimensionnelle :

$$PSF(u, v; s(x, y)) = \frac{1}{2\pi s^2(x, y)} \exp\left(-\frac{u^2 + v^2}{2s^2(x, y)}\right) \quad (\text{IV.27})$$

Ici $s(x, y) > 0$ et le paramètre de flou local au point considéré (x, y) . Il est à noter que $s(x, y)$ varie très faiblement. C'est pourquoi il est possible de considérer sa dérivée comme localement négligeable.

En utilisant les propriétés P1-P3, un calcul direct permet d'écrire l'expression de la scène imagée :

$$I(x, y) = s_1 + (s_2 - s_1)\Phi\left(\frac{x - x_0}{s(x_0, y)}\right)$$

avec s_1 et s_2 représentant respectivement la radiance sur les domaines \mathcal{X}_1 et \mathcal{X}_2 , et $\Phi(x) = (2\pi s^2(x, y))^{-1/2} \int_{-\infty}^x \exp(-u^2/(2s^2(x, y))) du$ représentant l'intégrale de la fonction gaussienne associée au flou de paramètre $s(x_0, y)$. C'est pourquoi, l'expansion de Taylor de I au point (x, y) permet d'écrire l'expression suivante :

$$\begin{aligned} I(x+\eta, y) \approx & s_1 + (s_2 - s_1)\Phi\left(\frac{x - x_0}{s(x, y)}\right) + \eta (s_2 - s_1) \frac{1}{\sqrt{2\pi} s(x, y)} \exp\left(-\frac{1}{2} \frac{(x - x_0)^2}{s^2(x, y)}\right) \\ & - \eta^2 \frac{(x - x_0)(s_2 - s_1)}{2\sqrt{2\pi} s^3(x, y)} \exp\left(-\frac{1}{2} \frac{(x - x_0)^2}{s^2(x, y)}\right) \end{aligned} \quad (\text{IV.28})$$

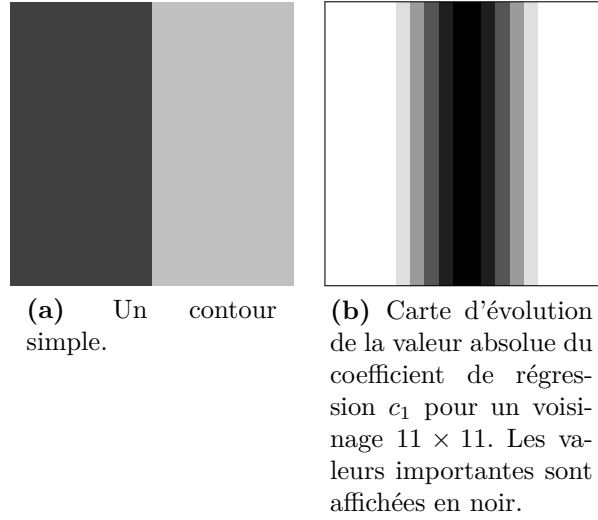


Figure IV.8 – Exemple trivial de contour

Cette dernière équation met en valeur le fait que lorsque $|x - x_0|$ croît, les coefficients de la série de Taylor décroissent rapidement au contraire du coefficient du premier ordre qui atteint son maximum en $x = x_0$. De même, l'équation IV.28 montre que les coefficients de la série dépendent linéairement de $(s_2 - s_1)$ qui représente l'intensité de la discontinuité. Notons finalement que les coefficients du polynôme décroissent linéairement à la vitesse de $s(x, y)^{-p}$, ce qui montre une conséquence directe du flou gaussien. Plus le flou est important, plus les coefficients sont faibles ce qui est lié au fait que les contours deviennent des variations moins marquées.

Lorsque la surface de l'image est approximée par une régression polynomiale bidimensionnelle, il faut chercher le coefficient de régression correspondant à la direction choisie pour effectuer le développement en série de Taylor. Pour l'exemple choisi, il s'agit de c_1 . De façon générale, lors de l'utilisation concrète du détecteur *RSD*, tous les coefficients sont calculés. L'analyse de l'amplitude des coefficients permet ainsi de déterminer la meilleure direction pour le développement en série de Taylor et donc de directement connaître l'orientation du contour. Il n'est donc pas nécessaire de réaliser cette étape pour une autre raison que celle de comprendre comment les coefficients sont liés à la surface.

De plus, le processus de flou affecte aussi la partie continue I_c de la même façon. Il est donc possible de la modéliser localement de façon précise en utilisant la régression proposée, et donc de modéliser localement toute l'image, voir figure IV.6.

Chaque orientation et forme de contours est donc une combinaison différente des coefficients c_k , excepté c_0 qui correspond à l'élément d'ordre 0 de la régression. Dans le modèle utilisé, ce

coefficient peut être vu comme l'influence de la partie continue de l'équation IV.18 qui apparaît en tant que I_c dans l'équation IV.20. Cela implique que le coefficient c_0 ne peut être utilisé. Tous les autres coefficients ont par contre une valeur significative en présence de contour.

La façon la plus simple de calculer la carte des contours sachant cela est de calculer la norme des coefficients c_k ($k > 0$). Dans notre implémentation, nous avons utilisé la norme euclidienne élevé au carré.

Dans le cas d'une image couleur, il faut calculer les coefficients de régression sur chaque canal séparément. Ces coefficients (excepté c_0) peuvent être directement fusionnés en calculant la norme du vecteur comportant tous les coefficients de tous les canaux, voir figure IV.7 pour un exemple couleur avec différents paramètres.

La moyenne implicite effectuée dans le calcul des coefficients par l'estimation au sens des moindres carrés IV.21 permet de filtrer les hautes fréquences, et cela d'autant plus que la taille du voisinage utilisé est grande. Cela permet de ne pas détecter de contours dans les zones texturées, voir figure IV.7.

Au contraire, accroître l'ordre de la surface de régression permet d'obtenir une surface mieux adaptée à la surface réelle. Cela permet donc de détecter plus de détails et notamment de contours sur la carte de contour. Ces deux paramètres (taille de fenêtre N et ordre de régression K) ne sont pas liés. La carte de contour calculée sur l'image *Barbara* (figure IV.7) permet de mettre en évidence en regardant les éléments diagonaux du tableau que le résultat change. Augmenter les deux paramètres permet de filtrer les hautes fréquences (et donc les textures) tout en restant suffisamment adaptable pour déterminer les contours principaux de l'image.

Pour des questions d'affichage, les résultats exposés aux figures IV.7 - IV.10 subissent une inversion de couleur seuillée :

$$\begin{cases} 255 - \lfloor \text{Norme}^2 \rfloor & \text{lorsque } \text{Norme}^2 < 255 \\ 0 & \text{sinon} \end{cases} \quad (\text{IV.29})$$

3.2.3 Descripteur de contours

Si les coefficients de régression (hors c_0) sont tous significativement proche de 0, la zone décrite est une zone homogène qui ne contient pas de contour. En pratique, nous nous servons donc juste de ce fait dans la détection de notre motif lumineux. Il est toutefois intéressant de noter qu'il est possible d'utiliser le détecteur présenté pour déterminer des contours de formes ou d'orientation particulière. Dans cette partie nous développerons cette approche en décrivant l'impact de chaque coefficient. De par cette étude, nous pourrions ainsi aisément déterminer

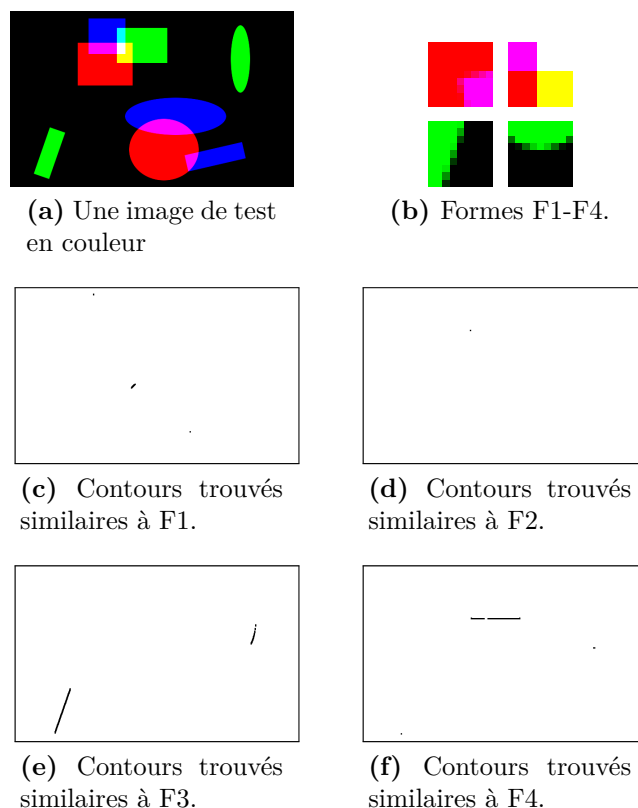


Figure IV.9 – Résultat de la méthode de reconnaissance de forme proposée élargi par une dilatation morphologique pour faciliter la visualisation.

l'orientation de chaque contour détecté.

L'analyse des coefficients de régression (hors c_0) permet de déterminer les circonstances dans lesquelles ils se retrouvent non-nuls. Les coefficients réagissent différemment selon la direction et la courbure du contour au point considéré. Ils tendent à être tous très faibles en l'absence de contour. Il est donc possible de décrire les points de contours grâce à eux.

Il est aussi possible de rechercher les contours similaires à une forme de contour donnée. Pour retrouver les contours similaires à une référence de coefficients C_K^* , il suffit de calculer la distance entre les deux : $\|C_K - C_K^*\|_2$. Les zones de correspondance sont alors celles où cette distance est faible. Des exemples de résultats sont exposés par les figures IV.9 et IV.10.

La figure IV.11 montre la réponse des coefficients sur une série de cas types. Seuls le signe et les valeurs proches de zéro ont besoin d'être considérés afin de décrire ces formes. La valeur de chaque coefficient dépend de l'amplitude de la rupture d'intensité entre les deux objets. C'est pourquoi le symbole " \neq " est utilisé car le signe peut changer lors d'un décalage d'un pixel. De plus ce symbole prend en compte un éloignement significatif de la valeur 0. Ces résultats

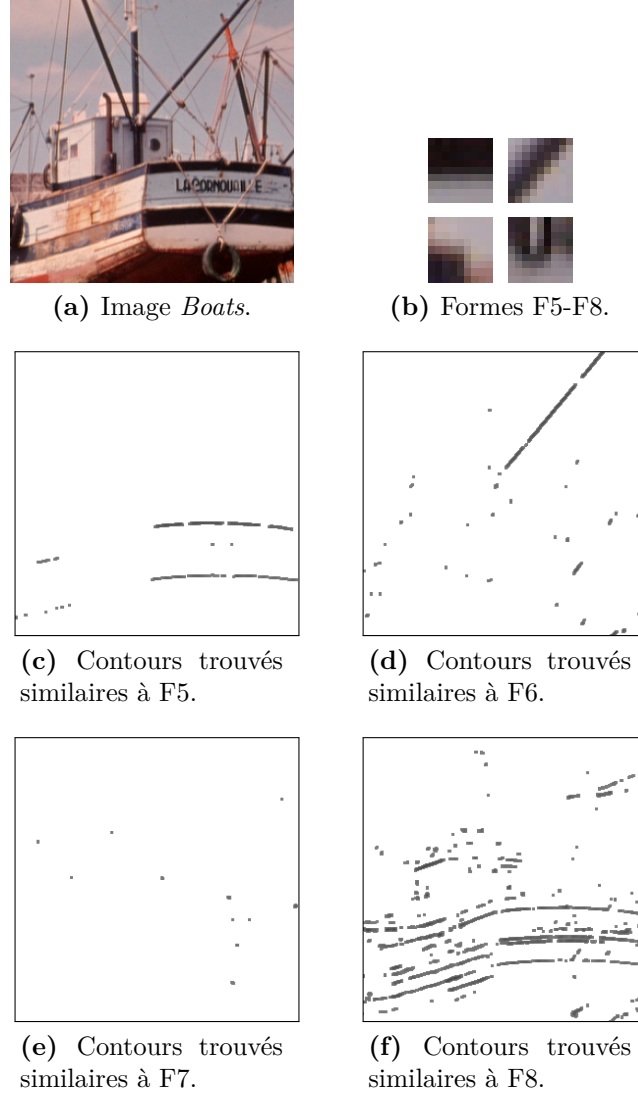


Figure IV.10 – Résultat de la méthode de reconnaissance de forme proposée élargi par une dilatation morphologique pour faciliter la visualisation.

peuvent être retrouvés par l'analyse des coefficients de Taylor ou directement en regardant les coefficients de la surface de régression.

Nous avons cependant constaté que les agencements complexes de contours mènent à de mauvaises détections, surtout dans le cas où l'ordre de régression n'est pas suffisant. Il est toutefois impossible de le fixer sans avoir d'indices sur l'image à analyser.

Les détecteurs de coins de Harris [HS88a] et de Shi-Tomasi [ST94a] utilisent une expansion de Taylor des différences élevées au carré. La matrice de Harris peut être retrouvée dans $A_K^T A_K$ sur la diagonale (avec un certain nombre d'informations additionnelles qui peuvent être consi-

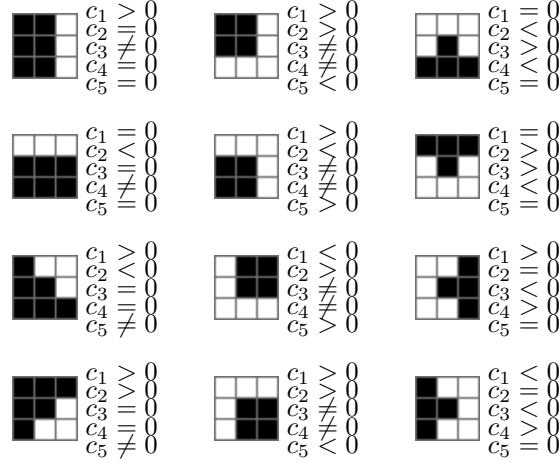


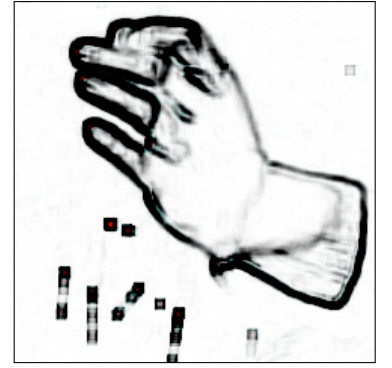
Figure IV.11 – Analyse de la variation des coefficients sur des exemples types.



(a) *La leçon d'anatomie*



(b) Carte de contour



(c) Zoom sur la main

Figure IV.12 – Évaluation visuelle du détecteur de contours sur *La leçon d'anatomie* de Rembrandt. Dans l'image de droite, un zoom sur la main nous permet d'observer en rouge la relative efficacité du détecteur de coins induit.

dérées ici comme du bruit) lors du calcul de la solution [IV.24](#). Les coefficients c_3 , c_4 et c_5 sont calculés en utilisant ces parties de $A_K^T A_K$. En pratique, ces coefficients réagissent fortement en présence de coins dans le voisinage. En remplaçant respectivement les éléments A , B et C de la matrice de Harris [[HS88a](#)] par les coefficients c_3 , c_4 et c_5 , il est possible d'appliquer la méthode de Harris pour déterminer la présence de coins. Des exemples de détecteurs de coins utilisant la méthode de la distance — ce qui revient à faire une comparaison par patch — peuvent être vus aux figures [IV.9c](#), [IV.9d](#) et [IV.10e](#). Un exemple utilisant la méthode de Harris avec les coefficients de la régression peut être observé sur la figure [IV.12c](#) (les points rouges).

3.3 Évaluation du détecteur

Il est difficile d'évaluer la précision d'un détecteur de contours avec pertinence. La précision et la robustesse doivent être pris en compte. C'est la comparaison la plus significative qui puisse être faite entre algorithmes. Ce problème a été étudié en profondeur dans [AP79, HSSB96].

Il a été choisi de l'évaluer par rapport à d'autres détecteurs, notamment en terme de robustesse face au bruit et face au flou. A cette fin, des images digitales ont été conçues afin d'obtenir une vérité terrain (carte des contours) non biaisée. La précision et la robustesse du détecteur sont calculées à partir de multiples échantillons pour plusieurs niveaux de bruit et de flou.

Les résultats obtenus sont ensuite analysés et comparés à ce qui est attendu au vu du modèle de représentation de la scène utilisée pour la méthode proposée.

3.3.1 Protocole expérimental

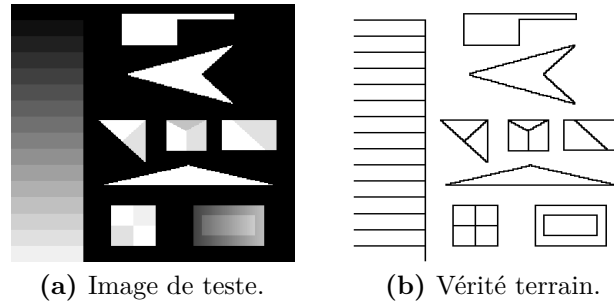


Figure IV.13 – Image de test utilisée pour l'évaluation du détecteur de contours et la carte de ses contours.

Afin d'évaluer le détecteur de contours résultant de la modélisation de la scène et du capteur utilisé, une étude comparative avec d'autres détecteurs a été faite. L'étude se porte notamment sur la robustesse face au bruit et face au flou.

Les détecteurs choisis comptent parmi les plus utilisés (Canny [Can86b], LoG⁴ [HM86]) et parmi les plus performants (Filtrage NL [LT10]). La méthode d'Haralick [Har84] y a aussi été ajoutée car c'est sans conteste la méthode qui se rapproche le plus de celle exposée. Elle diffère de celle exposée ci-dessus par le fait qu'elle utilise les coefficients polynomiaux dans sa recherche des changements de signe de dérivée directionnelle seconde. Certains de ces détecteurs nécessitent un prétraitement pour atténuer le bruit comme les filtres NL.

Afin de modéliser les résultats, nous avons effectué une simulation de Monte-Carlo avec 100 échantillons pour chaque algorithme et chaque niveau de bruit ou de flou. Sur chacune de ces

4. *Laplacian of Gaussian*

réalisations, un seuillage a été fait. Le seuil a été choisi de tel sorte que le nombre de pixel étiquetés “contour” correspond au nombre de pixels appartenant effectivement à des contours sur la vérité terrain. Il est à noter que de cette manière le nombre de pixels étiquetés “contour” alors qu’ils n’en sont pas est le même que le nombre de pixel étiquetés “non contour” alors qu’ils en sont. C’est pourquoi nous parlerons dans nos résultats de pixels mal détectés. Le pourcentage de pixels mal détectés est ensuite calculé en faisant la moyenne des simulations. Pour le cas du flou, un bruit a été ajouté après l’application du flou avec un paramètre $\sigma = 4$. Ces résultats sont exposés dans la table IV.4.

La seconde expérience menée consiste à calculer l’écart d’imprécision. Un bon détecteur doit non seulement bien détecter mais aussi bien localiser les résultats. A cette fin, nous avons calculé les courbes d’une FoM^5 [AP79] qui prend en compte la précision (bonne localisation) de la détection, voir aussi [JKS95, Chap.5.8.2] et [Pra07, Chap.15.5]. Elle se définit ainsi :

$$FoM = \frac{1}{\max(|I_E|, |D_E|)} \sum_{(x,y) \in D_E} \frac{1}{1 + \alpha d(x,y)} \quad (\text{IV.30})$$

où I_E et D_E représentent respectivement le contour “idéal” (vérité terrain) et le contour détecté. En utilisant le cardinal de ces ensembles et la distance $d(x,y)$, $(x,y) \in D_E$ entre le contour détecté en (x,y) et le contour idéal le plus proche, la FoM a été réalisée. Le paramètre α est un poids permettant de prendre plus ou moins en compte la distance dans la FoM . Plus le résultat est élevé, plus le détecteur a une bonne précision. La FoM permet donc de comparer les méthodes entre elles avec le critère de bonne détection et celui de bonne localisation qui sont tous les deux décrit dans [Dem02].

Le calcul direct de la FoM est complexe lorsque la vérité terrain comporte de nombreux contours avec une géométrie compliquée. C’est pourquoi nous avons utilisé des images de test simplifiées pour son calcul, voir figure IV.14. La différence d’intensité utilisée dans la construction de ces images de test est de 8 et ces images ont subi un flou et un bruit avant analyse. Si la vérité terrain a des contours de la largeur de deux pixels, cela est dû au fait que le contour réel est en réalité entre les deux pixels de l’épaisseur. Selon la définition du détecteur utilisé, il peut soit trouver la frontière sous-pixel, soit considéré les pixels d’un côté de la frontière, soit ceux de l’autre, ou encore dans le cas d’une définition symétrique, les deux côtés.

Une fois encore, une simulation de Monte-Carlo avec 100 échantillons pour chaque flou et variance de bruit a été effectuée. Pour la FoM avec un bruit variant, un flou constant a été appliqué avec le paramètre $s = 2$, voir figure IV.16. Pour la FoM de flou variant, un bruit gaussien

5. *Figure of Merit*

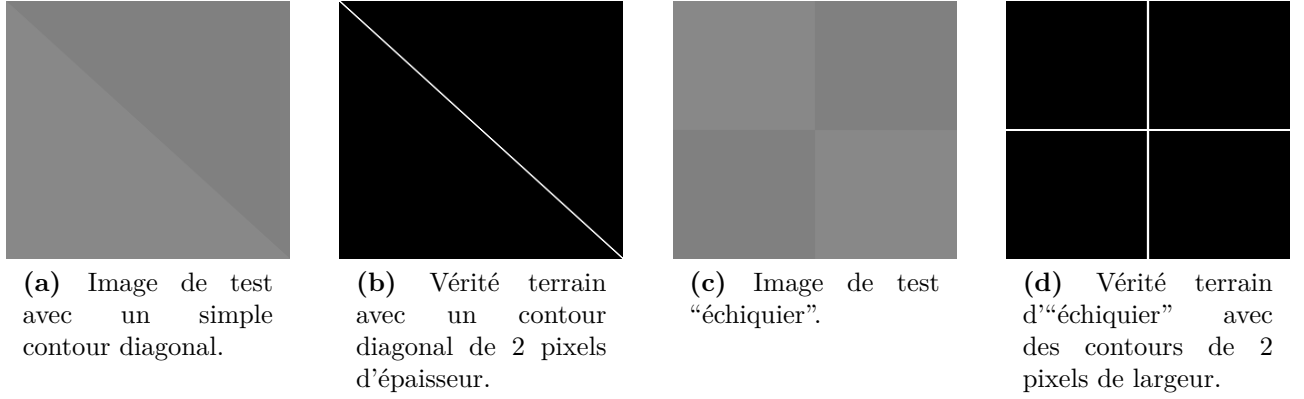


Figure IV.14 – Les images de test (avant bruit et flou) utilisées pour évaluer la localisation des contours dans la construction de la FoM.

	Rayon de flou	$s=1$	$s=2$	$s=4$	$s=7$	$s=10$
Détecteurs	PSNR	25.80	21.52	18.83	16.41	14.81
Canny [Can86b] (échelle=1)		40.36%	59.55%	66.90%	78.37%	81.95%
Canny [Can86b] (échelle=2)		36.44%	60.81%	70.79%	86.37%	89.04%
LoG [HM86] (échelle=1)		12.94%	51.09%	57.82%	56.40%	54.90%
LoG [HM86] (échelle=2)		24.12%	64.91%	77.06%	80.35%	84.02%
Filtre NL [LT10]		19.61%	43.65%	56.86%	74.78%	81.05%
Filtre NL [LT10] *		4.26%	24.85%	43.11%	56.52%	63.24%
Haralick Facet [Har84]		54.68%	73.82%	80.44%	98.80%	99.95%
Haralick Facet [Har84] *		<i>55.52%</i>	<i>75.29%</i>	<i>80.58%</i>	<i>98.85%</i>	<i>99.96%</i>
Méthode proposée α		0.05%	21.39%	46.42%	58.61%	68.88%
Méthode proposée β		5.15%	28.44%	46.46%	58.76%	65.12%

Tableau IV.4 – Comparaison de la robustesse des détecteurs face à un flou gaussien. Le pourcentage P_{MD} exposé est le pourcentage de détection manquées. Les meilleurs résultats sont montrés en bleu en gras, et les plus mauvais en rouge en italique. * : un pré-filtrage de type médian a été effectué. α : (3×3 , ordre 2). β : (5×5 , ordre 2)

constant a été appliqué avec le paramètre $\sigma = 2$, voir figure IV.15. Pour les méthodes nécessitant une régularisation un filtre médian a été appliqué comme lors de l'expérience précédente.

Enfin une dernière expérience à été menée afin de montrer l'impact des paramètres N et K . A cette fin, nous avons construit la FoM avec différents paramètres K et N en présence de flou IV.18 et en présence de bruit IV.17, le but étant de déterminer le meilleur paramétrage possible.

En conclusion, plusieurs échantillons visuels appuieront les résultats exposés.

Corrélation appuyée sur un maillage

Détecteurs	Déviatiion standard PSNR	$\sigma=3$ 40.37	$\sigma=4$ 37.88	$\sigma=6$ 34.36	$\sigma=8$ 31.87	$\sigma=10$ 29.96
Canny [Can86b] (échelle=1)		8.760%	8.789%	8.989%	9.550%	10.417%
Canny [Can86b] (échelle=2)		0.142%	0.403%	1.254%	6.109%	11.62%
LoG [HM86] (échelle=1)		35.024%	36.438%	37.401%	37.784%	38.033%
LoG [HM86] (échelle=2)		3.758%	6.161%	14.64%	25.51%	35.18%
Filtrage NL [LT10]		0.958%	6.139%	35.43%	62.94%	82.47%
Filtrage NL [LT10] *		39.682%	43.831%	48.525%	50.978%	52.544%
Haralick Facet [Har84]		11.364%	11.561%	11.926%	12.201%	12.483%
Haralick Facet [Har84] *		9.887%	10.128%	10.519%	10.792%	11.065%
Proposed method α		0.143%	0.147%	0.885%	5.54%	13.52%
Proposed method β		0.056%	0.057%	0.105%	0.542%	1.958%

Tableau IV.5 – Comparaison de la robustesse des détecteurs face à un bruit gaussien. Le pourcentage P_{MD} exposé est le pourcentage de détection manquées. Les meilleurs résultats sont montrés en bleu en gras, et les plus mauvais en rouge en italique. * : un pré-filtrage de type médian a été effectué. α : (3×3 , ordre 2). β : (5×5 , ordre 2)

3.3.2 Résultats

La méthode proposée est naturellement robuste au flou de par son modèle sous-jacent. De plus, la taille du noyau utilisé implique un système avec plus de contrainte ce qui se traduit par une moyenne implicite lors de sa résolution (moindre carré) sur plus d'échantillons. Cela implique une robustesse naturelle au bruit gaussien. Ces deux points sont largement soulignés par les tableaux IV.4–IV.5.

Les résultats obtenus pour le détecteur de Canny et celui d'Haralick montrent une certaine robustesse face au bruit mais une sensibilité évidente face au flou. Au contraire, le détecteur plus récent par filtrage non linéaire est très robuste au flou mais non face au bruit. Il est clair au vu de ces tables que le détecteur proposé est très robuste face au bruit et qu'il est aussi robuste face au flou en le comparant à ses pairs.

Le flou tend à atténuer les coefficients en les étalant. Les paramètres sont donc moins sensibles au contours, mais les plus fortes variations permettent toujours la bonne localisation de celui-ci. L'impact du flou sur l'image est donc sensiblement le même que celui d'augmenter la taille du noyaux N , voir figure IV.8b. Augmenter N permet de contrebalancer le bruit. De fait, un calcul direct montre que l'estimation par les moindres carré des coefficients c_i a une déviation standard de $O(\frac{\sigma}{N})$ où σ est la déviation standard locale, voir équation IV.23.

Le bruit impulsionnel est plus problématique. Ce type de bruit tend à déformer localement l'image. Il génère de fortes variations locales S_d , voir équation IV.19. Le détecteur élaboré s'appuie sur une modélisation de l'image et est donc naturellement faible face à cela. Il détecte ce type de bruit comme des contours. Pour pallier ce problème, il est nécessaire d'effectuer un

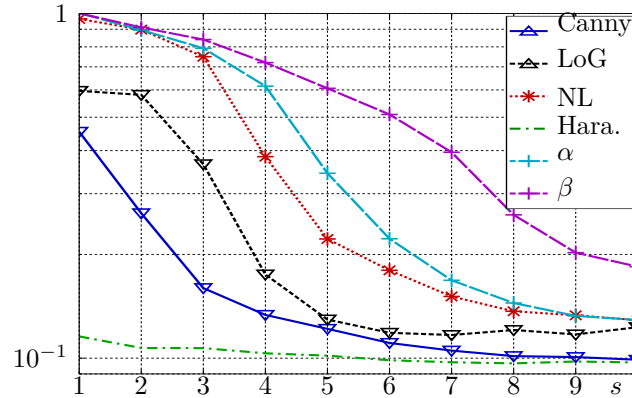


Figure IV.15 – Comparaison des détecteurs de contours par FoM calculé sur l'image IV.14c représentée en tant que fonction dépendant du rayon de flou s utilisé avec un bruit de déviation standard $\sigma = 2$. Les méthodes de Canny et LoG sont d'échelle 2. α et β sont les mêmes que dans les tables IV.4–IV.5.

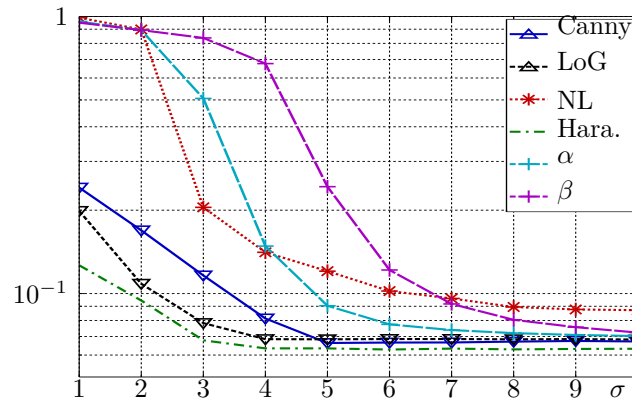


Figure IV.16 – Comparaison des détecteurs de contours par FoM calculé sur l'image IV.14a représentée en tant que fonction dépendant de la déviation standard σ du bruit utilisé avec un rayon de flou de $s = 2$. Les méthodes de Canny et LoG sont d'échelle 2. α et β sont les mêmes que dans les tables IV.4–IV.5.

prétraitement. Un filtrage médian est particulièrement efficace pour éliminer ce type de bruit. Il est intéressant de rappeler que ce type de bruit est généralement le plus problématique pour la grande majorité des détecteurs de contours existant.

Pour la seconde expérience, la figure IV.16 montre la FoM des détecteurs en tant que fonction dépendant de la déviation standard du bruit. La figure IV.16 montre quant à elle la FoM des détecteurs en tant que fonction dépendant du rayon de flou. Ces deux figures confirment que le détecteur d'Haralick [Har84] est bien plus sensible à la présence du flou qu'au bruit blanc, voir tables IV.4 et IV.5. De même, ces figures montrent que le détecteur LoG [HM86] est plus robuste que celui de Canny [Can86b] face au flou gaussien ; ce dernier est cependant meilleur face au bruit. Finalement, il peut être noté que le filtrage non-linéaire [LT10] reste particulièrement précis face au flou gaussien, bien que le détecteur proposé ait une meilleure FoM que ses compétiteurs. Cela est principalement dû au fait que les faux positifs détectés par

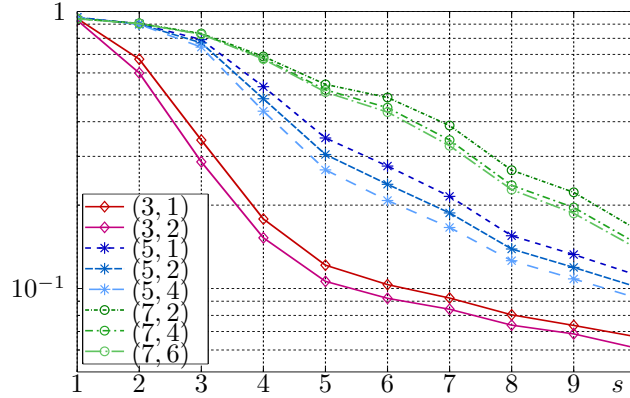


Figure IV.17 – Illustration de l’influence du couple de paramètres (N, K) par FoM en fonction du flou avec un bruit fixe $\sigma = 2$.

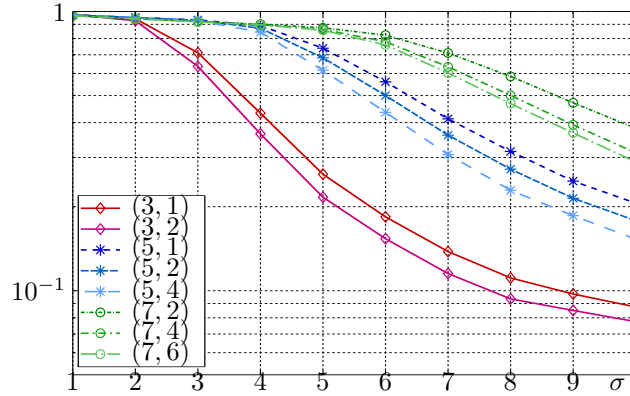


Figure IV.18 – Illustration de l’influence du couple de paramètres (N, K) par FoM en fonction du bruit avec un flou fixe $s = 2$.

la méthode proposée sont en moyenne plus proche des vrais contours que dans la méthode du filtrage non-linéaire.

La troisième expérience a permis de produire les figures IV.17 et IV.18 qui montrent l’évolution de la performance en tant que fonction FoM dépendant respectivement de la déviation standard du bruit gaussien et du rayon de flou. Ces deux figures montrent que la robustesse croît avec l’accroissement de K et de N — ce qui confirme nos précédentes observations. Toutefois, l’image de test étant très simple, elle ne nous permet pas de conclure sur la bonne détection de contours de faible intensité proche de contours d’intensité plus forte.

Il est toujours plus facile d’apprécier visuellement des résultats. A cette fin, une comparaison visuelle a été faite en utilisant les très célèbres images *Lena* et *boats*, voir figures IV.19. L’image *boats* a été bruité avec un bruit gaussien d’écart-type $\sigma = 10$.

Cette figure représente les résultats obtenus par les détecteurs comparés précédemment. Elle illustre particulièrement la sur-sensibilité du filtrage non linéaire [LT10] qui tend à faire



(a) Image *Lena*.



(b) Résultat du détecteur de Canny [Can86b].



(e) Résultat du détecteur LoG [HM86].



(f) Résultat du détecteur par filtrage NL [LT10] (sans pré-traitement).



(i) Résultat du détecteur proposé $N = 3$, $K = 2$.



(j) Résultat du détecteur proposé $N = 5$, $K = 2$.



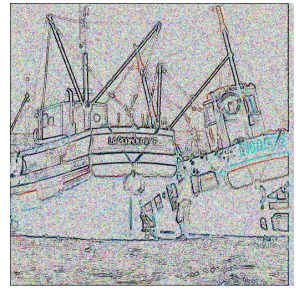
(c) Image *boats* bruité ($\sigma = 10$).



(d) Résultat du détecteur de Canny [Can86b].



(g) Résultat du détecteur LoG [HM86].



(h) Résultat du détecteur par filtrage NL [LT10] (sans pré-traitement).



(k) Résultat du détecteur proposé $N = 3$, $K = 2$.



(l) Résultat du détecteur proposé $N = 5$, $K = 2$.

Figure IV.19 – Comparaison visuelle sur les images *Lena* et *boats*.

ressortir tout pixel légèrement différent de ses voisins. Pour les détecteurs LoG [HM86] et de Canny [Can86b] ce comportement est atténué mais se retrouve lorsqu'on le confronte au bruit. La méthode proposée quant à elle permet de retrouver les contours aussi bien marqués que légers en modifiant les paramètres N et K , voir figures IV.19i et IV.19j. De plus, en adaptant la taille du noyau, la méthode proposée est plus robuste au bruit, voir figure IV.19k et IV.19l. Les résultats obtenus confirment clairement les résultats présentés précédemment. Une détection sur une plus grande image est disponible en annexe 3.

Les travaux sur ce détecteur et descripteur de contours ont fait l'objet de publication [SCGMB13].

4 Extraction des maillages projetés

Tout d'abord il convient de définir la notion de maillage projeté. Le terme maillage est ici abusif dans le sens où il s'agit en réalité d'un pavage non régulier ne comportant pas nécessairement de métrique liée. Le pavage pour qu'il puisse être projeté se doit d'avoir des frontières épaisses. Ce qui est projeté est donc un agencement plus ou moins complexe de segments épais. Le problème ici est de les retrouver après la déformation liée à la projection sur une surface non-plane.

Le problème dual revient donc à segmenter chaque "pavé" du pavage. C'est avec cette approche que nous avons construit notre algorithme d'extraction. Nous n'aurons dans un premier temps aucun a priori sur la forme des "pavés".

Notons que l'image initiale avant projection pourrait être utilisée. Nous avons fait le choix de ne pas l'employer directement dans l'algorithme d'extraction présenté ici afin de pouvoir la conserver comme élément de contrôle lors de l'appariement de clefs.

4.1 Extraction du motif

Afin de pouvoir extraire le motif, il est nécessaire de pouvoir déterminer avec précision la position de ces contours. Nous présenterons la méthode retenue dans la section suivante. Le but de cette partie est de définir les points importants, nécessaires à obtenir via le détecteur de contours.

Le détecteur présenté dans la partie précédente nous fournit un certain nombre de coefficients dont l'analyse nous indique bien plus que la simple présence de contours. En effet, nous avons les propriétés suivantes sous réserve d'utiliser à minima l'ordre $K = 2$:

- Détection : si la norme des coefficients est élevée, il y a un contour.

- Localisation : la bonne localisation du contour se situe au maximum de la norme suivant la perpendiculaire au contour.
- Orientation : les deux premiers coefficients exploitables c_1 et c_2 permettent de déterminer la direction du contour.
- Courbure : elle est fournie par les coefficients d'ordre 2. C'est-à-dire c_3 , c_4 et c_5 .

L'orientation est une propriété précieuse qui nous permet de déterminer un algorithme de recherche suivant le contour. Cela permet notamment de ne pas avoir à effectuer la convolution sur toute l'image mais sur une partie très limitée seulement. Le principe de la méthode est qu'à chaque fois qu'un pixel est détecté comme étant un contour, nous avons une information sur la localisation du prochain contour. Cette information est légèrement biaisée par la discrétisation de l'image. Il convient alors de tester non pas un unique pixel mais un segment de quelques pixels (typiquement 3) en utilisant la propriété de localisation. En suivant ainsi un contour à l'intérieur d'un pavage comme décrit précédemment, on finit par fermer un "pavé".

La segmentation de chaque "pavé" est donc relativement aisée sous réserve d'avoir un point de départ sur le contour. Outre le fait que cette méthode permet d'économiser des calculs inutiles lors de la convolution, elle permet aussi d'obtenir l'ensemble des pixels d'un contour de "pavé" et ces ensembles sont séparés les uns des autres. Cette propriété, qui peut sembler triviale n'est pas garantie lors d'une détection classique de contours, et il faut alors employer d'autres algorithmes afin de regrouper les pixels appartenant au même contour.

Afin de pouvoir déterminer quand s'arrêter, il suffit de déterminer si l'on repasse sur le contour après en avoir suivi tout le tour. Toutefois, nous avons préféré utiliser un indicateur s'appuyant sur le sens horaire (ou antihoraire) qu'est sensé décrire le parcours lors de la détection. Cela nous permet entre autre de stopper la détection plus rapidement lorsque nous sommes sur des zones déformées, ou avec des occlusions, entraînant des "pavés" parfois non-fermés. Pour cela, le calcul en temps réel de la moyenne des positions des pixels du contour nous sert de centre P_c pour la rotation. La rotation peut alors être testée grâce au produit scalaire et au produit vectoriel des vecteurs $(P_0 - P_c)$ et $(P_n - P_c)$. Les retours en arrière dans le parcours horaire (avec une marge de seuil) permettent d'arrêter la tentative de détection, vouée à l'échec.

4.1.1 Analyse du motif : recherche des coins

Ces contours de "pavé" sont ensuite traités afin de déterminer les nœuds du motif. Les nœuds ne sont pas sur les contours des "pavés" à cause de l'épaisseur des contours du motif

projeté. Il est néanmoins raisonnable de prendre l'hypothèse suivante :

H3 : *Les nœuds du maillage sont proches de coins dans le contour de chacun des “pavés” de leur voisinage.*

Cette hypothèse nous permet de construire un estimateur de la bonne localisation des points en utilisant la moyenne des pixels de type coins. Nous verrons plus tard que cette estimation n'est pas la seule envisageable, ni forcément la meilleure. Néanmoins cette recherche est obligatoire pour pouvoir segmenter le tour des “pavés” en segments.

Afin de détecter les coins, il est possible d'utiliser la version modifiée du détecteur de Harris avec nos coefficients comme expliqué dans la section 3.2.3. Toutefois le détecteurs de Harris n'est réellement performant que pour des angles proches de 90°, et cela se retrouve dans notre modification.

Il est possible d'obtenir les angles à partir de la classification en a paquets des directions des contours, a étant ici le nombre d'arêtes d'un “pavé”. On calcule ensuite les séparations des paquets en supprimant les pixels d'étiquettes isolées. Cette méthode, bien que plus adaptée à des pavage de type triangulation, n'en reste pas moins peu fiable lorsque les contours sont fortement déformés. Or le cas se produit systématique lors de la projection d'un maillage sur le bord d'un mannequin de par la surface trop inclinée recevant cette projection.

Enfin, une troisième approche consiste à localiser les maxima locaux de la fonction suivante, qui correspond à la distance par rapport au centre :

$$\begin{aligned} f : \llbracket 1, N \rrbracket &\longrightarrow \mathbb{N} \\ n &\longmapsto f(n) = \left\| P_n - \frac{1}{N} \sum_{i=1}^N P_i \right\|_2^2 \end{aligned} \quad (\text{IV.31})$$

Il est à noter que le centre calculé ne correspond pas totalement au centre de gravité à cause de la discrétisation. Cette fonction devrait avoir a maxima sauf très forte déformation. Ces maxima ont pour positions les coins des pavés. Cette méthode est d'autant plus robuste que le nombre d'arêtes du “pavé” est faible. Il peut être délicat de déterminer avec précision ces maxima locaux. Nous avons implémenté cette détection en utilisant un algorithme basé sur la ligne de partage des eaux [Beu90].

Cette recherche particulière a été implémentée dans le but de pouvoir extraire les maxima locaux d'une courbe bruitée directement. En effet, là où la majeure partie des algorithmes basés sur le gradient perdent leur efficacité, la ligne de partage des eaux [Beu90] effectue correctement la détection. A la base, la ligne de partage des eaux est généralement utilisée afin de segmenter les images selon les lignes de crêtes de la luminance des pixels. Il existe plusieurs implémen-

tations de l'algorithme. Nous avons choisi d'utiliser une ligne de partage des eaux en élevant petit à petit le niveau d'eau. Nous n'utilisons donc pas ici de ruissellement.

Sous réserve d'une initialisation correcte, la ligne de partage des eaux recherchant a maxima doit démarrer avec a bassins correctement situés dans les a minima. Cela revient donc à partir d'un type d'extrema, pour trouver les extrema de l'autre type. Cependant, la localisation des extrema de départ (minima) n'a pas besoin d'être précise. En fait, n'importe quel point de la cuvette formée par la courbe en cet endroit correspond, voir figure IV.20.

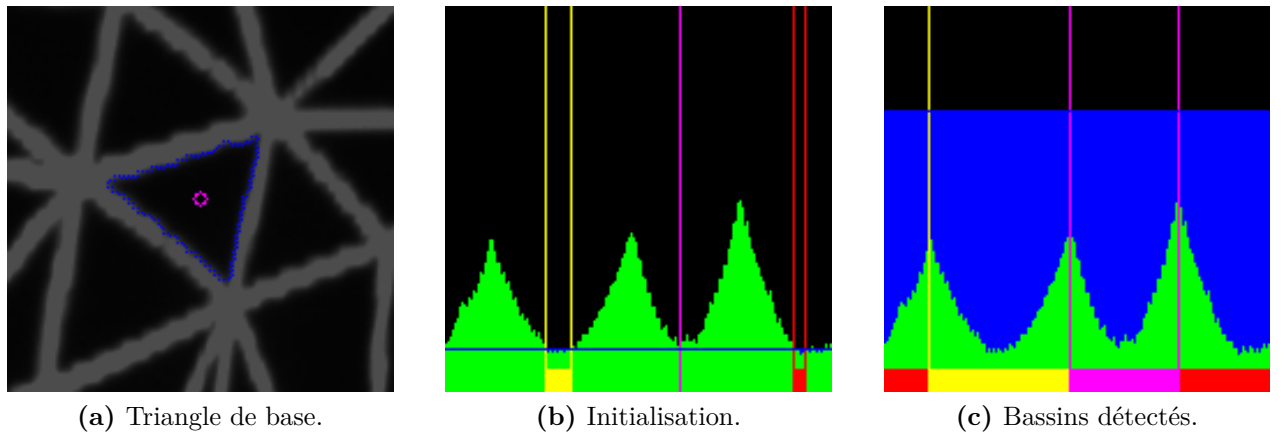


Figure IV.20 – Ligne de partage des eaux : principe. Son application dans le cadre de notre thèse se fait sur la distance des pixels des contours détectés du triangle par rapport à son centre.

Dans cette figure, la ligne horizontale bleue représente le niveau de l'eau tandis que la courbe verte est représenté comme une montagne. Le but est donc dans un premier temps d'initialiser l'algorithme à l'aide de a bassins (ici $a=3$ pour des triangles) représentés par de larges bandes horizontales jaune, mauve et rouge. Afin de mieux visualiser, le bord des bassins est symbolisé par les lignes horizontales de la couleur des bassins, voir figure IV.21a. Plus le niveau de l'eau monte, plus les bassins s'étendent. Afin de trouver les extrema, il faut donc déterminer la position où les bassins se rejoignent lorsque le niveau de l'eau monte, voir figure IV.20c.

Dans l'implémentation classique de cette version de l'algorithme de ligne de partage des eaux, l'eau monte petit à petit et à chaque montée on étend les bassins qui peuvent l'être.

Afin de diminuer le temps de calcul, nous avons choisi de nous inspirer des recherches dichotomiques. En effet, nous pouvons voir l'ordonnée comme un tableau fini d'éléments ordonnés, l'ordonnancement étant ici que les bassins de niveaux plus élevés incluent les bassins de niveaux moindres. Grâce à cela, nous pouvons effectuer une recherche dichotomique sur le niveau de l'eau adéquate pour que deux bassins se rejoignent. Nous avons donc eu la possibilité par ce biais d'augmenter plus rapidement la montée des eaux.

Si les bassins ne se sont pas rejoints, plusieurs étapes de calcul inutiles sont alors gagnées. En effet, la plupart des pentes sont suffisamment raide pour que la montée du niveau de l'eau de plusieurs pixels ne modifie pas les limites des bassins.

Si deux bassins se rejoignent, la montée a été trop rapide et il faut alors sous-échantillonner la montée en deux. On reprend donc les valeurs des bassins et de la ligne d'eau que l'on avait à l'étape précédente et divisons le pas par deux. On relance alors le processus jusqu'à ce que le pas se limite à la valeur 1. Il faut noter que comme l'espace est décomposé en deux à chaque fois, un seul test est nécessaire. En effet, dans cette configuration si au cours d'une décomposition les bassins ne se rejoignent pas, c'est qu'ils se rejoindront nécessairement dans la partie supérieure. On peut donc la décomposer directement en deux.

Cette seconde situation pose le problème de l'extension inutile du bassin lors de l'étape précédente puisque l'on revient à l'état précédent. L'algorithme reste tout de même généralement plus rapide que la ligne de partage des eaux par élévation classique, du moins dans l'utilisation que nous en avons fait.

Dans notre implémentation actuelle, la séparation entre les bassins est calculée séparément pour chaque couple de bassins, entre chaque pallier de montée des eaux. Bien que plus facile à implémenter, nous perdons ici en efficacité car des étapes dans la recherche dichotomique pour la séparation de deux bassins peuvent servir à la recherche de la frontières entre un autre couple. Il est donc possible d'optimiser l'implémentation actuellement utilisée sur ce point.

Bien que cette méthode fonctionne parfaitement dans notre application, il reste un problème de taille qui est de déterminer le point de départ des a bassins lors de l'initialisation. En effet, les bassins ne sont pas tous aussi profonds, il est donc difficile de déterminer aisément les positions adéquates pour que la segmentation des contours du triangle se fasse automatiquement.

Une première approche a été de monter progressivement le niveau de l'eau de façon classique jusqu'à la constitution de a bassins. Malheureusement, cette approche ne fonctionne pas dans le cas où les cuvettes ne sont pas toutes de même hauteur. En effet, à cause du bruit sur la courbe, ce genre de situation a tendance à créer plusieurs graines de bassins dans la même cuvette et entraîne automatiquement une mauvaise segmentation.

Nous avons aussi testé une séparation arbitraire en trois partie égale sur les abscisses mais les côtés d'un triangle générique n'étant pas identique, cela crée le même problème de placement des graines de bassins.

À ce jour, bien que cette méthode soit la plus précise que nous ayons trouvée, nous n'avons pas encore de méthode optimale pour le placement de tous bassins. De par l'algorithme de propagation, qui commence vers le milieu d'un segment du triangle, nous pouvons correctement positionner une des graines sans trop d'effort. La position des deux autres reste encore

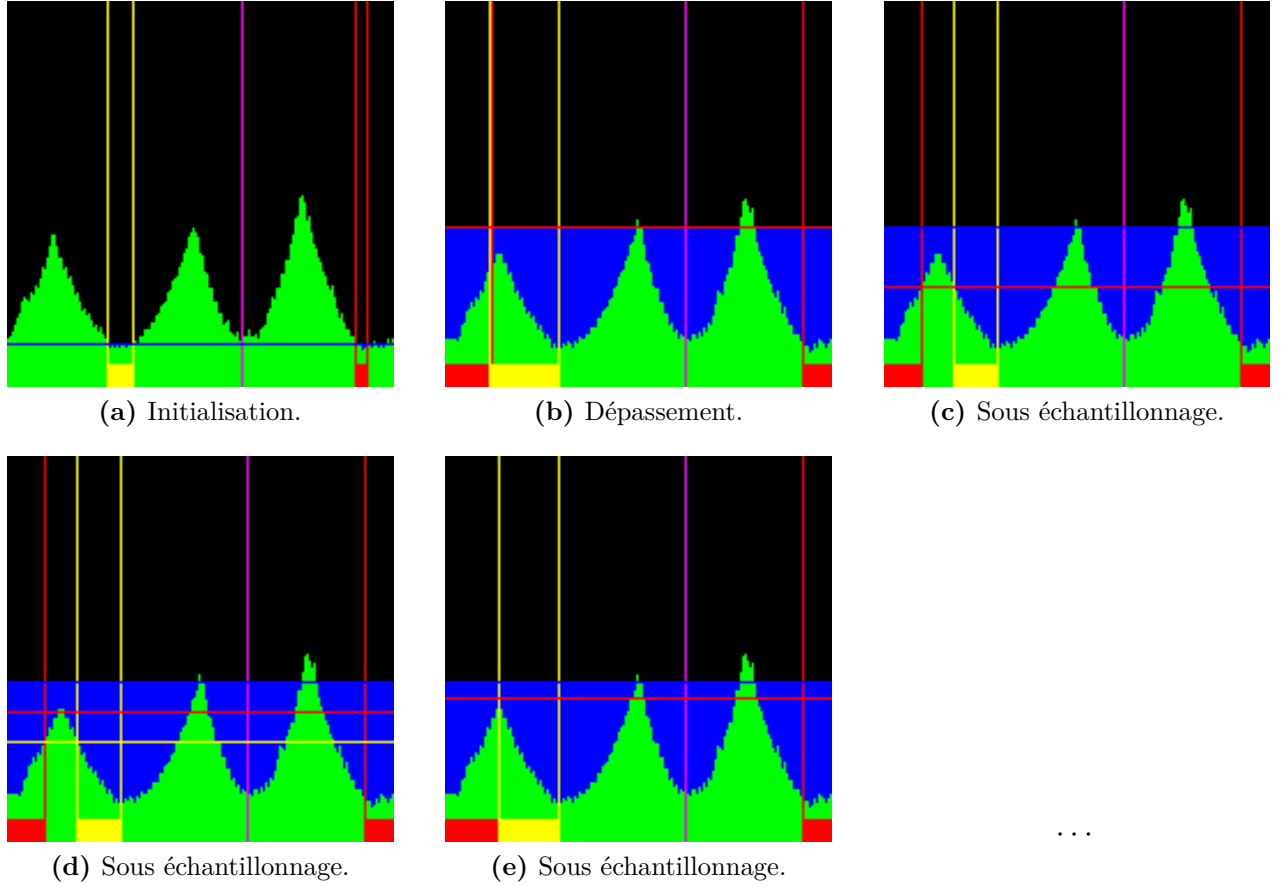


Figure IV.21 – Ligne de partage des eaux : recherche dichotomique. Ici, on essaye de déterminer la frontière entre les bassins jaunes et rouges uniquement, suite à une montée trop importante.

problématiques.

Une solution efficace mais couteuse consiste à séparer les pixels de contour en a paquet selon l'orientation de leur gradient, de prendre le pixel médian de chacun des a paquets et de faire de son correspondant dans la courbe de distance une graine pour notre algorithme de ligne de partage des eaux.

4.1.2 Analyse du motif : recherche des nœuds

La localisation de la position de chaque nœud est une étape délicate et cruciale dans notre application. C'est en effet cette position qui sera utilisée par la suite en tant que points à apparier, voir section 2. Il faut prendre en compte plusieurs paramètres liés à la construction du motif et à la déformation lors de sa projection.

Lors de la construction du motif, il est nécessaire de choisir une épaisseur pour ses traits.

Cette épaisseur implique que la position du coin le plus proche sur le tour d'un pavé voisin est dépendante de l'angle entre ses segments, voir [H3](#) et figure [IV.22](#).

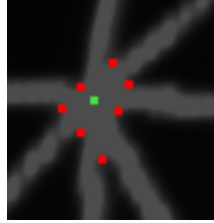


Figure IV.22 – Positionnement des coins (en rouge) par rapport au nœud (en vert). La distance varie selon l'angle entre les arrêtes.

Prendre uniquement la position des coins dans le voisinage est dès lors complexe et inefficace. En effet, comment choisir si le coin détecté est lié au nœud que l'on cherche à extraire, et comment choisir la bonne taille pour la zone de recherche des coins ? Cette question délicate n'a pas été traitée pour la simple et bonne raison que les positions des coins ne suffisent pas. Prenons par exemple le cas de la figure [IV.22](#). Les contours des “pavés” ont des angles relativement proches du nœud que l'on souhaite extraire, à l'exception de l'un des “pavés”. Le coin détecté pour ce “pavé” est donc plus éloigné du nœud que les autres. La moyenne est donc déviée dans cette direction.

On peut choisir de calculer l'intersection par extrapolation de l'affinement de chaque contour. Cette opération est relativement couteuse et est fondamentalement fautive. L'erreur entre le résultat et le point obtenu correspond alors à la déformation des segments droits en segments courbes par la projection sur une surface non-plane.

Nous avons opté pour une approche similaire à la détection de coins sous-pixels utilisée dans *OpenCV*⁶ lors de la détection des coins dans une mire de type damier. Cette détection est en réalité une amélioration en localisation d'un point déjà détecté. La délocalisation se fait en déterminant le minimum d'une énergie $E_{\text{localisation}}$ basée sur le gradient dans le voisinage de l'image :

$$\forall P_i \in V(P_c), E_{\text{localisation}}^{P_c} = \nabla I_{P_i}^T \cdot (P_c - P_i) \quad (\text{IV.32})$$

Autrement dit, on recherche les coordonnées du point P_c étant le plus proche de l'orthogonalité entre chaque vecteur formé du centre P_c de la fenêtre de recherche et du pixel P_i du voisinage d'une part et le vecteur gradient de l'image en P_i d'autre part. Si l'énergie est nulle, cette

6. *Open Computer Vision*

équation peut se réécrire ainsi :

$$P_c = \sum_{P_i \in V(P_c)} \left(\nabla I_{P_i} \cdot \nabla I_{P_i}^T \right)^{-1} \cdot \sum_{P_i \in V(P_c)} \left(\nabla I_{P_i} \cdot \nabla I_{P_i}^T \cdot P_i \right) \quad (\text{IV.33})$$

Le centre P_c peut ainsi être trouvé par approximation successive en décalant à chaque fois la nouvelle fenêtre de recherche $V(P_c)$ sur le point P_c trouvé jusqu'à stabilité de P_c (à un seuil près).

Cette méthode d'estimation est bien mieux adaptée à notre cas que les méthodes précédentes, comme déterminé empiriquement. Cela s'explique par la malléabilité liée à la considération de tout le voisinage et non seulement de certains éléments. De fait, le centre ainsi optimisé n'est plus autant dépendant de la bonne localisation des coins mais uniquement des gradients locaux et donc de la disposition des contours localement.

Au vu des résultats et du fait que la totalité des algorithmes ont été développé avec *OpenCV* [Lag11] pour sa gestion des images et son algorithmie matricielle, la fonction d'*OpenCV* “*cornerSubPix*” à été détournée pour l'implémentation de notre algorithme. Le point de départ choisi est donc simplement la moyenne des coins détectés dans un voisinage de taille arbitrairement fixé. En effet, si le voisinage est trop petit pour prendre en compte tous les coins, l'optimisation sous-pixel tendra vers une correction de cette erreur.

4.2 Extraction du motif par propagation

Puisque l'extraction d'un pavé et la détermination de ces segments est possible, il suffit de mettre en place un algorithme par propagation.

Pour l'initialisation, il convient de trouver le contour d'un premier “pavé”. Cela peut se résumer à trouver un point à l'intérieur d'un “pavé” avant de partir dans une direction (par exemple le haut) jusqu'à croiser notre premier point de contour. Ensuite l'extraction du contour du “pavé” se fait comme décrit précédemment.

Le choix du premier pavé extrait ne change aucunement les performances de l'algorithme. Qu'il soit sur le bord ou au centre, la rapidité sera la même de par l'algorithme de propagation utilisé qui est basé sur une pile d'instruction et non sur un algorithme récursif. La seule chose à prendre en compte est de vérifier que l'image n'est pas séparée en deux parties par une zone d'occultation. Dans un tel cas, il est alors nécessaire de lancer plusieurs fois l'algorithme : une fois par partie connexe (avec triangle de départ dans cette partie).

Suite à cela, en trouvant les segments du contour (voir section 4.1.1) et en prenant à chaque fois le pixel central au segment décalé vers l'extérieur du pavé, nous avons a extractions de

Algorithme 2 Algorithme d'extraction du motif de type triangulation

Donnée: Une image acquise de la scène avec le motif projeté dessus

Résultat: Une partie de la triangulation est retrouvée et stockée en mémoire

Déterminer un pixel P de départ aléatoirement mais étant dans un triangle

$PILE_EXECUTION \leftarrow \emptyset$

$TRIANGULATION \leftarrow \emptyset$

Tant que P n'est pas sur un contour **faire**

$P \leftarrow P - \vec{y}$ {Décalage vers le haut}

Fin tant que

Empiler($PILE_EXECUTION, P$)

Tant que $PILE_EXECUTION$ non vide **faire**

Dépiler($PILE_EXECUTION, P$)

Si P n'est pas sur un triangle déjà extrait **alors**

$Q \leftarrow P$

$CONTOUR \leftarrow \emptyset$

Répéter

$CONTOUR \leftarrow CONTOUR \cup Q$

On déplace Q dans le sens horaire sur le contour

Tant que $Q = P$ OU le contour décrit un angle de plus de 190°

Si $P = Q$ **alors**

Segmenter $CONTOUR$ pour trouver les 3 sommets A , B et C

Pour tout $S \in \{A, B, C\}$ **faire**

Si S est lié à un groupe de coins $V(S)$ déjà existant **alors**

$V(S) \leftarrow V(S) \cup S$

Sinon

$V(S) \leftarrow S$ {Création}

Fin si

Fin pour

Insérer ($V(A), V(B), V(C)$) dans la Triangulation

Pour tout $Q \in \{\text{Milieu}[AB], \text{Milieu}[BC], \text{Milieu}[CA]\}$ **faire**

$P \leftarrow Q$

Répéter

$P \leftarrow P + \frac{\nabla Q}{|\nabla Q|}$ {On déplace P vers l'extérieur du triangle}

Tant que P n'est pas sur un contour

Empiler($PILE_EXECUTION, P$)

Fin pour

Fin si

Fin si

Fin tant que

Retourner $TRIANGULATION$

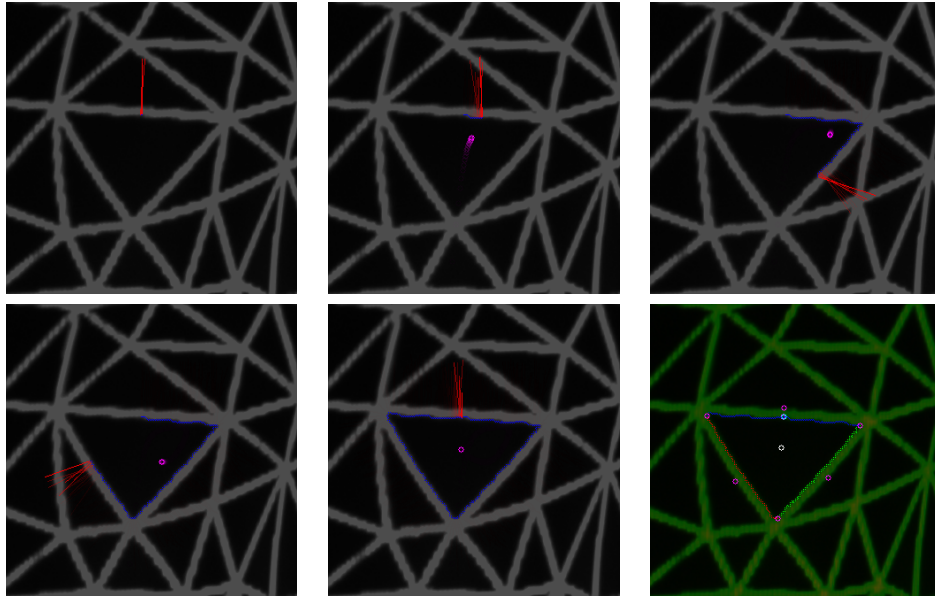


Figure IV.23 – Extraction d’un triangle. Le contour est détecté au fur et à mesure. La détection s’arrête lorsque l’on a fait le tour (sens horaire par rapport au centre de gravité du triangle, aussi détecté au fur et à mesure, ici en violet sur les 5 premières images). Le gradient est indiqué par le vecteur rouge. Dans la dernière image, on peut voir la segmentation du contour obtenue par ligne de partage des eaux. Les trois points violets situés sur les contours sont empilés suite à cette détection dans l’algorithme d’extraction.

“pavé” à effectuer. Ces extractions sont exactement la même chose que ce qui a été fait lors de l’initialisation. Le choix que nous faisons ici de prendre le pixel central est arbitraire et concrètement tout pixel un minimum éloigné des bords du segment conviendrait. Si le pixel était trop proche du bord d’un segment en revanche, il pourrait être possible que la recherche se poursuive non pas sur le voisin direct mais sur le voisin du voisin du fait de l’épaisseur des contours du motif.

L’implémentation utilisée se fait sous forme de pile d’extraction à effectuer. Un marquage est utilisé afin de déterminer si un contour a déjà été extrait ou non. Lors de chaque extraction, il est nécessaire de reconstruire la structure de stockage adéquate afin de garder le motif en mémoire.

Dans le cas où l’algorithme n’arriverait pas à faire le tour d’un “pavé”, il est préférable d’indiquer qu’à cet endroit existe un contour mais que nous ne pouvons l’utiliser. Nous sommes très probablement face à un problème d’occultation. Les nœuds correspondant aux sommets du “pavé” impossible à extraire doivent alors être retrouvés par voisinage en remontant la pile d’exécution. Il est de fait impossible de calculer correctement une signature pour ces nœuds.

Nous nous sommes surtout intéressés au motif de type triangulation de Delaunay. Le critère

de Delaunay n'est en réalité pas spécifiquement utilisé, mais permet de limiter les écrasements de triangles ce qui est utile dans notre cas. La structure qui se doit d'être reconstruite est dès lors une structure de données permettant de stocker une triangulation.

Si les sauts entre “pavés” permettent de restituer l'adjacence des “pavés” (triangles) dans la structure de donnée, la bonne localisation des points doit se faire au fur et à mesure. Dans la section 4.1.2, il est devenu évident que la meilleure localisation pour les nœuds devait se faire avec la méthode proposée, basée sur le gradient du voisinage. Cette méthode demande un point de départ et nous nous sommes proposé d'utiliser la moyenne des coins détectés. C'est cette étape qui se fait au fur et à mesure de l'extraction par propagation. L'étape d'affinage en utilisant le gradient du voisinage ne vient qu'à la fin, afin d'affiner la localisation de chaque nœud.

A la fin de l'exécution de l'algorithme d'extraction proposé, un pavage partiel est donc obtenu ; ou plus concrètement dans notre utilisation, une triangulation partielle. Il est donc possible de calculer les clefs introduites précédemment sur la triangulation partielle.

5 Conclusion

La cabine d'acquisition telle qu'elle a été conçue répond à un certain nombre de contraintes techniques et budgétaires pour une application donnée, voir chapitre II. Dans sa conception, un angle important a été induit entre les capteurs. Cela se traduit lors du procédé classique d'appariement par des résultats peu probants, c'est-à-dire la plupart du temps mal appariés. Les contraintes testées permettent effectivement d'augmenter le pourcentage de bon appariement, mais le nombre total de couples décroît, rendant la reconstruction de moins en moins précise. L'idée était donc de structurer la répartition de ces points.

Dans ce chapitre a été élaboré un motif particulièrement adapté à la reconstruction tridimensionnelle de par son unicité. Sa détection n'est cependant pas des plus simples et demande un certain temps de calcul. Toutefois, l'appariement peut se baser entièrement sur ce motif sans faire appel aux notions du chapitre précédent, ce qui le rend particulièrement intéressant pour des applications stéréovision avec de gros écarts entre appareils de capture.

Ce motif est donc la solution envisagée au problème de mauvaise détection que nous avons avec la cabine de reconstruction telle qu'elle a été conçue. Malgré ces résultats statistiquement prouvés, son application n'est pas des plus évidentes du fait de la nécessité de localiser les nœuds. Enfin, la reconstruction obtenue n'étant qu'éparse, on a pu noter qu'une méthode de raffinement du maillage doit être implémentée afin d'obtenir une meilleure précision pour la

surface reconstruite.

Dans le prochain chapitre, un raffinement possible pour obtenir des points tridimensionnels supplémentaires va être proposé. Nous profiterons aussi de ce chapitre dédié au passage à la 3D pour expliquer comment, à partir des couples de points, il sera possible de retrouver la position tridimensionnelle.

Chapitre V

Reconstruction tridimensionnel et raffinement

Après avoir exposé les différentes méthodes de mesures tridimensionnelles utilisables dans le cadre de la cabine d’acquisition proposée dans le chapitre II, le problème de la reconstruction tridimensionnelle a été abordée au chapitre III. Ce problème est constitué des deux parties “mise en correspondance” de points homologues et “positionnement tridimensionnelle”. Les chapitres III–IV ont largement exposé la problématique de la détection de points homologues afin de les appairer.

Ce chapitre expose le traitement tridimensionnel. Le but de ce traitement est d’obtenir les points tridimensionnels correspondant à la réalité de la scène. Il existe différents niveaux de reconstructions tridimensionnelles. Ce chapitre débute donc avec le positionnement tridimensionnel des couples de points supposés bien appariés et s’étend sur le traitement de ces données tridimensionnelles. En effet, un nuage de point seul est difficilement exploitable. Il convient donc de déterminer des relations de voisinage entre chaque point de ce nuage afin de reconstituer la surface tridimensionnelle des objets constituant la scène acquise. De plus, il est nécessaire d’étudier cette surface afin de déterminer les zones de celle-ci nécessitant plus ou moins de points afin de correspondre au mieux à la réalité et ainsi éviter la présence de plusieurs points n’apportant aucune information.

Ce chapitre se conclue donc sur les idées prometteuses envisagées afin de raffiner l’acquisition éparse obtenue avec le motif du chapitre précédent. Ce motif comporte des avantages évidents lorsqu’il s’agit d’établir les relations de voisinage nécessaires lors de l’utilisation d’un nuage de point en tant que surface. De plus, la plupart des modèles de numérisation partent d’une acquisition fine avant de retirer les points inutiles à la description de la surface. Le motif proposé nous permet d’introduire une stratégie consistant à raffiner uniquement les zones le nécessitant. C’est donc l’inverse de la stratégie couramment utilisé qui part d’un nuage dense à clairsemer.

1 Reconstruction tridimensionnelle

La reconstruction tridimensionnelle commence avec l'estimation des paramètres internes et externes de chaque caméra (étalonnage) et s'arrête avec la restitution tridimensionnelle des couples de points image m . Cette restitution peut être de diverses natures suivant l'application voulue.

Une hiérarchisation des différentes reconstructions va être dans un premier temps abordée. S'en suivra un bref retour sur l'étalonnage des capteurs qui est en réalité dépendant du niveau de hiérarchisation souhaité. Enfin, le problème de la triangulation spatiale sera traité.

1.1 Stratification des reconstructions tridimensionnelles

Faugeras [Fau95] a défini une hiérarchisation de l'espace tridimensionnel. Selon les informations à extraire d'une reconstruction, le calibrage peut varier entre quatre niveaux distincts :

$$\text{projectif} \Leftarrow \text{affine} \Leftarrow \text{métrique} \Leftarrow \text{euclidien}$$

Cette hiérarchisation est depuis utilisée par de nombreux scientifiques, comme Pollefeys [PVG99]. La figure V.1 illustre cette hiérarchisation par les différentes représentations d'un cube dans chacune des différentes reconstructions. Le niveau d'une reconstruction définit le type de transformation géométrique liant l'objet et son modèle tridimensionnel.

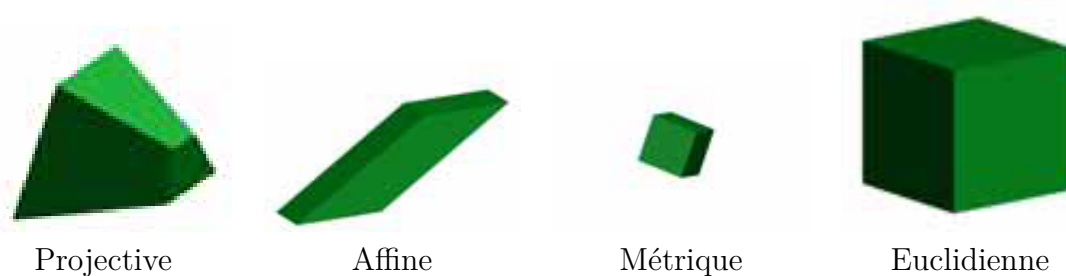


Figure V.1 – Hiérarchisation des reconstructions.

La reconstruction projective ne permet d'évaluer que le birapport entre les images d'un même point tridimensionnel. La reconstruction affine permet aussi de restituer le parallélisme existant entre les contours. La reconstruction métrique permet de restituer les angles et par ce biais, il est possible d'effectuer des rapports entre mesures dans toutes les directions et non uniquement des rapports sur les droites parallèles entre-elles. Le dernier niveau fixe un facteur d'échelle permettant de mesurer directement des valeurs correspondant à la réalité.

Dans le cadre de la cabine d'acquisition souhaitée, la reconstruction doit au moins être métrique. Et dans ce cas, il est nécessaire d'avoir alors deux points précis permettant d'estimer le facteur d'échelle. Cela reviendrait à simuler une reconstruction directement euclidienne.

1.1.1 Reconstruction projective

Le premier niveau de cette hiérarchie est la reconstruction projective. A ce stade, la reconstruction est fidèle à la réalité à une transformation projective près. Ce type de transformation est représenté par une matrice carré non-singulière de taille 4 dans l'espace tridimensionnel. Du fait de la projection, cette transformation est définie à un facteur d'échelle près. Elle possède donc 15 degrés de liberté. Il est possible d'obtenir la reconstruction projective à partir d'un ensemble de correspondances entre deux vues grâce à une décomposition en matrice fondamentale [Har92b].

1.1.2 Reconstruction affine

Le deuxième niveau est la reconstruction affine. Elle est fidèle à la réalité à une transformation affine près. Une telle transformation comporte 12 degrés de liberté :

$$\hat{T}_A \propto \left(\begin{array}{c|c} A_{3 \times 4} & \\ \hline 0 & 1 \end{array} \right) \quad (\text{V.1})$$

Pour passer du niveau projectif au niveau affine, il faut restituer le parallélisme. Cela passe par la détermination d'un plan particulier. Dans l'espace projectif, il s'agit du plan défini par l'ensemble des points dont le facteur d'échelle des coordonnées homogènes est nul. Ceci correspond donc à l'ensemble des points d'"intersection" de droites parallèles. C'est-à-dire le point de fuite situé à l'infini.

Il suffit alors de déterminer les coordonnées du plan contenant les points de fuite puis d'estimer la transformation projective qui l'appliquera à sa position canonique $\hat{\Pi}_\infty = (0 \ 0 \ 0 \ 1)^T$. Une fois les coordonnées du plan déterminées dans l'espace projectif, la transformation est de la forme :

$$\hat{T}_{PA} \propto \left(\begin{array}{c|c} A_{3 \times 3} & 0 \\ \hline \hat{\Pi}_\infty^T & \end{array} \right) \quad \text{avec } \det A_{3 \times 3} \neq 0 \quad (\text{V.2})$$

Puisqu'il n'y a aucune hypothèse sur la matrice $A_{3 \times 3}$ autre que l'inversibilité, on peut choisir la matrice identité :

$$\hat{T}_{PA} \propto \left(\begin{array}{c|c} Id_{3 \times 3} & 0 \\ \hline \hat{\Pi}_{\infty}^T & \end{array} \right) \quad (V.3)$$

Seules les distances selon une même direction peuvent être mesurées sur une reconstruction affine.

1.1.3 Reconstruction métrique

Le niveau directement supérieur est la reconstruction métrique. Il consiste à restituer les angles lors de la reconstruction. Ce type de reconstruction est fidèle à la scène réelle à une similarité près.

Les similarités sont des compositions de rotations (3 degrés de liberté), de translations (3 degrés de liberté) et d'homothéties (1 degré de liberté). Elles comportent donc 7 degrés de liberté et sont donc définies par :

$$\hat{T}_M \propto \left(\begin{array}{c|c} hR_{3 \times 3} & t \\ \hline 0 & 1 \end{array} \right) \quad (V.4)$$

où R est une matrice de rotation, h est le facteur d'échelle de l'homothétie et t le vecteur de translation.

Pour permettre une reconstruction métrique, il faut identifier une conique particulière Ω appelée la conique absolue. De façon générale, une conique se représente dans le plan par une matrice carré C de taille 3×3 définissant l'ensemble des points m par l'équation :

$$\hat{m}^T C \hat{m} = 0 \quad (V.5)$$

La conique définie par l'ensemble des droites ℓ tangentes à une conique C est la conique duale C^* de C . Elle respecte l'équation suivante :

$$\ell^T C^* \ell = 0 \quad (V.6)$$

Lorsque C n'est pas dégénérée, elle est de rang 3 et on a :

$$C^* \propto C^{-1} \quad (\text{V.7})$$

La conique absolue est définie par l'ensemble des points $M = (x \ y \ z \ w)^T$ ayant la propriété suivante :

$$x^2 + y^2 + z^2 = 0 \quad \text{et} \quad w = 0 \quad (\text{V.8})$$

L'objet dual de Ω considéré est alors la quadrique duale absolue Q^* composée de plan tangents à Ω . Son expression canonique est donc :

$$Q^* \propto \left(\begin{array}{c|c} Id_{3 \times 3} & 0 \\ \hline 0 & 0 \end{array} \right) \quad (\text{V.9})$$

En utilisant l'équation [V.7](#) et [V.8](#), l'expression mathématique de Ω_∞ dans Π_∞ a donc pour image :

$$\omega_\infty \propto Id_{3 \times 3} \quad (\text{V.10})$$

La conique absolue se projette également en une conique dans les images. Il est plus simple d'utiliser la duale de la conique absolue pour réaliser cette projection, sachant qu'une seule matrice la définit. Il vient alors que la duale de l'image de la conique absolue est donnée par la relation suivante pour le capteur d'image c :

$$\omega_{\infty,c}^* \propto P_c Q_\infty^* P_c^T \quad (\text{V.11})$$

Lorsque Ω_∞^* est à sa position canonique, la relation suivante est vérifiée [[HZ04](#)] :

$$\begin{aligned} \omega_{\infty,c} &\propto (P_c Q_\infty^* P_c^T)^{-1} \\ &\propto \left((K_{\text{int},c} R_c)^T Q_\infty^* K_{\text{int},c} R_c \right)^{-1} \\ &\propto K_{\text{int},c}^{-T} R_c \Omega_\infty R_c^T K_{\text{int},c}^{-1} \\ &\propto (K_{\text{int},c} K_{\text{int},c}^T)^{-1} \end{aligned} \quad (\text{V.12})$$

où $K_{\text{int},c}$ est la matrice des paramètres intrinsèques et R_c la matrice de rotation des paramètres extrinsèques du capteur d'image c . La translation se projetant sur le plan infini, elle n'influe pas dans la détermination de l'image.

Pour passer d'une reconstruction affine à une reconstruction métrique, il faut déterminer la

transformation affine qui placera la conique absolue sur sa position canonique. Cette transformation peut être choisie comme étant [Har94] :

$$\hat{T}_{AM} \propto \left(\begin{array}{c|c} K_{\text{int}}^{-1} & 0 \\ \hline 0 & 1 \end{array} \right) \quad (\text{V.13})$$

1.1.4 Reconstruction euclidienne

Ce type de reconstruction diffère du niveau métrique par le fait que l'échelle est fixée. Ainsi, les distances peuvent être mesurées de façon absolue. La reconstruction est connue à une transformation rigide près, c'est-à-dire la composition d'une rotation (3 degré de liberté) et d'une translation (3 degré de liberté). Ces transformations n'ont donc que 6 degrés de liberté :

$$\hat{T}_E \propto \left(\begin{array}{c|c} R & t \\ \hline 0 & 1 \end{array} \right) \quad (\text{V.14})$$

Pour passer d'une reconstruction métrique à une reconstruction euclidienne, il suffit de réaliser une mise à l'échelle. Le facteur d'échelle ne peut être connu qu'en disposant d'une distance de référence dans la scène.

1.2 Rappel sur la calibration

La calibration à partir d'une mire de référence permet d'estimer directement une reconstruction euclidienne au contraire des méthodes d'*autocalibration* (étalonnage automatisé). Lors d'une calibration par mire, un certain nombre de correspondances entre les images de deux capteurs sont nécessaires. Ces correspondances sont déterminées par l'extraction de points spécifiques de la mire dans chaque vue. Une seule image est donc nécessaire mais de multiples prises de l'objet permettent plus de points de correspondance ; et donc une approximation plus stable des paramètres.

L'*autocalibration* étant de plus impossible dans notre cas (voir Chapitre II section 3), son implémentation n'a pas été plus étudiée.

1.3 Triangulation à partir de deux vues

La dernière étape d'une reconstruction tridimensionnelle est la restitution de la scène. Pour ce faire, on calcule la droite de vue issue du centre optique et passant par chaque projection m . Ces droites correspondent au trajet effectué par les photons allant de la scène vers le capteur d'image. Pour deux projections homologues, celle-ci se coupent en un point tridimensionnel M unique.

La détermination de ces droites est cependant dépendante de la bonne estimation des paramètres internes (pour le centre optique) et externes des appareils de captures. De plus, la discrétisation des matrices d'acquisition CCD provoquent de légers écarts entre ces droites et cela même suite à une localisation sous-pixel (entre les coordonnées discrètes relatives aux pixels) des pixels homologues. En ajoutant à cela le bruit d'acquisition, il est clair que ces droites de vues ne peuvent qu'être approximée.

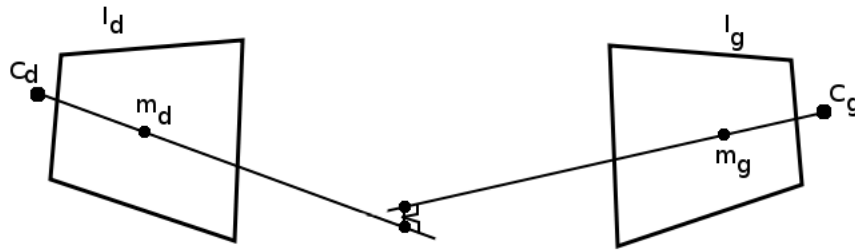


Figure V.2 – Triangulation spatiale : les droites ne se coupent pas forcément.

Ces droites de vues ne se coupent donc pas rigoureusement dans la pratique mais passent très proches l'une de l'autre. Il s'agit alors de déterminer le point tridimensionnel le plus probable.

Hartley et Sturm ont comparé les différentes méthodes existantes de triangulation dans [HS97]. La méthode du point moyen (*midpoint*) qui est totalement déconseillée par les auteurs du fait de ses performances pour une reconstruction projective ou affine, se comporte plutôt bien dans une reconstruction métrique ou euclidienne. Nous avons choisi d'implémenter cette méthode pour sa simplicité et parce que nous travaillons en reconstruction euclidienne. Elle dispose aussi de l'avantage non négligeable de se généraliser aisément à une reconstruction à partir de plus que deux appareils de captures. Il n'est toutefois pas inutile d'envisager de la remplacer à l'avenir par une adaptation multi-vues de la méthode *poly abs* décrite par Hartley et Sturm afin de rendre la cabine plus apte à gérer d'autres type de reconstruction.

La méthode du point moyen est très simple dans sa conception. Il s'agit de déterminer sur chacune des droites le point le plus proche de l'autre droite et de prendre la moyenne des

deux. En pratique l'éloignement entre ces droites est suffisamment faible pour que les écarts de mesure induits soient minimes. Cela est d'autant plus vrai que les appareils de captures de la cabine possèdent un angle important entre eux. Outre le problème de mauvais appariement, l'erreur tridimensionnelle a donc bien plus de chance d'être engendré majoritairement par une mauvaise localisation des points homologues que par la méthode utilisée.

Notons enfin que l'avantage de la méthode original de détermination des couples d'images que nous avons développé au chapitre précédent est qu'elle nous permet d'obtenir directement une surface tridimensionnelle. En effet, les relations de voisinage des points avant projection sur la scène sont utiles pour déterminer les nouvelles relations de voisinage sur la surface tridimensionnelle.

Une triangulation de Delaunay plane projetée sur une surface plane et frontale créera une triangulation de Delaunay plane dans l'espace tridimensionnel. Il en est de même pour la projection sur toutes surfaces frontale dont la courbure est moindre que la distance entre les points. Sur des surfaces plus accidentées, il faut donc porter une attention particulière aux relations de voisinage qui ne correspondent plus forcément à une approximation de la surface par manque de points mais qui dans tous les cas correspondent à des voisinages sur l'approximation de la surface. Cela simplifie énormément l'étape de conception de la surface à partir du nuage de points tridimensionnels obtenu par la triangulation spatiale. C'est en effet un problème couteux auquel sont confrontées toutes les autres méthodes de numérisation.

1.4 Reconstruction par bloc de capteur

Au sein de la cabine d'acquisition, il existe plusieurs blocs d'éléments. Ceux-ci sont composés de quatre appareils de captures disposés en carré et d'un vidéo-projecteur au centre de ce carré, tous orientés vers la scène. La cabine est donc composée de 6 de ces blocs disposés en cercle tout autour de la scène.

Il est alors possible d'effectuer deux traitements. Le premier consiste à trianguler avec deux appareils de captures, puis d'effectuer la même opération avec les deux restants (voire d'effectuer toutes les combinatoires possibles) et de fusionner les résultats. Cette disposition des blocs a été introduite pour limiter les zones cachées dans une vue ; et le cas échéant pouvoir s'appuyer sur les vues restantes. Avec ce premier choix, certains points seront détectés par certains capteurs uniquement. L'étape de fusion des données sera alors complexe.

La seconde possibilité est la triangulation à partir des quatre appareils de capture simultanément. En effet, plutôt que de fusionner l'information en aval, il est bien plus simple de la fusionner en amont dans la méthode. Il est de plus possible d'étendre la méthode du point

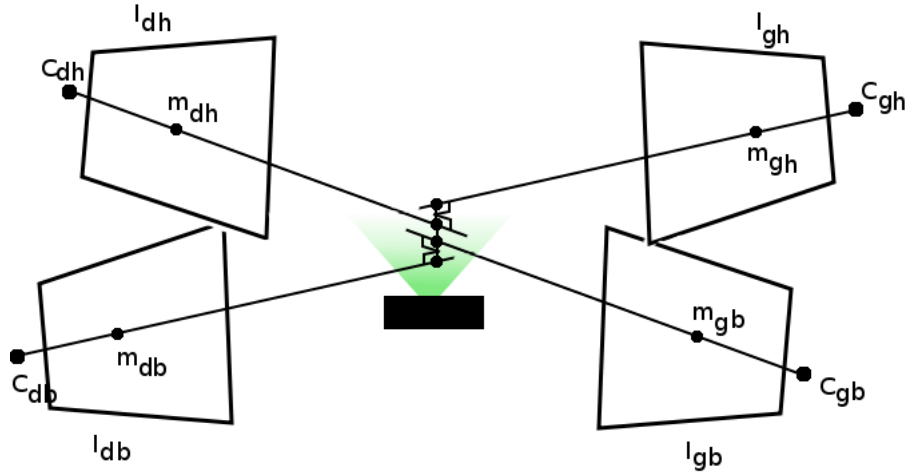


Figure V.3 – Schéma d'un bloc de capteur de la cabine dans le cadre de la triangulation spatiale.

milieu à plus que deux droites. L'idée étant de trouver l'ensemble des points appartenant à une droite de vues et étant le plus proche de l'une des autres droites, puis d'en faire la moyenne. Si le point est visible dans les 4 appareils du bloc, il y aura donc 4 droites et donc 12 points dans l'ensemble. Pour 3 images du point tridimensionnel, il y aura 6 points dans l'ensemble et pour le cas classique stéréovision, on se retrouve de nouveau avec deux points. Le fait de prendre la moyenne sur plus d'éléments implique une meilleure approximation de la position tridimensionnelle du point. De plus, en étudiant l'écart-type de l'ensemble, il est possible de rejeter un mauvais appariement ou encore de renforcer la bonne localisation spatiale des points de la scène.

Cette idée n'a pas eu le temps d'être poussé plus en avant par manque de temps et actuellement tous les appareils de capture ne sont pas utilisés. Nous avons toutefois vérifié la faisabilité de cette démarche et son implémentation pourra être réalisée dans de futurs travaux.

1.5 Assemblage des maillages surfaciques de bloc

Il s'agit ici de fusionner les maillages correspondant à l'acquisition de la scène par des blocs de capteurs différents. Ce problème est souvent rencontré en maillage surfacique. Plusieurs algorithmes permettent de réaliser ce type de fusion [KSMK99, YZX⁺04, JLW⁺06, LPMV10].

On notera que puisque chaque maillage correspond à une partie de la même surface, les maillages se recouvrent au niveau de leur intersection. Il faut alors déterminer la zone de recouvrement afin d'ajouter les points de l'une des surfaces incluse dans cette zone de recouvrement dans l'autre surface tronquée de cette zone de recouvrement, puis de souder les bords.

2 Raffinement

2.1 Détection des zones nécessitant raffinement

Le système d'acquisition présenté permet donc d'obtenir une triangulation surfacique tridimensionnelle de la surface de la scène. Cette triangulation étant relativement grossière, une analyse de celle-ci est ensuite nécessaire afin de définir les zones à raffiner en lui ajoutant des nœuds, ainsi que les zones pouvant être déraffinées pour alléger le nombre de points et ainsi la taille mémoire.

Il faut noter qu'après la triangulation spatiale, la triangulation obtenue ne respecte plus nécessairement le critère de Delaunay : certaines arêtes peuvent se trouver inversées par rapport à celui-ci. Ces zones sont les premières à nécessiter un raffinement. En effet, le changement effectué en passant de la projection d'une surface plane à la surface tridimensionnelle que l'on souhaite numériser entraîne une modification des distances entre les points qui devrait donc modifier la topologie de la triangulation. Cela est dû à une forte courbure dans les zones problématiques. Ce problème de courbure dans la surface entraîne aussi des imprécisions légères dans les zones où il manque des points pour décrire correctement la surface.

La solution consiste donc à analyser la courbure discrète calculée en chaque nœud de la triangulation surfacique. Nous faisons ainsi appel au modèle tridimensionnel obtenu afin d'évaluer les zones nécessitant raffinement. Les seuls supports que nous ayons étant les images d'origine, le travail consiste alors à extraire un point homologue dans la projection de la zone nécessitant un raffinement dans chaque vue. Puisque les projections de ces zones forment les triangles déformés du motif dans chaque vue, la zone de recherche correspond donc à l'intérieur de ces triangles.

Lors de l'extraction de la triangulation dans chaque vue, la détection du contour de chaque triangle, puis de sa segmentation en arrêtes, a été effectuée. Les données ayant été stockées, il est facile d'estimer le milieu de chaque arrête. Il est donc possible de subdiviser chaque triangle de la triangulation en 4 triangles au moins une fois aux endroits où cela est nécessaire. Il n'est malheureusement pas possible de raffiner plus par ce procédé sans changer la triangulation projetée de base. De plus, même en la modifiant et en effectuant des acquisitions successives de la scène avec des maillages plus adaptés — ce qui est contreproductif dans notre cas puisque pour éviter les erreurs liées au mouvement on doit se limiter à une unique acquisition — il reste la limite induite par l'épaisseur des arrêtes de la triangulation.

Afin de pouvoir raffiner successivement, il est donc nécessaire d'introduire des fonctions de formes décrivant la déformation des triangles nécessitant raffinement depuis un triangle

de référence. Cette transformation nous permet de déterminer les contours vraisemblables du motif de type triangulation s'il avait été adapté à la surface par un raffinement en ce point. Afin d'approximer au mieux, il est nécessaire de travailler avec des fonctions de formes d'ordre minimum 2. Cela est possible en utilisant les 3 sommets du triangle ainsi que les 3 milieux des arêtes et leur positions respectives après déformation.

Toutefois, il faut aussi tenir compte du fait que les fonctions de forme ainsi définies ne donnent qu'une approximation de la surface et que la surface réelle ne la suit pas nécessairement. Nous disposons cependant d'un outil intéressant à ne pas négliger qui est la géométrie épipolaire.

Dans le cadre du raffinement nous avons donc un triangle de référence T , un triangle déformé T_g acquis dans une vue I_g , pour lequel on a estimé une fonction de forme d'ordre 2 $f_{T \rightarrow T_g}$, et un triangle déformé T_d acquis dans une autre vue I_d , pour lequel on a aussi estimé une fonction de forme d'ordre 2 $f_{T \rightarrow T_d}$. Afin de raffiner, on introduit un sommet S au centre du triangle T . Les coordonnées de celui-ci dans I_g et I_d sont alors données par $S_g = f_{T \rightarrow T_g}(S)$ et $S_d = f_{T \rightarrow T_d}(S)$. La géométrie épipolaire ajoute que S_d doit être sur la droite épipolaire ℓ_{S_g} engendrée par S_g . Mais de par l'approximation faite lors de l'utilisation des fonctions de forme, ceci n'est plus nécessairement respecté. On peut cependant corriger la position du point S_d en le projetant sur ℓ_{S_g} en S'_d . Le couple (S_g, S'_d) respecte alors de nouveau la contrainte épipolaire. La géométrie étant symétrique, le couple (S'_g, S_d) la respecte aussi. Nous avons donc deux approximations de la réalité que l'on cherche à obtenir et que l'on notera ici (S_g^*, S_d^*) .

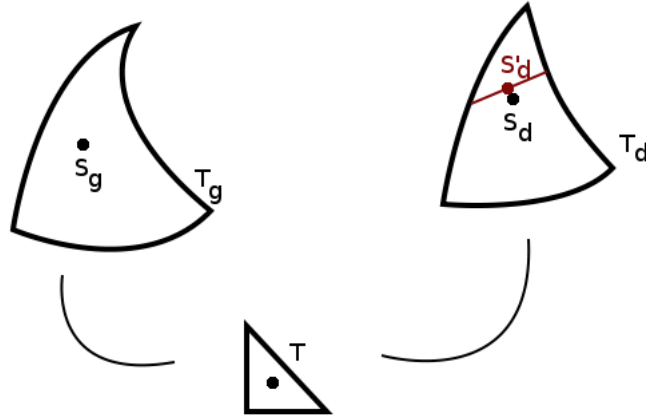


Figure V.4 – Correction de la fonction de forme par la géométrie épipolaire. En rouge, on voit la droite épipolaire engendrée par S_g .

Notons que la position de S_d sur le segment de ℓ_{S_g} inclus dans T_d n'est pas sûre. Nous avons introduit les fonctions de forme pour obtenir une solution proche de la réalité. Il n'est pas ici possible d'utiliser les techniques de corrélation classiques car nous travaillons à l'intérieur des

triangles et donc en dehors du motif lumineux. La luminosité émise par cette zone est donc très faible au sein du système et les capteurs d'acquisition souffrent alors d'un bruit gaussien amplifié. En ajoutant à cela l'homogénéité probable de la zone de recherche, les techniques de corrélation ne peuvent que donner ici de mauvais résultats. La méthode par fonction de forme reste donc notre meilleure approximation.

Grâce à elle, nous pouvons déterminer sur quelle droite épipolaire est sensée se trouver le point d'une vue par rapport à l'autre. Il est donc nécessaire de compléter l'information en s'appuyant sur les différentes vues afin d'augmenter le nombre de contraintes liées à la géométrie et d'optimiser l'approximation. Le même raisonnement s'effectue donc sur chaque couple de capteurs du bloc. Nous obtenons ainsi pour un bloc de quatre capteurs, 12 approximations de (S_g^*, S_d^*) .

Afin d'estimer correctement le point M^* engendré par la triangulation spatiale du couple (S_g^*, S_d^*) , nous avons opté pour une approximation statistique directement sur les points tridimensionnels $M_{(i,j)}$ engendrés par chaque couple (S'_i, S'_j) possible. Nous avons notamment envisagé d'utiliser la moyenne :

$$M^* \approx \sum_{\substack{i \in [1,4] \\ j \in [1,4] \\ i \neq j}} M_{(i,j)} \quad (\text{V.15})$$

Nous avons aussi envisagé d'utiliser la médiane. Pour cela, il faut pouvoir définir une relation d'ordre sur les points $M_{(i,j)}$. Au vu de l'application, qui est de déterminer une profondeur correcte en chaque point de la surface, il convient d'établir une relation d'ordre permettant de classer les éléments par rapport à leur profondeur. Pour cela il faut calculer la distance des points $M_{(i,j)}$ par rapport à la camera. Le problème est alors de choisir quelle caméra doit être la référente des caméras qui ont été utilisées lors de la triangulation spatiale. Nous avons contourné ce problème en fixant un point de référence au centre de la cabine, c'est à dire en $(0, 0, z)$. Ce point qui est alors situé au centre de la zone de reconstruction devrait être situé au centre de l'objet (dans le plan vertical). Ainsi la relation d'ordre RO utilisée est la comparaison de la distance entre les points $M_{(i,j)}$ et la droite verticale au centre de la cabine $(0, 0, z)$.

$$M^* \approx \text{MEDIANE}(\{M_{(i_1,j_1)} \dots M_{(i_{12},j_{12})}\}, \text{RO}) \quad (\text{V.16})$$

Nous n'avons cependant pas encore pu déterminer quelle approche donnait les résultats les plus précis. De plus d'autres approximations restent possibles.

3 Conclusion

La méthode présentée consiste donc à retrouver la position spatiale de chaque nœud de la triangulation projeté grâce à leurs images détectées. Au chapitre précédent, ces images ont été détectées comme points homologues. Afin de restituer la position spatiale, un calcul par bloc de capteur est effectué consistant à déterminer la position la plus probable du point tridimensionnel en connaissant approximativement les droites de projection de celui-ci dans les différents capteurs. Lors de la restitution spatiale de chaque nœud, les relation de voisinage déterminée sur la triangulation sont conservé. Elles peuvent être re-déterminée au besoin à partir du motif d'origine. La surface obtenue est donc non pas un nuage de point mais directement un ensemble de morceau de la surface de la scène. Afin d'obtenir une reconstruction globale, il faut alors recoller ces surfaces.

Le résultat obtenu est donc déjà une surface. Chaque point de cette surface est précisément localisé grâce au motif utilisé. La répartition de ces points n'est toutefois pas spécialement adaptée à la surface et il convient alors de raffiner cette dernière. Le raffinement présenté ne s'effectue que dans les zones où l'ajout de point impacte significativement la qualité de la reconstruction. Généralement, il s'agit des zones à forte courbure.

Les algorithmes d'assemblage des différents maillages et de raffinement n'ont pas pu être implémentés par manque de temps. Le développement d'une méthode de corrélation adaptée au projet BRAMMS a finalement demandé beaucoup plus de temps que prévu initialement.

Conclusion générale et perspectives

La thèse qui précède a donc débuté par un état de l'art complet des différentes technologies de mesures tridimensionnelles existantes. Cet état de l'art a été analysé dans le cas précis de notre étude. C'est à dire dans le cas où l'objet d'étude est le buste féminin. Ceci est directement lié au projet BRAMMS dans lequel s'inscrivent les présents travaux.

Nous avons vu au cours des premiers chapitres que ce sujet d'étude posait des contraintes supplémentaires en comparaison d'une mesure tridimensionnelle telle que pratiquée classiquement. Travailler sur du vivant — ou du mobile — c'est s'exposer aux mouvements. Ce problème implique que l'acquisition doit se faire très rapidement — si possible en une seule fois — afin de limiter les imprécisions induites par le mouvement. Il est à noter qu'un être vivant n'est jamais totalement immobile. En effet, le système nerveux génère en permanence des micro-mouvements afin de détecter d'éventuelles défaillances. Dans notre étude, nous nous sommes confrontés à un problème bien plus visible encore, lié à la respiration de la personne numérisée.

Lors de l'étude des technologies actuelles permettant l'acquisition de mesures tridimensionnelles, ce problème a été pris en compte. Nous avons pris comme postulat de départ la nécessité d'acquérir la scène rapidement et en une unique acquisition. Ceci est possible par le biais des méthodes utilisant des systèmes de capture d'image. Nous avons donc élaboré une cabine d'acquisition surfacique à partir d'appareils photographiques. Et nous avons dû de fait concevoir un système de déclenchement permettant l'acquisition de chaque appareil de capture en un même instant.

Un second problème imposé par l'aspect de la peau humaine a contraint la méthode utilisée à employer des projecteurs. En effet, la peau est uniforme en texture. Cela n'est généralement pas gênant sur un visage, riche en courbures diverses, mais c'est plus problématique dès lors que l'on se trouve dans une zone relativement plane. Il est très difficile par exemple de reconstruire le dos d'une personne par les techniques de stéréovision conventionnelles. Nous avons alors ajouté de l'information sur le sujet à acquérir en projetant un motif. Pour cela nous avons

Conclusion générale et perspectives

étudié en plus des appareils photographiques sur le marché, les différents projecteurs utilisables pour cette étude.

Le résultat de cette étude est donc la conception et la création d'une cabine d'acquisition photographiques à déclenchement simultané actuellement entièrement opérationnelle. Afin de prendre des mesures sur les surfaces de la scène à acquérir, il était nécessaire de développer un algorithme permettant la reconstruction de la scène avec ce dispositif. Plusieurs méthodes de reconstruction tridimensionnelle ont été implémentées et étudiées. Ces travaux ont été complétés par une stratégie consistant à projeter des motifs spécifiques afin d'apporter de l'information dans les zones homogènes de la scène.

Plusieurs motifs aléatoires ont été testés. Nous avons commencé par un nuage de tâches elliptiques dont l'orientation, les dimensions et la textures étaient aléatoires. Avec ce premier motif, nous avons essayé d'utiliser l'algorithme classique de reconstruction tridimensionnelle. Cette reconstruction comporte différentes étapes : la localisation des points d'intérêt, l'appariement de points d'intérêt entre images d'une même acquisition et enfin la triangulation spatiale de chacun de ces appariements. Le résultat de ces opérations est un nuage de points tridimensionnels. Il n'est pas forcément dense ni régulier, et ceci dépend directement de l'étape d'appariement.

L'appariement, qui est donc l'étape cruciale lors de la reconstruction tridimensionnelle, a été étudiée en profondeur dans le chapitre III. Il est apparu que la disposition des appareils d'acquisition était très importante au bon fonctionnement de cet algorithme de par la façon dont il aborde le problème. L'algorithme classique cherche des zones similaires dans l'image. Or, en cas de changement important dans la position des appareils photographiques lors d'une acquisition, de nombreux facteurs — notamment la position des sources lumineuses qui sont fixes et l'écrasement spatiale lors de la projection lorsque le capteur n'est pas perpendiculaire à la surface — entraîne un biais évident dans la recherche de zones similaires. On arrive fréquemment à la conclusion que le point homologue n'est pas dans la zone la plus similaire à celle contenant le point de départ. Les mauvais appariements induits par ce biais entraînent donc un mauvais départ pour l'étape suivante qui triangulera donc des points ne se correspondant pas pour arriver à un point tridimensionnel parfois très éloigné de la réalité.

Plusieurs mesures de corrélations et plusieurs contraintes sur l'algorithme ont été testées, sans toutefois parvenir à pallier ce problème.

Une solution simple aurait par exemple été d'augmenter drastiquement le nombre d'appareil d'acquisition. Cependant, outre la contrainte budgétaire, cette solution ne permet plus d'avoir une symétrie vis-à-vis de la source lumineuse induite par la projection. Or cela était une contrainte nécessaire afin de limiter le nombre d'artefact visuel dans les images obtenues

au cours des acquisitions. Nous avons donc décidé d'explorer d'autres pistes, plus complexes mais plus efficaces et pertinentes vis-à-vis du projet BRAMMS et ses impératifs, en cherchant non pas à déterminer des zones semblables, mais plutôt à construire un motif de façon intelligente. L'idée n'était plus tant d'avoir de l'information aléatoire projetée sur la scène, mais plutôt d'avoir une information structurée permettant d'être utilisée comme support dans la reconstruction.

Plusieurs motifs ont été testés mais un seul à su se démarquer de ses paires : celui correspondant à un maillage. Ce terme est cependant un peu abusif dans le sens où aucune métrique particulière n'est considéré. On peut donc ici tout aussi bien parler de pavage.

Le premier intérêt d'un maillage en tant que motif est qu'il existe des relations de voisinage entre ses éléments. Ceci est un critère déterminant qui permet d'établir un algorithme ne se reposant pas uniquement sur la ressemblance, mais aussi sur les résultats des voisins dans le maillage. Cela permet une utilisation même en cas d'écart important entre les capteurs d'image, ce qui est notre cas. L'algorithme ainsi modifié offre moins de chances de mauvais appariements, sous réserve que la triangulation ne forme pas un pavage trop régulier. En effet, un quadrillage amène certes de l'information de voisinage, mais il retire l'intérêt premier du motif qui était aussi d'apporter de l'information permettant la différenciation des points dans des zones normalement homogènes.

Le deuxième intérêt de ce choix est le fruit d'une étude approfondie de maillages plus aléatoires de type triangulation. Le chapitre IV est entièrement dédié à cette étude. C'est dans ce chapitre que nous amenons la majeure partie de nos travaux qui sortent de l'état de l'art actuel. L'analyse des triangulations nous a permis de créer une méthode originale consistant à générer des clefs associées aux différents nœuds telles que ces clefs soient uniques sur toute la triangulation.

Cette unicité est un avantage important par rapport aux méthodes existantes. En effet, il n'est dès lors plus nécessaire d'apparier des points d'intérêts par leur ressemblance, car le fait qu'ils soient ou non homologues est connu.

Enfin, le troisième intérêt d'un maillage de type triangulaire est que puisque nous connaissons déjà les relations de voisinage entre ses nœuds, nous n'obtenons pas un nuage de points qu'il faut ensuite traiter, afin de retrouver la surface, mais directement une surface.

Cependant, l'extraction d'une image déformée de triangulation n'est pas aisée et une méthode d'extraction a été nécessaire. Dans le chapitre IV, nous exposons les éléments fondamentaux à retrouver. Si pour la définition des clefs il était nécessaire de considérer une triangulation par ses nœuds, lors de son extraction c'est la recherche des triangles qui nous permettra cette fois de la récupérer. En effet, nous avons appuyé notre extraction sur la recherche des contours,

Conclusion générale et perspectives

recherche qui donne généralement de meilleurs résultats que la recherche de coins dans l'image afin de localiser les nœuds. En effet, de par leur nature aléatoire, les nœuds sont sujets à une forte variation de leur forme.

Lors de la présentation au chapitre III de l'algorithme d'appariement, nous avons expliqué son fonctionnement en commençant par une explication complète sur la géométrie épipolaire et la géométrie au sein d'un capteur d'image. La modélisation qui a été proposée pour le capteur d'image a été réutilisée afin de définir un détecteur de contours réellement adapté à notre problématique. Ce détecteur est présenté dans le chapitre IV. Lors de l'utilisation de la cabine, il y a peu de luminosité ambiante afin de pouvoir mieux faire ressortir le motif, et cela implique un bruit d'acquisition plus important. Le détecteur proposé permet de détecter les contours tout en atténuant le bruit gaussien intensifié par la faible luminosité. De plus, il permet de déterminer automatiquement des informations sur le type de contours rencontrés comme leur orientation.

Nous avons donc basé notre algorithme d'extraction sur ce détecteur de contour afin de détecter des contours fermés. Un traitement sur l'ensemble de ces contours permet donc de retrouver la triangulation. Le résultat n'est pas une triangulation complète, mais plusieurs parties de triangulation qu'il faut ensuite utiliser afin de construire des clefs uniques, puis de s'étendre par voisinage.

L'utilisation de ce motif provoque donc un gain de temps et de précision considérable. La recherche de points homologues est grandement simplifiée par les propriétés d'unicité du motif. Ces mêmes propriétés empêchent les mauvais appariements, principaux responsables des erreurs lors de la reconstruction tridimensionnelle. La reconstruction obtenue n'est de plus pas un nuage de points mais directement des morceaux de surfaces grâce encore une fois au motif utilisé.

Au final, nous avons donc conçu et supervisé la construction d'une cabine d'acquisition de surface tridimensionnelle par le biais d'appareils photographiques et de projecteurs. Un protocole d'acquisition a été mis au point, nécessitant une technique d'appariement spécifique basée sur un motif particulier dont l'extraction a été réalisée par l'élaboration d'un détecteur de contour résistant au bruit d'acquisition. La cabine est opérationnelle et se trouve actuellement dans les locaux de l'UTT.

Les travaux effectués sont sujet à amélioration. Des travaux futurs pour ce projet consisteraient à améliorer l'extraction du motif projeté. De plus, il est peut-être possible d'optimiser la méthode de raffinement envisagée.

J'ajouterai en ultime conclusion que lors de la fin de la rédaction de ce manuscrit, je me suis rendu compte de l'existence d'une thèse similaire qui a eu lieu en parallèle. Les résultats de la thèse de Christian Piccolo sont brièvement décrits à la fin du premier chapitre. Même si

nos avis sur la méthode à employer diffèrent, la similarité des cabines développées en termes de matériel et d'approche globale reste un point intéressant à souligner.

Annexe A : Détecteur de contours

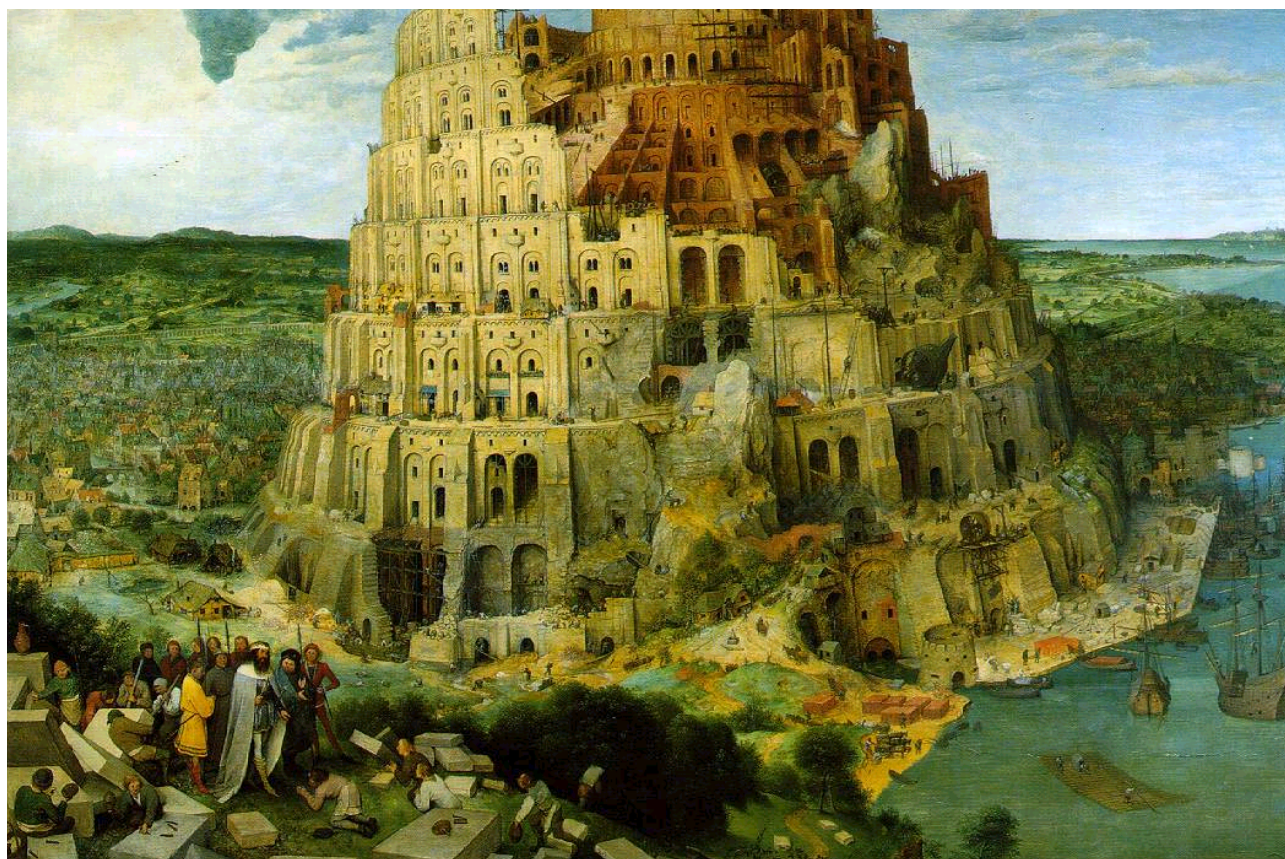


Figure A.1 – Brueghel : *La tour de Babel*, 1563. Image utilisée pour la détection suivante, voire figure [A.2](#).

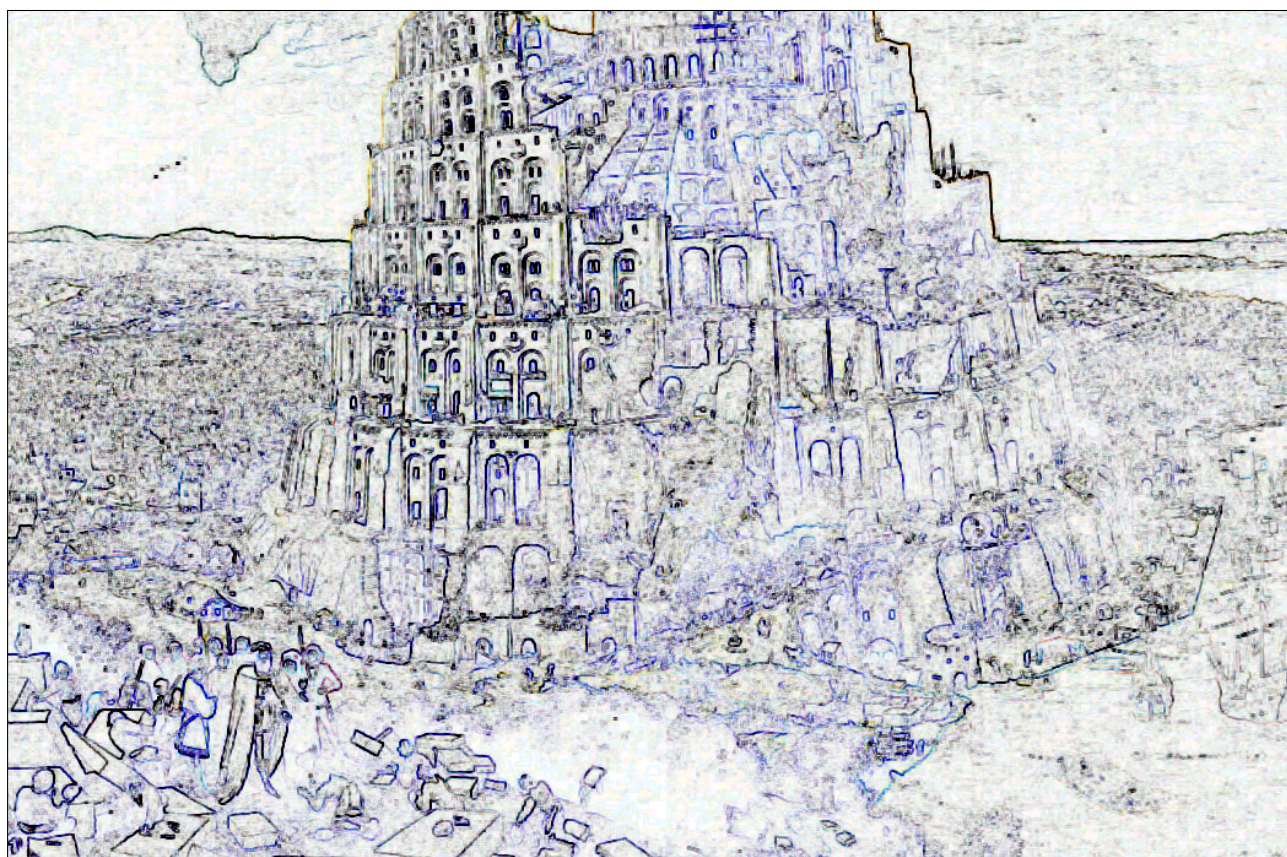


Figure A.2 – Les contours sont extrait à l'aide d'une quadrique dans un voisinage 3×3 depuis [A.1](#).

Annexe C : Documentation de la cabine

Ceci est la documentation technique de la cabine d’acquisition. Pour de plus amples détails concernant les choix technologiques faits, le lecteur est prié de se reporter à la thèse sur le sujet, chapitre [II](#).

B.1 L’agencement de la cabine

Nous abordons donc ici la partie purement matérielle de la cabine. La partie logicielle sera traitée par la suite.

B.1.1 Appareils de captures d’images

La cabine est équipée de 24 appareils photographiques numériques *reflex* CANON 600D. Ces appareils ont été choisis en fonction des contraintes budgétaires et techniques nécessaires à l’élaboration du processus de reconstruction. Chaque appareil est équipé d’un objectif à focale fixe “Canon EF 50mm f/1.6 II”, la focale fixe étant un critère déterminant lors du processus de *calibration*.

Les caractéristiques de chaque appareil et de chaque monture sont donc les suivantes :

- ISO : Auto ou 100 à 6400 (12800 théorique)
- Vitesse d’obturation : 1/4000sec à 30sec plus mode BULB (ouvert de la durée de l’appui).
- Taille du capteur CCD : $22,3 \times 14,9$
- Résolution maximale : 5182×4356
- Images en mode rafale : 3,7 par seconde.
- Focale : 50mm
- Ouverture minimum : 22
- Ouverture maximal : 1,8

Annexe C : Documentation de la cabine

- Angle de champ diagonale : 46°
- Champ à 1m : $0,47\text{m} \times 0,30\text{m}$
- Champ à 1,5m : $0,7\text{m} \times 0,46\text{m}$

Il est recommandé d'utiliser chaque appareil en mode manuel avec les réglages suivants dans le cadre de l'utilisation de la cabine :

- ISO : 6400
- Vitesse d'obturation : $1/4\text{sec}$
- Ouverture : 2
- Résolution : maximale non compressée (RAW voir RAW+L mais perte inutile d'espace)

Les 24 appareils sont fixés à la structure métallique sur deux étages par le pas de vis utilisé normalement pour leur pied. Chacun d'eux possède un adaptateur pour l'alimenter directement sur le secteur, et une carte de stockage (SD). Il est donc nécessaire que les appareils soit branchés sur le secteur, possède leur carte de stockage avec le formatage adéquat (voir section B.2.3) et soit reliés via USB à l'ordinateur et via prise jack au système de déclenchement.

Avant chaque utilisation de la cabine, il est nécessaire de vérifier que chaque appareil est bien allumé et que les caches des objectifs aient été retirés.

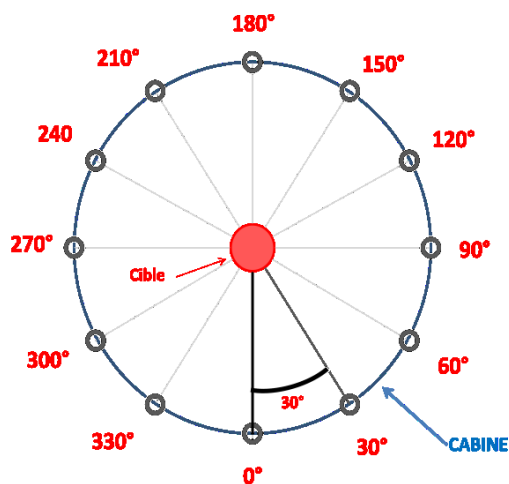


Figure B.1 – Disposition des appareils photographiques numériques. Diamètre de la cabine : 3m. Diamètre de la zone de reconstruction (centrale) : 1m.

B.1.2 Appareils de projection

La cabine est équipée de 6 vidéoprojecteurs PHILIPS PPX 1430. Ce choix a été déterminé en fonction des différentes technologies de vidéoprojection disponibles actuellement sur le marché

et de nos besoins.

Leurs caractéristiques sont les suivantes :

- Contraste : 500:1
- Luminosité : 301um
- Résolution : 1280×768
- Distance d’affichage : 0,2 à 3m
- Technologie : LCoS

Chaque vidéoprojecteur possède une carte mémoire (SD) stockant le motif à projeter sur la scène. Normalement, les motifs doivent être alternés dans leur couleur.

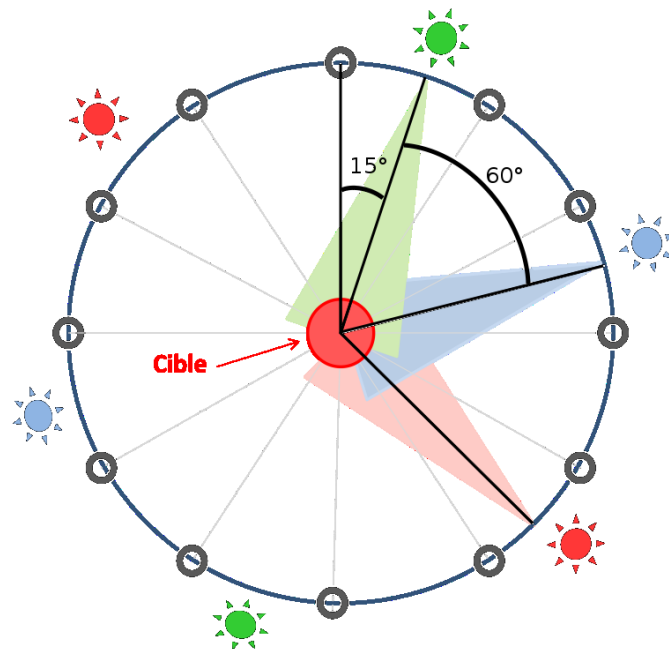


Figure B.2 – Disposition des vidéo-projecteurs.

B.1.3 Système de déclenchement

Les appareils 600D sont équipés d'un port pour périphérique de déclenchement. Bien qu'il soit possible de déclencher l'acquisition par un appareil via USB, il est impossible d'en déclencher plusieurs simultanément. Un dispositif de déclenchement a donc été élaboré à partir d'une unité de commande Arduino qui simule l'appui sur la télécommande de déclenchement de l'appareil photographique, et qui peut donc en déclencher plusieurs en même temps. Ce dispositif permet de plus d'empêcher l'autofocus qui se déclenche normalement automatiquement avant

Annexe C : Documentation de la cabine

une acquisition. L'Arduino supporte donc un montage particulier permettant d'effectuer cette simulation. Cette plaquette est située au-dessus de la cabine.

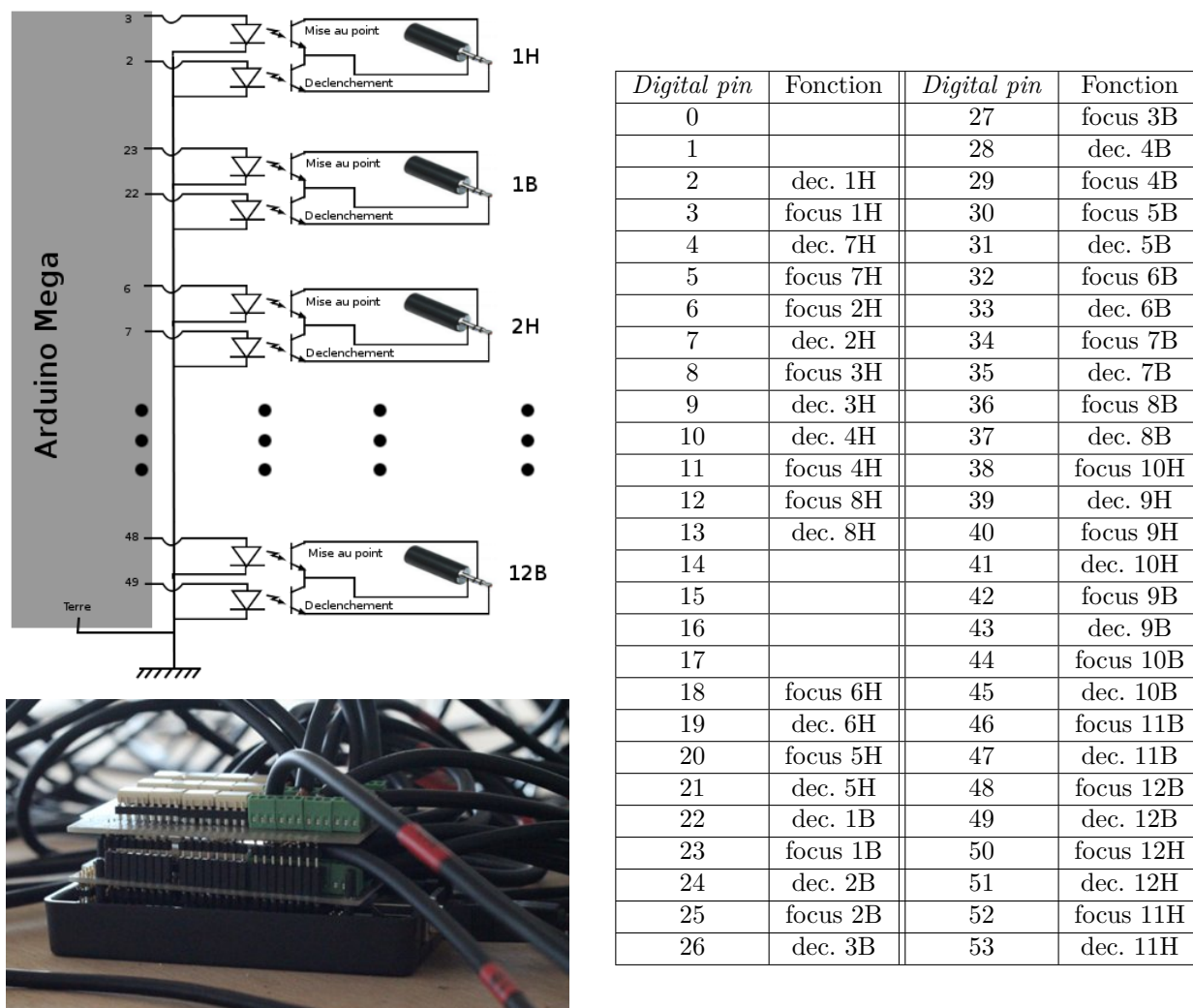


Figure B.3 – Circuit de déclenchement simultané utilisant une interface Arduino et brochage de ses broches de type digitale (*digital pin* sur l'Arduino). “dec.” correspond à la commande de déclenchement et “focus” à la commande de mise au point. Les appareils sont notés en fonction du numéro de leur colonne et de la hauteur sur celle-ci (“H” pour haut et “B” pour bas).

Pour bloquer l'autofocus, il faut que l'Arduino exécute l'opération “déclencher le focus de l'appareil X” qui revient dans notre montage à mettre la sortie correspondante à 1. Ensuite, après un temps fixé relativement long, laissant tous les appareils faire leur mise au point, l'ordre “déclencher l'acquisition de l'appareil X” qui revient encore une fois à mettre la sortie correspondante à 1 doit être effectué. Enfin, on peut remettre à zéro toutes les sorties.

B.1.4 Disposition

Afin de pouvoir déplacer aisément la cabine ou la modifier pour d'autres applications, elle a été pensée de façon modulaire dès sa conception, en utilisant les composants universels du système modulaire de PROFILES de Bosch Rexroth, voir Figure B.4. Ces composants ont divers intérêts tels que la facilité et la rapidité de montage, tout en offrant une structure solide et facilement démontable.

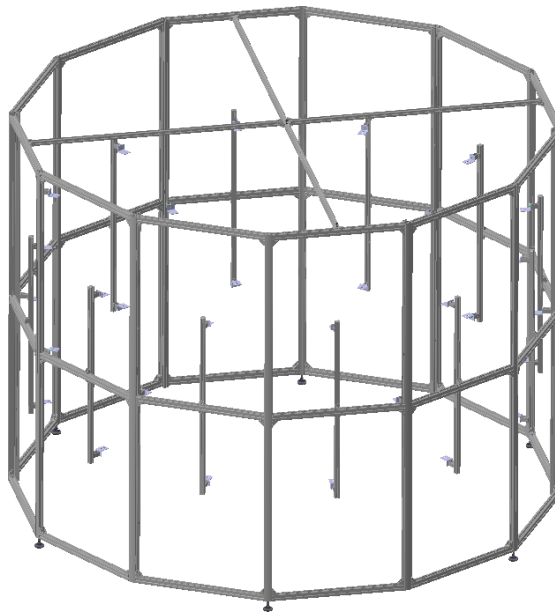


Figure B.4 – Structure de la cabine.

Afin de pouvoir aisément faire rentrer les objets constitutifs de la scène à numériser, l'un des pans du cylindre à douze faces de la cabine est amovible, et permet un basculement selon un axe vertical vers l'intérieur. Les interactions avec l'extérieur de la cabine, essentiellement lumineuses, ont été limitées à l'aide d'un tissu noir épais entourant la cabine. La pose d'un plafond en contreplaqué a permis de finir l'isolement optique du système, tout en ayant l'avantage d'offrir un support stable à l'ensemble de la câblerie. De plus, des prises de courant ont été aménagées au sein de la structure afin de pouvoir brancher les appareils de captures et de projection directement sur secteur. Ainsi, seuls trois câbles relient les prises internes de la cabine au réseau électrique, limitant donc les manipulations nécessaires à sa mise en route.

La liaison entre la cabine et l'ordinateur est assurée via USB. Pour des considérations techniques, une topologie en étoile a été réalisée. A cette fin, deux *hub* USB ont été fixés sur le toit de la cabine. Il y a donc trois câbles USB qui relient la cabine à l'ordinateur : un pour le

système de déclenchement et deux pour la transition des photos acquises.

Enfin, le toit de la cabine est équipé d'une source lumineuse afin de permettre l'installation de la scène. Un système de ventilation y a aussi été intégré afin de pallier la chaleur dégagée par l'ensemble du dispositif. Ces deux ajouts sont pilotés par deux interrupteurs situés près de la porte de la cabine.

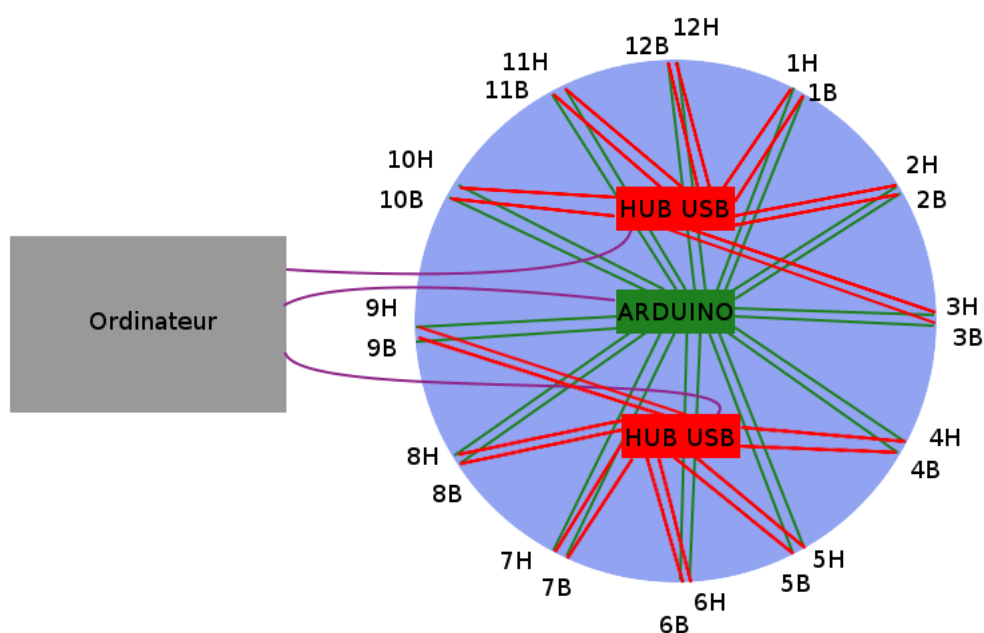


Figure B.5 – Cablage de la cabine.

B.2 L'interface

Nous traitons ici de la partie logicielle de la cabine. C'est-à-dire son interface de pilotage. Le matériel utilisé est décrit dans la section précédente.

B.2.1 IHM

La bibliothèque EDSDK 2.1 permet l'accès à l'appareil photographique et à ses fonctions. L'appareil dispose via ce biais des mêmes options de configuration qu'en mode manuel. Il est donc possible de le configurer entièrement par l'interface qui utilise cette bibliothèque. Une configuration identique des appareils étant un point important, l'interface qui a été développée pour le pilotage de la cabine via l'ordinateur se devait aussi de respecter cette contrainte. Pour

ce faire, chaque modification des paramètres d'acquisition entrée via l'interface (Figure B.6) est reportée sur chaque appareil photographique.

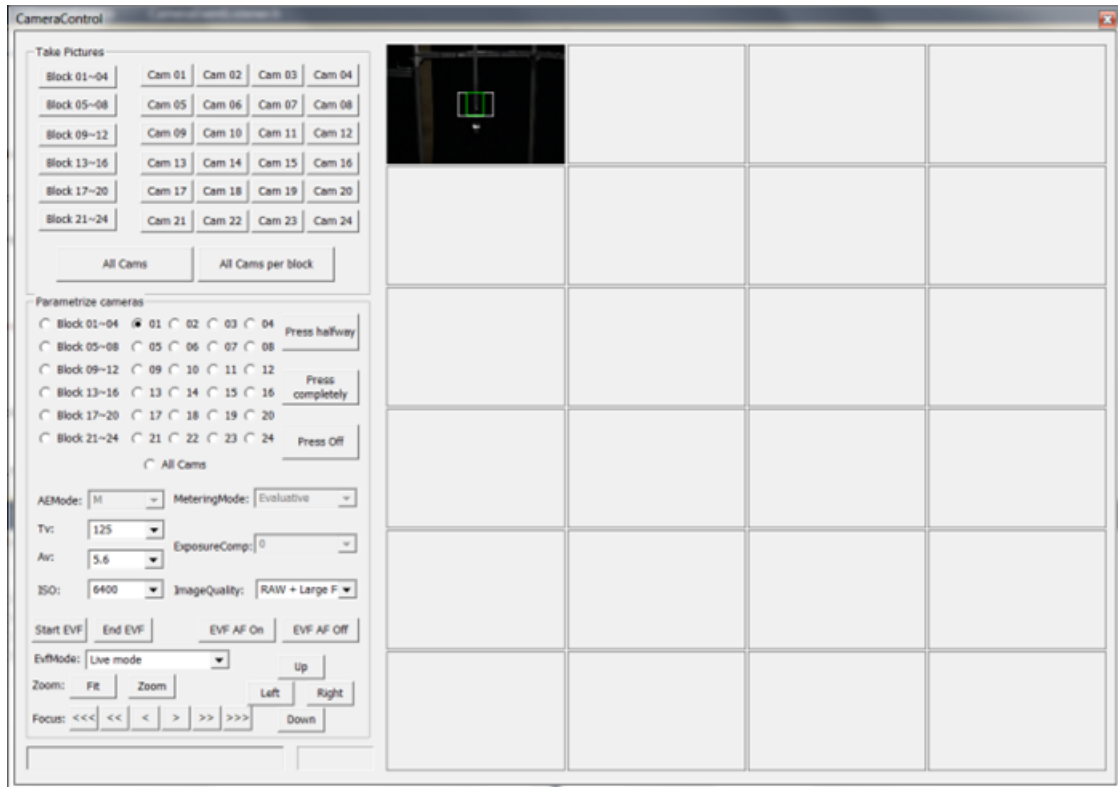


Figure B.6 – Interface. Mode EVF détourné et affiché.

Le mode EVF¹, qui permet normalement d'afficher sur l'écran LCD d'un appareil photographique ce qui est dans le champ optique de l'appareil, peut être détourné pour être affiché directement à l'écran. Même s'il est malheureusement impossible de le faire pour plus d'un appareil à la fois, il est possible de s'en servir sur un appareil (par défaut, celui sur la porte en haut), voir Figure B.6. Il est ainsi possible de contrôler ce qui se passe à l'intérieur de la cabine. Ceci peut, entre autre, être utilisé afin de ne pas déclencher d'acquisition inopportune si les conditions de prise ne sont pas réunies.

A terme, l'interface devrait aussi assurer le rapatriement des données qui est pour l'instant assuré par une autre routine. Il est techniquement possible de récupérer une miniature de la dernière photographie sur chaque appareil bien que cela soit couteux en temps. Cette fonctionnalité a donc été abandonnée, car elle empêche l'acquisition rapide de plusieurs scènes successives.

1. *Electronic View Finder*

B.2.2 Calibration

Afin d’obtenir avec précision la géométrie exacte qui relie chaque capteur — paramètres extrinsèques (position et orientation relatives) — ainsi que la géométrie interne de chaque capteur — paramètres intrinsèques (focale, coordonnées du centre optique sur le capteur, facteur d’orthogonalité) — il est nécessaire de procéder à une étape de *calibration* de la cabine d’acquisition. La précision de ces paramètres influe directement sur la précision de la reconstruction tridimensionnelle. C’est donc une étape capitale dans le procédé.

Il n’est toutefois pas nécessaire d’effectuer cette étape avant chaque acquisition. Cette opération n’est nécessaire qu’après chaque changement géométrique. Cependant, comme dans notre cas nous utilisons des focales fixes, cela correspond à un changement de position ou d’orientation des appareils. Il convient toutefois de réévaluer régulièrement les appareils photographiques situés sur la porte de la cabine. Ces derniers sont soumis à plus de vibrations et ont donc plus facilement tendance à se dérégler en orientation.

Afin d’étalonner le système, plusieurs méthodes existent. Nous avons choisi d’utiliser une mire plane de type damier. Son utilisation dans un ensemble de vues permet de déterminer la position relative des appareils les uns par rapport aux autres ainsi que d’estimer ses paramètres internes. Afin d’obtenir les paramètres, la “*Camera Calibration Toolbox*” [Bou10] est utilisée.

Il faut donc prendre une série d’acquisition avec une mire se déplaçant dans la cabine. Afin d’avoir une certaine précision, il est nécessaire de prendre plus d’acquisitions qu’il n’y a d’appareils photographiques. Il faut aussi que la mire se déplace entre chacune des acquisitions, qu’elles soient toujours visibles par au moins deux appareils de capture.

B.2.3 Identification automatique des appareils photographiques et arborescence

Afin d’interagir avec les appareils d’acquisitions, l’interface inspecte à son lancement les appareils connectés via USB à l’ordinateur. Les interfaces de ces appareils sont stockées dans un tableau de données nous permettant d’y avoir accès séparément. Il est toutefois plus problématique de déterminer l’ordre dans lequel les appareils ont été acquis. En effet, cet ordre peut varier à chaque démarrage.

La conception des appareils 600D fait qu’il n’est pas possible d’ouvrir un document créé par un autre moyen que par la prise de photographie par ledit appareil, lorsque celui-ci est vu comme un périphérique de stockage de données par l’ordinateur. N’ayant pas directement accès au format de stockage “CR2” qui est le format de donnée brut de Canon, il n’est pas possible

de stocker l'information au sein des pixels. De même, malgré le fait que le 600D puisse lire directement des données au format de compression "JPEG", il est impossible de stocker des informations précises dans les pixels de l'image. En effet, lors de la compression, ces données sont altérées, et il n'est alors plus possible de différencier les appareils les uns des autres, et ce même avec un très faible niveau de compression. La seule possibilité trouvée à l'heure actuelle consiste à modifier une image créée lors d'une photographie par l'appareil concerné. La modification à faire consiste à modifier les dimensions de l'image car celles-ci ne sont nullement altérées par les formats de compression. De plus cette image doit pouvoir être sauvegardée à part, afin de ne pas être effacée par erreur, ce qui est heureusement possible en créant un répertoire de stockage séparé pour cette image d'identification. Le codage utilisé est donc le suivant : l'image `"/DCIM/101CANON/IMG_0001.JPG"` possède l lignes et c colonnes. c correspond au numéro de la colonne de l'appareil photographique. l est mis à 1 si ce sont les appareils du bas et à 2 sinon.

Lors de la récupération des images, il convient d'instaurer une architecture dans l'arborescence. Ainsi, cette arborescence contient des dossiers de type `"NA"` où N est le numéro de la colonne en prenant comme norme le sens direct partant de la colonne sur la porte d'accès de la cabine (celle-ci porte donc la valeur "1") et A correspond soit à B si c'est un appareil du bas ou à H si c'est un appareil du haut. Lors du transfert, une renumérotation des fichiers est à effectuer afin de toujours commencer par le fichier 1.

B.2.4 Flux de récupération des données

Lors de l'acquisition, les photographies sont d'abord stockées sur les cartes de stockage de chaque appareil. Il est donc possible de prendre plusieurs scènes avant de les transférer sur l'ordinateur. Le transfert est une opération qui peut être longue. Sur une campagne d'acquisition d'une quarantaine de scènes, le transfert des photographies peut prendre deux heures lors d'un traitement manuel. Il est donc nécessaire d'automatiser ce procédé en profitant du fait que chaque appareils photographique est connecté via USB à l'ordinateur, et de l'accélérer au maximum.

L'identification de chaque appareil est décrite dans la partie précédente. Il ne reste donc qu'à transférer les photographies dans une arborescence correspondant à notre cabine (voir section précédente : [B.2.3](#)). Puis on procède à l'effacement desdites acquisitions sur les cartes mémoires des appareils photographiques.

Une fois les photographies transférées, tout est en place pour assurer le bon déroulement de la partie algorithmique de la cabine. Ce transfert peut toutefois être optimisé par une attention

porté aux spécifications de l'ordinateur fonctionnant de pair avec la cabine. Outre le fait de préférer de l'USB3 à l'USB2 et l'USB2 à l'USB1, il faut aussi vérifier que les deux ports USB utilisés — nous parlons ici des deux câbles USB reliés aux appareils photographiques, il n'y a aucun prérequis pour celui relié à l'interface de déclenchement — correspondent à des ports USB matériels différents (à ne pas confondre avec la notion de port physique, nous parlons ici de sorties gérées séparément sur la carte-mère, et non de ports de sortie différents. En effet, une sortie USB sur carte mère — port matériel — peut donner naissance à plusieurs sorties extérieures de connexion — port physique. Si les ports matériels sont différents, ils seront gérés par des *thread* (file d'exécution) différents ce qui permettra de diviser le temps de transfert des données par deux.

Bibliographie

- [AEM11] Christian D Austin, Emre Ertin, and Randolph L Moses. Sparse signal methods for 3-d radar imaging. *Selected Topics in Signal Processing, IEEE Journal of*, 5(3) :408–423, 2011. [10](#)
- [AP79] I.E. Abdou and W. Pratt. Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE*, 67(5) :753–763, 1979. [97](#), [98](#)
- [BB01] Jasmine Banks and Mohammed Bennamoun. Reliability analysis of the rank transform for stereo matching. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 31(6) :870–880, 2001. [58](#)
- [BBH03] Myron Z Brown, Darius Burschka, and Gregory D Hager. Advances in computational stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(8) :993–1008, 2003. [48](#)
- [BCPF00] T Belli, M Cord, and S Philipp-Foliguet. Colour contribution for stereo image matching. In *International Conference on Color in Graphics and Image Processing*, pages 317–322. Citeseer, 2000. [61](#)
- [Bes88] Paul J Besl. Active, optical range imaging sensors. *Machine vision and applications*, 1(2) :127–152, 1988. [7](#), [11](#)
- [BETVG08] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3) :346–359, 2008. [61](#)
- [Beu90] Serge Beucher. *Segmentation d’images et morphologie mathématique*. PhD thesis, École Nationale Supérieure des Mines de Paris, 1990. [106](#)
- [BGM⁺04] I Bloch, Y Gousseau, H Matre, D Matignon, B Pesquet-Popescu, F Schmitt, M Sigelle, and F Tupin. Le traitement des images. *Polycopié du cours ANIM, Département TSI-Télécom-Paris*, 2004. [85](#)

Bibliographie

- [BL05] Matthew Brown and David G Lowe. Unsupervised 3d object recognition and reconstruction in unordered datasets. In *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, pages 56–63. IEEE, 2005. 61
- [Bou10] Jean-Yves Bouguet. Camera calibration toolbox for matlab, 2010. http://www.vision.caltech.edu/bouguetj/calib_doc/. 34, 36, 146
- [BTVG06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf : Speeded up robust features. In *Computer Vision–ECCV 2006*, pages 404–417. Springer, 2006. 61, 62
- [Can86a] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6) :679–698, 1986. 61
- [Can86b] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6) :679–698, nov. 1986. 97, 99, 100, 101, 103, 104
- [Cha05] Sylvie Chambon. *Mise en correspondance stéréoscopique d’images couleur en présence d’occultations*. PhD thesis, Université toulouse III-Paul Sabatier, 2005. 63, 64
- [CJ04] Gustavo Carneiro and Allan D Jepson. Pruning local feature correspondences using shape context. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 16–19. IEEE, 2004. 62
- [CMP96] Alain Crouzil and Serge Massip-Pailhes, Louis ans Castan. A new correlation criterion based on gradient fields similarity. In *Pattern Recognition, Proceedings of the 13th International Conference on*, volume 1, pages 632–636. IEEE, 1996. 56, 66
- [Col71] Robert Collier. *Optical holography*. Elsevier, 1971. 9
- [Cox95] Greg S Cox. Template matching and measures of match in image processing. *University of Cape Town, South Africa*, 1995. 65
- [CR12] Rimi Cогranne and Florent Retraint. Statistical Detection of Defects Using a Parametric Adapted Model of Tomography Projections. submitted to *Medical Imaging, IEEE Transactions on*, 2012. 86
- [Cut90] LJ Cutrona. Synthetic aperture radar. *Radar Handbook, second edition*, ed. M. Skolnik, McGraw-Hill, New York, 1990. 9
- [CZF⁺11] Rimi Cогranne, Cathel Zitzmann, Lionel Fillatre, Igor Nikiforov, Florent Retraint, and Philippe Cornu. A cover image model for reliable steganalysis. In *Information Hiding*, vol.6958 of Lecture Notes in Computer Science, pages 178 – 192, Prague, Czech Republic, May 18–20, 2011. 86

- [CZR⁺12] Rimi Cогranne, Cathel Zitzmann, Florent Retraint, Igor Nikiforov, and Philippe Cornu. A Non-linear Local Model of Natural Images for Reliably Detect Hidden Information. submitted to *Image Processing, IEEE Transactions on*, 2012. 86
- [Dem02] D. Demigny. On optimal linear filtering for edge detection. *Image Processing, IEEE Transactions on*, 11(7) :728–737, 2002. 98
- [DRR03] James Davis, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Spacetime stereo : A unifying framework for depth from triangulation. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–359. IEEE, 2003. 71
- [Eng10] TOLA Engin. *DAISY : A Fast Descriptor for Dense Wide Baseline Stereo and Multiview Reconstruction*. PhD thesis, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, 2010. 61
- [Fau95] Olivier Faugeras. Stratification of three-dimensional vision : projective, affine, and metric representations. *JOSA A*, 12(3) :465–484, 1995. 118
- [FB81] Martin A Fischler and Robert C Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, 1981. 35
- [FJM92] Olivier D Faugeras, Q T Juong, and Stephen J Maybank. Camera self-calibration : Theory and experiments. In *Computer Vision ECCV’92*, pages 321–334. Springer, 1992. 35
- [Gab48] Dennis Gabor. A new microscopic principle. *Nature*, 161(4098) :777–778, 1948. 9
- [Gar01] Dorian Garcia. *Mesure de formes et de champs de déplacements tridimensionnels par stéréo-corrélation d’images*. PhD thesis, Institut National Polytechnique de Toulouse-INPT, 2001. 64
- [GBTD⁺09] Manuel Grand-Brochier, Christophe Tilmant, Michel Dhome, et al. Descripteur local d’image invariant aux transformations affines. In *ORASIS’09-Congrès des jeunes chercheurs en vision par ordinateur*, 2009. 61
- [Gen11] Jason Geng. Structured-light 3 d surface imaging : a tutorial. *Advances in Optics and Photonics*, 3(2) :128–160, 2011. 72
- [GNG⁺08] Jinwei Gu, Shree Nayar, Eitan Grinspun, Peter Belhumeur, and Ravi Ramamoorthi. Compressive structured light for recovering inhomogeneous participating media. In *Computer Vision–ECCV 2008*, pages 845–858. Springer, 2008. 71
- [GNY92] Paul M Griffin, Lakshmi S Narasimhan, and Soung R Yee. Generation of uniquely encoded light patterns for range data acquisition. *Pattern recognition*, 25(6) :609–616, 1992. 71

Bibliographie

- [Goo05] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company Publishers, 2005. [63](#)
- [Har84] Robert M. Haralick. Digital step edges from zero crossing of second directional derivatives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-6(1) :58–68, jan. 1984. [91](#), [97](#), [99](#), [100](#), [101](#)
- [Har92a] Kevin G Harding. Phase-grating use for slope discrimination in moiré contouring. In *Robotics-DL tentative*, pages 265–270. International Society for Optics and Photonics, 1992. [71](#)
- [Har92b] Richard I Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Computer Vision ECCV'92*, pages 579–587. Springer, 1992. [35](#), [119](#)
- [Har94] Richard I Hartley. An algorithm for self calibration from several views. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 908–912. IEEE, 1994. [122](#)
- [Har97] Richard I Hartley. In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(6) :580–593, 1997. [35](#)
- [Har99] Richard I Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2) :115–127, 1999. [54](#)
- [Hen75] Charles R Henderson. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, pages 423–447, 1975. [35](#)
- [HHJC99] Peisen S Huang, Qingying Hu, Feng Jin, and Fu-Pen Chiang. Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Optical Engineering*, 38(6) :1065–1071, 1999. [71](#)
- [HHR01] Olaf Hall-Holt and Szymon Rusinkiewicz. Stripe boundary codes for real-time structured-light range scanning of moving objects. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 359–366. IEEE, 2001. [71](#)
- [HKLP09] Vu Hoang Hiep, Renaud Keriven, Patrick Labatut, and J-P Pons. Towards high-resolution large-scale multi-view stereo. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1430–1437. IEEE, 2009. [47](#)
- [HM86] Andres Huertas and Gerard Medioni. Detection of intensity changes with subpixel accuracy using laplacian-gaussian masks. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(5) :651–664, sept. 1986. [97](#), [99](#), [100](#), [101](#), [103](#), [104](#)

-
- [HM95] Radu Horaud and Olivier Monga. *Vision par ordinateur : outils fondamentaux*. Traité des nouvelles technologies, Série informatique. Editions Hermès, 1995. [57](#), [58](#)
 - [HS88a] C. Harris and M. Stephens. A combined corner and edge detector. In *Fourth Alvey Vision Conference, Manchester, UK*, pages 147 – 151, 1988. [95](#), [96](#)
 - [HS88b] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988. [61](#)
 - [HS97] Richard I Hartley and Peter Sturm. Triangulation. *Computer vision and image understanding*, 68(2) :146–157, 1997. [123](#)
 - [HSH05] Pär Hammarstedt, Peter Sturm, and Anders Heyden. Degenerate cases and closed-form solutions for camera calibration with one-dimensional objects. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 317–324. IEEE, 2005. [34](#)
 - [HSSB96] M. Heath, S. Sarkar, T. Sanocki, and K. Bowyer. Comparison of edge detectors : a methodology and initial study. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, pages 143 –148, jun 1996. [97](#)
 - [HZ04] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*, volume 2. Cambridge University Press, 2004. [46](#), [121](#)
 - [IG98] Hiroshi Ishikawa and David Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *Computer Vision ?ECCV'98*, pages 232–248. Springer, 1998. [61](#)
 - [JKC85] Jeffrey A Jalkio, Richard C Kim, and Steven K Case. Three dimensional inspection using multistribe structured light. *Optical Engineering*, 24(6) :246966–246966, 1985. [14](#), [71](#)
 - [JKS95] Ramesh Jain, Rangachar Kasturi, and Brian G. Schunck. *Machine vision*. McGraw-Hill, Inc., New York, NY, USA, 1995. [85](#), [98](#)
 - [JLW⁺06] Xiaogang Jin, Juncong Lin, Charlie CL Wang, Jieqing Feng, and Hanqiu Sun. Mesh fusion using functional blending on topologically incompatible sections. *The Visual Computer*, 22(4) :266–275, 2006. [125](#)
 - [Jon97] G.A. Jones. Constraint, optimization, and hierarchy : Reviewing stereoscopic correspondence of complex features. *Computer Vision and Image Understanding*, 65(1) :57–78, 1997. [52](#)
 - [Kho11a] Kourosh Khoshelham. Accuracy analysis of kinect depth data. In *ISPRS workshop laser scanning*, volume 38, page 1, 2011. [12](#)

Bibliographie

- [Kho11b] Kourosh Khoshelham. Accuracy analysis of kinect depth data. In *ISPRS workshop laser scanning*, volume 38, page 1, 2011. [72](#)
- [KLCL05] Jae Chul Kim, Kyoung Mu Lee, Byoung Tae Choi, and Sang Uk Lee. A dense stereo matching using two-pass dynamic programming with generalized ground control points. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1075–1082. IEEE, 2005. [48](#)
- [KPT77] RJ Kauth, AP Pentland, and GS Thomas. Blob-an unsupervised clustering approach to spatial preprocessing of mss imagery. 1977. [61](#)
- [KRS04] Alexander Kaplan, Ehud Rivlin, and Ilan Shimshoni. Robust feature matching across widely separated color images. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 136–139. IEEE, 2004. [61](#)
- [Kru13] Erwin Kruppa. *Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung*. Hölder, 1913. [35](#)
- [KS03] Jana Kostková and Radim Sára. Stratified dense matching for stereopsis in complex scenes. In *BMVC*, volume 5, page 6. Citeseer, 2003. [56](#)
- [KSMK99] Takashi Kanai, Hiromasa Suzuki, Jun Mitani, and Fumihiko Kimura. Interactive mesh fusion based on local 3d metamorphosis. In *Graphics interface*, volume 99, pages 148–156, 1999. [125](#)
- [Lag11] Robert Laganière. *OpenCV 2 computer vision application programming cookbook*. Packt Publishing, 2011. [111](#)
- [LG96] CF Van Loan and GH Golub. Matrix computations, 1996. [89](#)
- [LH87] HC Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Readings in Computer Vision : Issues, Problems, Principles, and Paradigms*, MA Fischler and O. Firschein, eds, pages 61–62, 1987. [35](#)
- [LH06] Hongdong Li and Richard Hartley. Five-point motion estimation made easy. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 630–633. IEEE, 2006. [36](#)
- [LJ77] RA Lewis and AR Johnston. A scanning laser rangefinder for a robotic vehicle. In *IJCAI*, pages 762–768, 1977. [7](#)
- [LMW88] JJ Le Moigne and Allen Mark Waxman. Structured light patterns for robot mobility. *Robotics and Automation, IEEE Journal of*, 4(5) :541–548, 1988. [71](#)
- [Low04] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2) :91–110, 2004. [61](#), [62](#)

- [LPMV10] Ruding Lou, Jean-Philippe Pernot, Alexei Mikchevitch, and Philippe Véron. Merging enriched finite element triangle meshes for fast prototyping of alternate solutions in the context of industrial maintenance. *Computer-Aided Design*, 42(8) :670–681, 2010. [125](#)
- [LQ05] Maxime Lhuillier and Long Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(3) :418–433, 2005. [61](#)
- [LT10] O. Laligant and F. Truchetet. A nonlinear derivative scheme applied to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(2) :242–257, feb. 2010. [97](#), [99](#), [100](#), [101](#), [102](#), [103](#)
- [LU62] Emmett N Leith and Juris Upatnieks. Reconstructed wavefronts and communication theory. *JOSA*, 52(10) :1123–1128, 1962. [9](#)
- [LVD98] Jean-Marc Lavest, Marc Viala, and Michel Dhome. Do we really need an accurate calibration pattern to achieve a reliable camera calibration? In *Computer Vision ECCV’98*, pages 158–174. 1998. [34](#)
- [MC04] Zoltán Megyesi and Dmitry Chetverikov. Affine propagation for surface reconstruction in wide baseline stereo. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 76–79. IEEE, 2004. [61](#)
- [ML96] Samuel Matej and Robert M Lewitt. Practical considerations for 3-d image reconstruction using spherically symmetric volume elements. *Medical Imaging, IEEE Transactions on*, 15(1) :68–78, 1996. [61](#)
- [Mor80] Hans P Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical report, DTIC Document, 1980. [65](#)
- [MS89] Mumford, D. and Shah, J. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5) :577–685, 1989. [86](#)
- [Nac75] Myron L Nack. Temporal registration of multispectral digital satellite images using their edge images. 1975. [66](#)
- [Nis84] Herbert Keith Nishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5) :235536–235536, 1984. [66](#)
- [OCG⁺07] J-J Orteu, T Cutard, D Garcia, E Cailleux, and L Robert. Application of stereo-vision to the mechanical characterisation of ceramic refractories reinforced with metallic fibres. *Strain*, 43(2) :96–108, 2007. [47](#)

Bibliographie

- [OFN12] Yuji Oyamada, Pascal Fallavollita, and Nassir Navab. Single camera calibration using partially visible calibration objects based on random dots marker tracking algorithm. In *IEEE ISMAR 2012 Workshop on Tracking Methods and Applications (TMA)*, 2012. [x](#), [34](#), [35](#)
- [OLK⁺04] Thierry Oggier, Michael Lehmann, Rolf Kaufmann, Matthias Schweizer, Michael Richter, Peter Metzler, Graham Lang, Felix Lustenberger, and Nicolas Blanc. An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution (swissranger). In *Optical Systems Design*, pages 534–545. International Society for Optics and Photonics, 2004. [11](#)
- [Par94] Robert Ladislav Parker. *Geophysical inverse theory*. Princeton university press, 1994. [89](#)
- [PCZL09] Keju Peng, Xin Chen, Dongxiang Zhou, and Yunhui Liu. 3d reconstruction based on sift and harris feature points. In *Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on*, pages 960–964. IEEE, 2009. [61](#)
- [PdlC03] Gonzalo Pajares and Jesús M. de la Cruz. Stereovision matching through support vector machines. *Pattern Recogn. Lett.*, 24(15) :2575–2583, November 2003. [61](#)
- [PDM02] Euripides G. M. Petrakis, Aristeidis Diplaros, and Evangelos Milios. Matching and retrieval of distorted and occluded shapes using dynamic programming. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(11) :1501–1516, 2002. [61](#)
- [PHH10] Brian F Platt, Stephen T Hasiotis, and Daniel R Hirmas. Use of low-cost multistripe laser triangulation (mlt) scanning technology for three-dimensional, quantitative paleoichnological and neoichnological studies. *Journal of Sedimentary Research*, 80(7) :590–610, 2010. [14](#)
- [PKVG99] Marc Pollefeys, Reinhard Koch, and Luc Van Gool. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, 32(1) :7–25, 1999. [35](#)
- [PNF⁺08] Marc Pollefeys, David Nistér, J-M Frahm, Amir Akbarzadeh, Philippos Mordohai, Brian Clipp, Chris Engels, David Gallup, S-J Kim, Paul Merrell, et al. Detailed real-time urban 3d reconstruction from video. *International Journal of Computer Vision*, 78(2-3) :143–167, 2008. [61](#)
- [Pra78] W. K. Pratt. *Digital image processing*, 1978. [66](#)
- [Pra07] W.K. Pratt. *Digital Image Processing : PIKS Scientific Inside*. John Wiley & Sons, 4th edition, 2007. [98](#)

- [PVG99] Marc Pollefeys and Luc Van Gool. Stratified self-calibration with the modulus constraint. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(8) :707–724, 1999. [118](#)
- [RAMT01] Mohammed RZIZA, Driss ABOUTAJDINE, Luce MORIN, and Ahmed TAMTAOUI. Schéma multirésolution d’estimation d’un champ de disparités dense sous contrainte épipolaire pour les images bruitées. In *18th Colloque sur le traitement du signal et des images, FRA, 2001*. GRETSI, Groupe d’Etudes du Traitement du Signal et des Images, 2001. [65](#)
- [Rao73] C.Radhakrishna Rao. Representations of best linear unbiased estimators in the gauss-markoff model with a singular dispersion matrix. *Journal of Multivariate Analysis*, 3(3) :276 – 292, 1973. [89](#)
- [RAR⁺11] Caroline Rougier, Edouard Auvinet, Jacqueline Rousseau, Max Mignotte, and Jean Meunier. Fall detection from depth map video sequences. In *Toward Useful Services for Elderly and People with Disabilities*, pages 121–128. Springer, 2011. [47](#)
- [RHHL02] Szymon Rusinkiewicz, Olaf Hall-Holt, and Marc Levoy. Real-time 3d model acquisition. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 438–446. ACM, 2002. [71](#)
- [RWC⁺98] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future : A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM, 1998. [71](#)
- [SB97] Stephen M Smith and J Michael Brady. Susan : a new approach to low level image processing. *International journal of computer vision*, 23(1) :45–78, 1997. [61](#)
- [SC92] Jun Shen and Serge Castan. An optimal linear operator for step edge detection. *CVGIP : Graphical Models and Image Processing*, 54(2) :112–133, 1992. [61](#)
- [SCGMB13] Rémi Slys, Rémi Cогranne, Laurence Giraud-Moreau, and Houman Borouchaki. Une nouvelle méthode de détection de contours basée sur une régression locale de surface. Actes du Colloque GRETSI 2013, 2013. [104](#)
- [Sei89] Peter Seitz. Using local orientational information as image primitive for robust object recognition. In *1989 Advances in Intelligent Robotics Systems Conference*, pages 1630–1639. International Society for Optics and Photonics, 1989. [66](#)
- [SF68] Irwin Sobel and Gary Feldman. A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, pages 271–272, 1968. [66](#)

Bibliographie

- [SGMB13] R Slys, L Giraud-Moreau, and H Borouchaki. On uniqueness in triangulation based pattern for structured light reconstruction. In *Proc. of IEEE International Conference on 3D and Stereo (IC3D)*, 2013. 84
- [SM⁺95] Cordelia Schmid, Roger Mohr, et al. Matching by local invariants, 1995. Rapport de recherche RR-2644, Institut National de Recherche en Informatique et en Automatique, INRIA. 62
- [SM99] Peter F Sturm and Stephen J Maybank. On plane-based camera calibration : A general algorithm, singularities, applications. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1. IEEE, 1999. 34
- [SOY⁺09] Ryusuke Sagawa, Yuichi Ota, Yasushi Yagi, Ryo Furukawa, Naoki Asada, and Hiroshi Kawasaki. Dense 3d reconstruction method using a single pattern for fast moving object. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1779–1786. IEEE, 2009. 71, 72, 75
- [SP03] Christoph Stock and Axel Pinz. Similarity measure for corner redetection. In *Image Analysis*, pages 133–139. Springer, 2003. 62
- [SPB04] Joaquim Salvi, Jordi Pages, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4) :827–849, 2004. 71
- [Spe93] SA Spectec. Systeme de mesures sans contact. *Catalogue de présentation*, 1993. 11
- [SR09] N Senthilkumaran and R Rajesh. Edge detection techniques for image segmentation—a survey of soft computing approaches. *International Journal of Recent Trends in Engineering*, 1(2) :250–254, 2009. 85
- [SS02] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3) :7–42, 2002. 48
- [SS03] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–195. IEEE, 2003. 71
- [SSNK05] Henrik Stewénus, Frederik Schaffalitzky, David Nistér, and Fredrik Kahl. A minimal solution for relative pose with unknown focal length. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Volume 2*, volume 2, pages 789–794. IEEE Computer Society, 2005. 36

-
- [ST94a] Jianbo Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 593–600, jun 1994. [95](#)
- [ST94b] Jianbo Shi and Carlo Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994. [61](#)
- [STH⁺80] Chester C Slama, Charles Theurer, Soren W Henriksen, et al. *Manual of photogrammetry*. Number Ed. 4. American Society of photogrammetry, 1980. [33](#)
- [Stu02] Peter Sturm. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. *Image and Vision Computing*, 20(5) :415–426, 2002. [36](#)
- [Sze93] Richard Szeliski. Rapid octree construction from image sequences. *CVGIP : Image understanding*, 58(1) :23–32, 1993. [11](#)
- [TI04] M Trujillo and Ebroul Izquierdo. A robust correlation measure for correspondence estimation. In *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*, pages 155–162. IEEE, 2004. [65](#)
- [TLF08] Engin Tola, Vincent Lepetit, and Pascal Fua. A fast local descriptor for dense matching. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. [61](#), [62](#)
- [Tsa86] Roger Y Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition 1986*, 1986. [34](#)
- [UK04] Farhan Ullah and Shun'ichi Kaneko. Using orientation codes for rotation-invariant template matching. *Pattern recognition*, 37(2) :201–209, 2004. [66](#)
- [Ves79] Charles M Vest. Holographic interferometry. *New York, John Wiley and Sons, Inc., 1979. 476 p.*, 1, 1979. [9](#)
- [WCRZ11] Lijun Wei, Cindy Cappelle, Yassine Ruichek, and Frédérick Zann. Gps and stereovision-based visual odometry : Application to urban scene mapping and intelligent vehicle localization. *International Journal of Vehicular Technology*, 2011, 2011. [47](#)
- [XS03] Jiangjian Xiao and Mubarak Shah. Two-frame wide baseline matching. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 603–609. IEEE, 2003. [61](#)

Bibliographie

- [XZ96] Gang Xu and Zhengyou Zhang. *Epipolar geometry in stereo, motion and object recognition : a unified approach*, volume 6. Springer, 1996. [46](#), [61](#)
- [YZX⁺04] Yizhou Yu, Kun Zhou, Dong Xu, Xiaohan Shi, Hujun Bao, Baining Guo, and Heung-Yeung Shum. Mesh editing with poisson-based gradient field manipulation. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 644–651. ACM, 2004. [125](#)
- [ZCS02] Li Zhang, Brian Curless, and Steven M Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*, pages 24–36. IEEE, 2002. [71](#)
- [ZCS03] Li Zhang, Brian Curless, and Steven M Seitz. Spacetime stereo : Shape recovery for dynamic scenes. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–367. IEEE, 2003. [71](#)
- [ZH04] Song Zhang and Peisen Huang. High-resolution, real-time 3d shape acquisition. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pages 28–28. IEEE, 2004. [14](#), [71](#)
- [Zha98] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty : A review. *International Journal of Computer Vision*, 27(2) :161–195, 1998. [54](#)
- [Zha00] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11) :1330–1334, 2000. [34](#)
- [Zha04] Zhengyou Zhang. Camera calibration with one-dimensional objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(7) :892–899, 2004. [34](#)

Cette thèse s'inscrit dans le cadre du projet CPER Bramms, dont un des objectifs était de développer une méthode d'acquisition de la surface du buste féminin. Les travaux menés ont donc eu pour but la conception, le développement et la réalisation d'une machine de mesure tridimensionnelle adaptée aux objets vivants.

Parmi le nombre important de méthodes de mesures tridimensionnelles existantes, l'attention a été portée sur la mise en correspondance par stéréovision ainsi que sur l'utilisation de lumière structurée. La mise en correspondance par stéréovision consiste à retrouver les pixels homologues dans deux images d'une même scène, prise de deux points de vue différents. Une des manières de réaliser la mise en correspondance est de faire appel à des mesures de corrélation. Les algorithmes utilisés se heurtent alors à certaines difficultés : les changements de luminosité, les bruits, les déformations, les occultations, les zones peu texturées et les larges zones homogènes. L'utilisation de lumière structurée a permis essentiellement d'ajouter de l'information dans les zones homogènes lors des travaux menés. En développant cette approche, une méthode de reconstruction originale basée sur l'exploitation d'un motif particulier projeté sur la surface a ainsi été conçue. Un appariement basé sur la comparaison de signatures de points particuliers du motif a été mis en place. Ce procédé permet une reconstruction éparsée en une unique acquisition et simplifie l'étape de gestion du nuage de points pour en faire un maillage surfacique.

Mots clés : Vision par ordinateur ; Photographie stéréoscopique ; Appariement (statistique) ; Reconnaissance optique des formes (informatique)

This thesis is part of the CPER BRAMSS project, one of its objectives was to develop an surface's retrieval method applied to the female bust. Therefore the work has aimed at the design, development and implementation of a three-dimensional measuring machine adapted to living objects.

Among the large number of existing methods of three-dimensional measurements, attention was paid to the stereo matching as well as the use of structured light. Matching in stereovision is to find homologous pixels in two images of the same scene, taken from two different points of view. One way to achieve the mapping is to use correlation measurements. The algorithms used come up against certain difficulties : the changing light, noises, distortions, occlusions, low textured areas and large homogeneous areas. The use of structured light allows essentially the adding of information in homogeneous areas in this work. Developing this approach, an original method of reconstruction based on the exploitation of a particular pattern projected on the surface has been designed. A matching based on a comparison of the signatures of specific points in the pattern was implemented. This method allows a single sparse reconstruction acquisition step and simplifies the handling of the point cloud when transforming it in a surface mesh.

Keywords : Computer vision ; Stereo Photography ; Matching (statistics) ; Pattern Recognition (computing)

Rémi SLYSZ

Doctorat : Systèmes Mécaniques et Matériaux

Année 2014

Reconstruction de surface 3D d'objets vivants

Cette thèse s'inscrit dans le cadre du projet CPER Bramms, dont un des objectifs était de développer une méthode d'acquisition de la surface du buste féminin. Les travaux menés ont donc eu pour but la conception, le développement et la réalisation d'une machine de mesure tridimensionnelle adaptée aux objets vivants.

Parmi le nombre important de méthodes de mesures tridimensionnelles existantes, l'attention a été portée sur la mise en correspondance par stéréovision ainsi que sur l'utilisation de lumière structurée. La mise en correspondance par stéréovision consiste à retrouver les pixels homologues dans deux images d'une même scène, prise de deux points de vue différents. Une des manières de réaliser la mise en correspondance est de faire appel à des mesures de corrélation. Les algorithmes utilisés se heurtent alors à certaines difficultés : les changements de luminosité, les bruits, les déformations, les occultations, les zones peu texturées et les larges zones homogènes. L'utilisation de lumière structurée a permis essentiellement d'ajouter de l'information dans les zones homogènes lors des travaux menés. En développant cette approche, une méthode de reconstruction originale basée sur l'exploitation d'un motif particulier projeté sur la surface a ainsi été conçue. Un appariement basé sur la comparaison de signatures de points particuliers du motif a été mis en place. Ce procédé permet une reconstruction éparse en une unique acquisition et simplifie l'étape de gestion du nuage de points pour en faire un maillage surfacique.

Mots clés : vision par ordinateur - photographie stéréoscopique - appariement (statistique) - reconnaissance optique des formes (informatique).

3D Surface Reconstruction of Living Object

This thesis is part of the CPER BRAMSS project, one of its objectives was to develop an surface's retrieval method applied to the female bust. Therefore the work has aimed at the design, development and implementation of a three-dimensional measuring machine adapted to living objects.

Among the large number of existing methods of three-dimensional measurements, attention was paid to the stereo matching as well as the use of structured light. Matching in stereovision is to find homologous pixels in two images of the same scene, taken from two different points of view. One way to achieve the mapping is to use correlation measurements. The algorithms used come up against certain difficulties: the changing light, noises, distortions, occlusions, low textured areas and large homogeneous areas. The use of structured light allow essentially the adding of information in homogeneous areas in this work. Developing this approach, an original method of reconstruction based on the exploitation of a particular pattern projected on the surface has been designed. A matching based on a comparison of the signatures of specific points in the pattern was implemented. This method allows a single sparse reconstruction acquisition step and simplifies the handling of the point cloud when transforming it in a surface mesh.

Keywords: computer vision - stereo photography - matching (statistics) - pattern recognition (computing).

Thèse réalisée en partenariat entre :

