THESE

pour obtenir le grade de
DOCTEUR DE L'ECOLE CENTRALE DE LYON
Spécialité : Informatique

présentée et soutenue publiquement par

Karima OUJI

le 28 JUIN 2012

Numérisation 3D de visages par une approche de super-résolution spatio-temporelle non-rigide

---

JURY

Pr. Nikos PARAGIOS	Ecole Centrale de Paris	Président
Pr. Fabrice MERIAUDEAU	Institut Universitaire de Technologie Le Creusot	Rapporteur
Pr. Chokri BEN AMAR	Ecole Nationale d'Ingénieurs de Sfax	Rapporteur
Pr. Liming CHEN	Ecole Centrale de Lyon	Directeur de thèse
Pr. Faouzi GHORBEL	Ecole Nationale des Sciences de l'Informatique de Tunis	Directeur de thèse
Dr. Mohsen ARDABILIAN	Ecole Centrale de Lyon	Co-Directeur de thèse

Dédicace

Je ne remercierai sans doute jamais assez mes parents, mon frère et mes deux sœurs à qui je dois tout, je leur dédie mon travail.

Remerciements

Je souhaiterais remercier tout d'abord les membres du jury de cette thèse :

- **Pr. Nikos Paragios** pour avoir accepté la présidence du jury et d'avoir également été un examinateur de ce manuscrit ;
- **Pr. Fabrice Meriaudeau** et **Pr. Chokri BEN AMAR** pour le temps passé à l'analyse et à la critique de ce manuscrit en tant que rapporteurs ;
- **Pr. Liming Chen** et **Pr.Faouzi Ghorbel**, directeurs de cette thèse, dont l'expérience et les précieux conseils m'ont été très utiles, et qui ont su me motiver dans les moments d'encouragement avec une grande patience ;
- **Dr. Mohsen Ardabilian**, qui a codirigé cette thèse, et qui a grandement aidé à la réalisation des expérimentations et au bon déroulement de cette thèse avec un suivi régulier de l'état d'avancement de mes travaux de recherche, merci de m'avoir fait confiance.

Un grand merci également à toutes les personnes faisant partie du laboratoire LIRIS à l'Ecole Centrale de Lyon particulièrement **Mme Isabelle SanJosé**, **Mme Colette Vial** pour leur disponibilité et leur gentillesse. Je remercie également **Mr Emmanuel Dellandrea**, **Mme Elisabeth Mironescu**, **Mr Christian Vial** et **Mr Abdel-Malek Zine** pour leur amabilité.

Je remercie également tous les membres du laboratoire CRISTAL-GRIFT particulièrement **Mlle Randa Mansour** pour sa disponibilité et son amabilité.

Je tiens à remercier mes amis et collègues du laboratoire LIRIS particulièrement **Wael BenSoltana**, **Aliaksandr Paradzinets**, **Przemyslaw Szeptycki**, **Chu Duc Nguyen**, **Pierre Lemaire**, **Huang Di** et **Boyang Gao** ainsi que mes amis **Hajer Chamekh**, **Fatiha Benali**, **Aida Hadj Kassem**, **Manel Migaou**, **Charlotte Carré**, **Emilie Ferré**, **Mohamed Amine Ben Souf**, **R.carlos Mbou**, **Mohamed Abouobayd** et tous mes chers amis du **V2** qui ont partagé avec moi les moments de difficultés et de stress et m'ont soutenu tout au long de cette thèse.

Résumé

La mesure de la forme 3D du visage est une problématique qui attire de plus en plus de chercheurs et qui trouve son application dans des domaines divers tels que la biométrie, l'animation et la chirurgie faciale. Les solutions actuelles sont souvent basées sur des systèmes projecteur/caméra et utilisent de la lumière structurée pour compenser l'insuffisance de la texture faciale. L'information 3D est ensuite calculée en décodant la distorsion des patrons projetés sur le visage. Une des techniques les plus utilisées de la lumière structurée est la codification sinusoïdale par décalage de phase qui permet une numérisation 3D de résolution pixélique. Cette technique exige une étape de déroulement de phase, sensible à l'éclairage ambiant surtout quand le nombre de patrons projetés est limité. En plus, la projection de plusieurs patrons impacte le délai de numérisation et peut générer des artéfacts surtout pour la capture d'un visage en mouvement. Une alternative aux approches projecteur-caméra consiste à estimer l'information 3D par appariement stéréo suivi par une triangulation optique. Cependant, le modèle calculé par cette technique est généralement non-dense et manque de précision. Des travaux récents proposent la super-résolution pour densifier et débruiter les images de profondeur. La super-résolution a été particulièrement proposée pour les caméras 3D TOF (Time-Of-Flight) qui fournissent des scans 3D très bruités. Ce travail de thèse propose une solution de numérisation 3D à faible coût avec un schéma de super-résolution spatio-temporelle. Elle utilise un système multi-caméra étalonné assisté par une source de projection non-étalonnée. Elle est particulièrement adaptée à la reconstruction 3D de visages, i.e. rapide et mobile. La solution proposée est une approche hybride qui associe la stéréovision et la codification sinusoïdale par décalage de phase, et qui non seulement profite de leurs avantages mais qui surmonte leurs faiblesses. Le schéma de la super-résolution proposé permet de corriger l'information 3D, de compléter la vue scannée du visage en traitant son aspect déformable.

Mots clés

Numérisation 3D, Stéréovision active, codification sinusoïdale, décalage de phase, multi-caméras, Appariement 3D non-rigide, Super-résolution, Spatio-temporel.

Abstract

3D face measurement is increasingly demanded for many applications such as biometrics, animation and facial surgery. Current solutions often employ a structured light camera/projector device to overcome the relatively uniform appearance of skin. Depth information is recovered by decoding patterns of the projected structured light. One of the most widely used structured-light coding is sinusoidal phase shifting which allows a 3D dense resolution. Current solutions mostly utilize more than three phase-shifted sinusoidal patterns to recover the depth information, thus impacting the acquisition delay. They further require projector-camera calibration whose accuracy is crucial for phase to depth estimation step. Also, they need an unwrapping stage which is sensitive to ambient light, especially when the number of patterns decreases. An alternative to projector-camera systems consists of recovering depth information by stereovision using a multi-camera system. A stereo matching step finds correspondence between stereo images and the 3D information is obtained by optical triangulation. However, the model computed in this way generally is quite sparse. To upsample and denoise depth images, researchers looked into super-resolution techniques. Super-resolution was especially proposed for time-of-flight cameras which have very low data quality and a very high random noise. This thesis proposes a 3D acquisition solution with a 3D space-time non-rigid super-resolution capability, using a calibrated multi-camera system coupled with a non calibrated projector device, which is particularly suited to 3D face scanning, i.e. rapid and easily movable. The proposed solution is a hybrid stereovision and phase-shifting approach, using two shifted patterns and a texture image, which not only takes advantage of the assets of stereovision and structured light but also overcomes their weaknesses. The super-resolution scheme involves a 3D non-rigid registration for 3D artifacts correction in the presence of small non-rigid deformations as facial expressions.

Keywords

3D scanning, Active stereovision, Sinusoidal coding, Phase-shifting, Multi-camera, Non-rigid matching, Super-resolution, Spacetime.

Table des matières

CHAPITRE 1

Introduction

1.1 Contexte

La numérisation 3D de visages est une problématique qui attire de plus en plus de chercheurs et qui joue un rôle clé dans des domaines divers tels que la biométrie, le cinéma, le jeu vidéo, la santé et l'art.

Concernant la biométrie faciale d'abord, les contextes international et national placent actuellement le problème de la sécurité comme une priorité d'institutions, d'entreprises et du gouvernement. La vidéosurveillance se multiplie, la biométrie faciale s'impose comme une solution efficace et non-intrusive pour lutter contre les menaces terroristes et retrouver les auteurs de faits suspects ou délictueux. Avec l'essor des techniques de numérisation 3D, une nouvelle gamme d'approches de reconnaissance basées sur l'information de profondeur a vu le jour montrant l'apport significatif de l'information 3D pour pallier les problèmes liés au changement de pose et aux conditions d'éclairage rencontrés en reconnaissance faciale 2D. Dans un scénario d'identification de visage par exemple, la numérisation faciale 3D permet la construction hors-ligne d'une galerie de visages 3D et la capture enligne d'un visage 3D pour l'identifier.

Ensuite l'industrie du cinéma et de l'animation 3D qui constitue un des plus grands consommateurs de cette technologie et un des principaux moteurs économiques pour son émergence. Cette obsession technologique pour la restitution du relief a donné naissance à une nouvelle ère de cinéma 3D. Les visages et les corps des acteurs sont numérisés et remplacés dans le film par leurs personnages virtuels appelés avatars. Une des premières utilisations de la technologie de numérisation 3D dans le cinéma était dans le film Teminator 2, tourné en 1991. Pour la première fois, un personnage de synthèse à été animé par des mouvements humains réels. Aussi vers la fin 2009, le succès planétaire du film Avatar de

James Cameron, plébiscité dans sa version en relief 3D, a vaincu les dernières réticences et provoqué une forte accélération des investissements des cinémas pour s'équiper de la technologie 3D. Avatar est le premier film qui évolue entièrement dans un monde 3D virtuel avec des scènes synthétisées quasi-photographiques et des personnages 3D photoréalistes comme le montre la figure 1.1.

FIGURE 1.1 – Le film Avatar.

Récemment la 3D a attiré davantage l'industrie du jeu vidéo ; les concepteurs de jeu cherchent à offrir le maximum d'interaction physique entre le joueur et la scène. En août 2009, le fabriquant de jeu japonais Nintindo propose sa première console de jeu Wii qui fournit une interaction physique en utilisant une manette de jeu sans fil, une raquette de tennis ou un gant de boxe interactifs. En octobre 2010, Microsoft dévoile sa nouvelle console de jeu Xbox 360 qui permet cette fois de contrôler des jeux vidéo sans utiliser de manette. Elle dispose de Kinect, son système de capture 3D de la forme et du mouvement. Avec 10 millions d'unités vendus, Kinect est entré le 11 mars 2011 dans le livre Guinness des records comme étant l'accessoire high-tech le plus vendu dans un court laps de

temps. Le système Kinect a donné naissance à une nouvelle ère de jeu vidéo qui fournit une interaction directe et temps-réel entre le joueur et la scène de jeu 3D. Plusieurs jeux ont accompagné le lancement du Kinect tels que Kinect Adventures, Kinectimals, Fighters Uncaged, etc.

Dans le domaine de la santé, avec la possibilité d'une réduction de coût massive favorisée par l'augmentation de la puissance de calcul, des méthodes de numérisation 3D sont utilisées dans des applications exigeant cette fois la haute précision de numérisation. En chirurgie plastique par exemple, les résultats postopératoires peuvent être simulés à partir des scans 3D. L'intervention peut être planifiée à l'aide, en partie, de la numérisation 3D. La 3D est aussi utile en orthopédie pour la mesure des membres et la fabrication sur mesure des prothèses et des semelles orthopédiques. Cette visualisation permet de prévenir et de corriger la position défectueuse des dents et la mesure 3D de la moule dentaire imprimée. En dermatologie, la numérisation 3D de la surface du corps humain permet de capturer la topologie 3D de la surface de peau et de caractériser la cellulite à l'exemple de la technologie proposée par Cyberware.

Numériser un visage 3D offre aussi la possibilité d'un essayage virtuel d'accessoires, comme les lunettes, et la possibilité de choisir la coupe de cheveux adéquate comme le propose la compagnie américaine Stellure. Une interface interactive permet au client de choisir une coupe de cheveux et de l'ajouter sur son visage numérisé pour une visualisation préalable comme l'illustre la figure 1.2.

FIGURE 1.2 – Ajout virtuel de cheveux pour le choix d'une coupe adéquate proposée par la compagnie Stellure.

Finalement, le visage numérique 3D a inspiré des artistes. Il a ouvert un nouveau champ de création artistique pour la gravure sur cristal de portraits ou l'innovation d'œuvres d'art modernes. Une installation artistique appelée Jurisprudents réalisée par Helmick et Schech-

ter utilise la technologie de numérisation 3D. Elle est exposée au Melvin Prince Federal Courthouse. En utilisant des méthodes traditionnelles, les artistes ont sculpté douze grands portraits de citoyens américains ordinaires, représentant douze membres d'un jury. Les têtes ont ensuite été numérisées par un scanner laser permettant de générer environ 3000 petites sculptures suspendues pour former deux têtes monumentales en face à face.

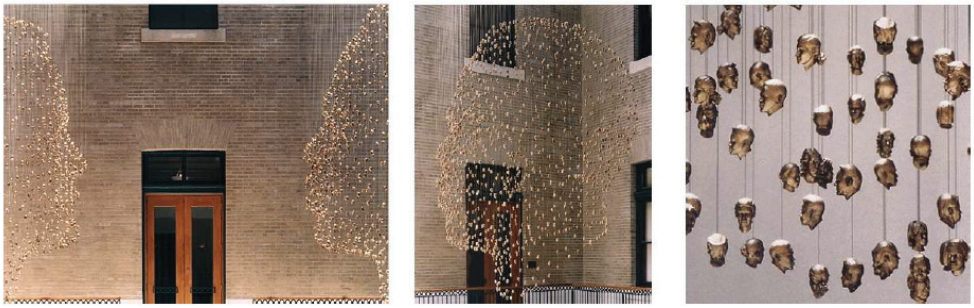


FIGURE 1.3 – L'installation artistique Jurisprudents réalisée par Helmick et Schechter, au Melvin Prince Federal Courthouse au Etats Unis.

Les travaux de recherche réalisés dans cette thèse s'inscrivent dans le cadre du projet FAR3D ANR-07-SESU-004 (Face Analysis and Recognition using 3D) qui a regroupé TELECOM Lille1/LIFL, Ecole Centrale de Lyon/LIRIS, Eurecom et Thales.

1.2 Problématique

Les techniques de numérisation 3D disponibles sur le marché permettent de reconstruire efficacement un modèle 3D d'un visage, mais le plus souvent, elles requièrent des équipements complexes et coûteux. Malgré une augmentation très rapide de l'offre de solutions performantes, il existe une grande demande pour des alternatives à faible coût. De plus, la numérisation 3D d'un visage animé constitue encore un défi surtout en présence de déformations non-rigides difficiles à modéliser.

Deux problématiques sont considérées dans ce travail de thèse. La première consiste à réaliser un système de numérisation 3D de formes et particulièrement de visages à faible coût, facile à mettre en place, rapide et capable de récupérer le maximum de détails sur le visage en palliant la platitude de la texture faciale. La deuxième problématique est l'aspect dynamique déformable de la forme faciale qui peut affecter la qualité de la forme 3D et

engendrer des erreurs de numérisation. Il s'agit de concevoir un schéma de super-résolution spatio-temporel apte à renforcer la qualité de la reconstruction 3D même en présence de variations non-rigide entre les trames 3D successives.

1.3 Objectifs et contributions

Dans ce manuscrit, nous proposons une solution de numérisation de séquences 3D texturées à faible coût et particulièrement adaptée à la numérisation 3D de visages, c.-à-d. rapide et mobile. Elle utilise deux ou plusieurs caméras étalonnées et un vidéoprojecteur non-étalonné et profite des atouts de la stéréovision et de la lumière structurée par décalage de phase pour une reconstruction 3D plus fidèle à la forme réelle du visage. La solution proposée intègre un schéma de super-résolution spatio-temporelle, avec un recalage non-rigide 3D, pour corriger l'information 3D et compléter la vue scannée du visage tout en considérant son aspect déformable. Une vue globale de notre système de numérisation est illustrée par la figure 1.4.

Nous utilisons une lumière structurée codée par un multiplexage temporel. Deux patrons sinusoïdaux en opposition de phase et un troisième patron blanc sont successivement projetés sur le visage. Selon notre approche, un échantillonnage 2D est d'abord appliqué pour chaque caméra séparément pour localiser les points d'intersection de franges. Un modèle 3D non-dense du visage est ainsi estimé pour chaque couple de caméras par un appariement stéréo entre les primitives obtenues et une triangulation optique. La résolution spatiale de ce modèle dépend du nombre de franges formant le patron utilisé. Un modèle 3D dense est ensuite obtenu par une estimation de l'information de phase des points situés à l'intérieur des franges, séparément pour chaque couple caméra-projecteur utilisées. Les approches classiques à base de lumière structurée par décalage de phase nécessitent un étalonnage hors-ligne du vidéoprojecteur avec les caméras et une étape de déroulement de phase. Contrairement à ces approches et pour apporter plus de flexibilité à notre système de numérisation, nous suggérons une estimation enligne des paramètres du vidéoprojecteur. En outre, l'étape de déroulement de phase n'est plus nécessaire grâce à l'étape d'appariement stéréo. Finalement, pour réduire la complexité temporelle de la numérisation, nous proposons une nouvelle approche de segmentation de la région faciale par une analyse de

l'amplitude du signal distordu sur le visage. Les nuages de points 3D obtenus pour chaque vue stéréo sont ensuite fusionnés pour densifier le modèle 3D et renforcer sa précision.

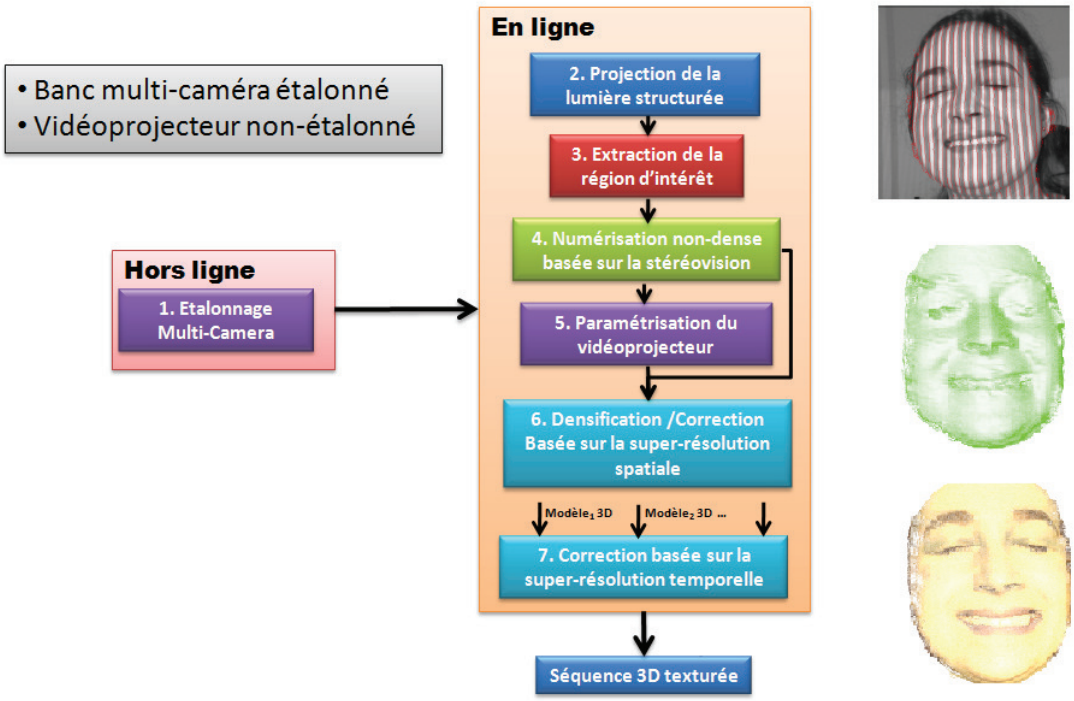


FIGURE 1.4 – Architecture du système.

La super-résolution spatio-temporelle permet de compléter et corriger les modèles 3D de visages puisque la numérisation 3D peut engendrer des distorsions et des artéfacts causés essentiellement par des occultations, par une variation de la pose ou de l'expression faciale, ou même par une réflexion de la lumière sur la surface faciale. Ainsi, en utilisant d'une part les différents modèles 3D fournis par tous les couples de caméras et d'autre part la trame 3D calculée à l'instant $t-1$, nous obtenons un modèle 3D de haute résolution corrigé à l'instant t. La super-résolution spatio-temporelle est assurée par une première étape d'appariement 3D suivie par une étape de fusion et de débruitage. Nous proposons une approche non-rigide d'appariement 3D pour traiter une éventuelle distorsion ou déformation faciale non-rigide. Un maillage du nuage 3D obtenu et un plaquage de la texture permettent de finaliser la trame 3D texturée du visage de l'instant t.

1.4 Organisation de la thèse

Le manuscrit de thèse s'articule autour de cinq parties essentielles. Le chapitre 2 présente un état de l'art sur la numérisation 3D. D'abord, nous introduisons les approches passives qui récupèrent l'information 3D sans assistance par une source lumineuse. Ensuite, nous étudions les approches actives qui estiment l'information 3D en analysant le profil d'une lumière projetée sur la surface au moment de la numérisation. Dans ce chapitre, nous discutons aussi les techniques hybrides qui remédient aux défaillances des approches de numérisation existantes par un modèle générique, une fusion d'approches, une analyse spatio-temporelle ou par super-résolution. Nous suggérons également d'étudier le problème de la modélisation 3D de la déformation non-rigide nécessaire pour renforcer la numérisation 3D d'un visage animé. La dernière partie de ce chapitre est consacrée à une analyse des systèmes de numérisation 3D commercialisés et à déceler leur points forts et leurs limites.

Dans le chapitre 3, nous introduisons une approche de numérisation 3D basée sur la stéréovision active. Une projection successive d'un patron binaire de franges noires et blanches alternées et de son patron complémentaire permet de pallier la platitude de la texture du visage. Le système de numérisation emploie deux caméras étalonnées et un vidéoprojecteur non-étalonné. Ainsi, une présentation du modèle sténopé de la caméra, de l'étalonnage stéréo et de la géométrie épipolaire s'avère nécessaire. Ce chapitre décrit ensuite le principe de l'échantillonnage 2D appliqué sur les deux vues stéréo gauche et droite séparément pour localiser les points d'intersection de franges sur les deux vues gauche et droite du visage. Les coordonnées 3D sont obtenues par un appariement stéréo suivi d'une triangulation optique. Une interpolation par les splines cubiques permet de densifier le nuage de points 3D obtenu. Finalement, un maillage suivi d'un placage de texture fournit un modèle 3D texturé du visage. Ce chapitre suggère une validation expérimentale qualitative et quantitative de l'approche proposée.

Nous proposons dans le chapitre 4, de projeter une lumière structurée sinusoïdale et de profiter des atouts de la stéréovision et de la lumière structurée par décalage de phase pour une reconstruction 3D précise et dense. D'abord, nous introduisons notre approche de segmentation de la région faciale. Ensuite, il s'agit de calculer un premier modèle non-

dense par stéréovision, formé par les points d'intersection de franges. Nous suggérons une approche d'estimation en ligne des paramètres du vidéoprojecteur basée sur le modèle non-dense obtenu. La lumière structurée sinusoïdale permet par la suite de calculer la phase locale pour chaque pixel à l'intérieur des franges. Nous proposons aussi une nouvelle technique de conversion directe de la phase locale en profondeur, remplaçant l'étape de déroulement de phase. Une étude expérimentale est enfin développée pour valider les étapes de la segmentation 2D, de la localisation 3D du vidéoprojecteur, et pour évaluer la qualité de la reconstruction 3D obtenue.

Dans le chapitre 5, l'objectif est de renforcer la qualité de la numérisation 3D d'un visage animé par une fusion spatio-temporelle des différents modèles 3D obtenus par chaque couple de caméras aux instants t et $t-1$. Nous proposons un schéma de super-résolution/correction spatio-temporelle qui traite les déformations non-rigides affectant le visage animé à travers le temps. D'abord, nous décrivons le principe de notre technique qui emploie deux ou plusieurs caméras préalablement étalonnées et un vidéoprojecteur non-étalonné. Ensuite, nous introduisons notre approche de super-résolution spatio-temporelle qui renforce la qualité du rendu facial numérisé par une correction des aberrations/artéfacts crées au moment de la numérisation. Les étapes d'appariement 3D non-rigide, de fusion et de débruitage nécessaires pour la reconstruction du modèle 3D de haute résolution final sont développées. Enfin, nous proposons d'évaluer les performances du système en utilisant en une première étape un système bi-caméra et en une deuxième étape un système tri-caméra. Nous suggérons une évaluation qualitative et quantitative de l'approche ainsi qu'une étude de la consistance de l'étape d'appariement non-rigide.

Le chapitre 6 est consacré à la conclusion et aux perspectives de ce travail. Dans la première section, nous décrivons nos contributions pour la numérisation 3D de visages fixes ou animés en valorisant les atouts de la technique proposée. Ensuite, nous introduisons les différentes perspectives envisageables pour améliorer la solution développée.

CHAPITRE 2

La numérisation optique 3D

2.1 Introduction

Ce chapitre est consacré à l'étude des techniques de numérisation sans contact optiques. Une taxonomie globale de la numérisation 3D est d'abord mise en œuvre. Ensuite, une étude de l'état de l'art de la numérisation 3D passive et active est réalisée. Nous discutons ensuite des techniques hybrides qui remédient aux défaillances des approches de numérisation existantes par une fusion d'approches, une analyse spatio-temporelle ou par super-résolution. Une attention particulière est accordée à l'influence de la déformation non-rigide des objets sur la qualité de la numérisation 3D. Diverses solutions ont été proposées pour surmonter le défi de la déformation non-rigide qui peut être une variation d'expression lors de la numérisation d'un visage. Finalement, nous présentons brièvement quelques systèmes de numérisation 3D disponibles sur le marché.

2.2 Taxonomie

La numérisation 3D distingue les techniques qui nécessitent un contact physique entre le système d'acquisition et l'objet à numériser des techniques capables de récupérer l'information 3D sans contact. La figure 2.1 présente une taxonomie des techniques de numérisation 3D avec et sans contact [Curless 2000]. Les techniques avec contact sondent le sujet grâce à un contact physique et fournissent une bonne précision. Elles se partagent entre des techniques qui détruisent l'objet pour le numériser en le découpant et des techniques non-destructives à l'exemple du bras articulé de palpage et de la MMT : Machine à Mesurer Tridimensionnelle. Le bras de palpage est un capteur positionné au bout d'un bras que l'on peut déplacer dans toutes les directions. On vient placer le capteur sur l'objet

à numériser pour prendre les coordonnes x, y, z dans l'espace de ce point. La reconstruction de l'objet s'obtient par une multiplication des prises de points. Cette technologie est destinée à reconstituer des objets techniques de géométrie simple. La numérisation 3D par palpage est lente puisque la mesure s'effectue point par point. La technologie MMT utilise une tête de mesure, une table sur laquelle la pièce à mesurer est immobilisée, trois liaisons glissières permettant de positionner la tête de mesure en tout point de l'espace et des règles graduées optiques ou électriques permettant de connaître la position de chacune des glissières. Les techniques de numérisation avec contact constituent des outils de métrologie utilisés essentiellement en industrie pour leur haute précision qui est de l'ordre du micron.

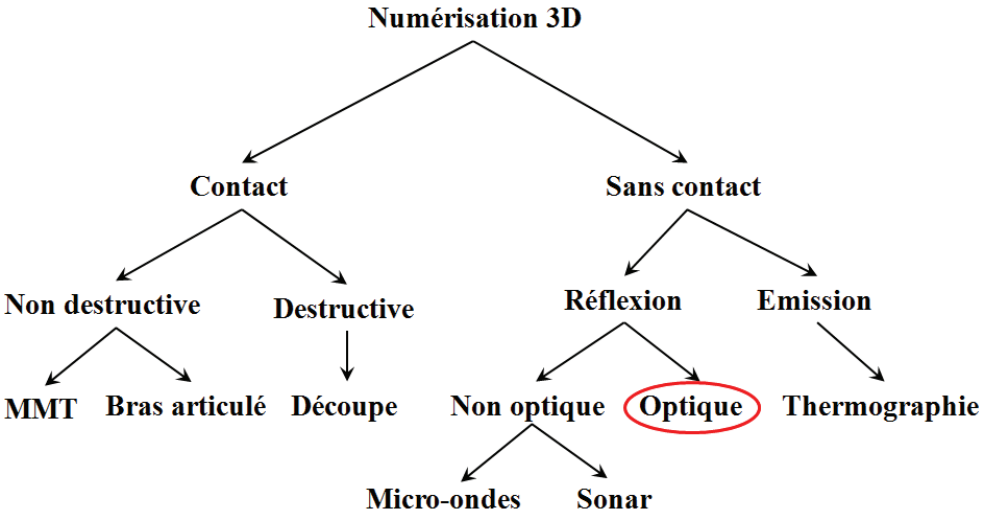


FIGURE 2.1 – Taxonomie de la numérisation 3D.

La famille de techniques de numérisation 3D sans contact se base sur le principe de l'émission ou de la réflexion d'ondes. La thermographie récupère l'information 3D en utilisant le principe de l'émission des ondes thermiques en évaluant les performances thermiques de l'objet. Elle est utilisée par exemple pour une localisation précise 3D des sources de déperdition de chaleur dans un bâtiment. Le principe de la réflexion permet aussi de récupérer l'information 3D en analysant le profil d'ondes optiques ou non-optiques réfléchies par la surface de l'objet. Les ondes non-optiques peuvent être des ondes sonar ou des micro-ondes.

La numérisation 3D optique peut être assurée par des méthodes passives ou actives. Les méthodes passives récupèrent l'information 3D en utilisant l'apparence observée de la

scène et des objets dans des images sans assistance par une source de lumière. Par contre, les méthodes actives nécessitent la projection d'une lumière sur la surface de l'objet pour le numériser. La figure 2.2 présente une classification non-exhaustive des techniques optiques. Dans les deux sections suivantes, nous décrivons les différentes approches passives et actives et nous discutons leurs atouts ainsi que leurs limitations.

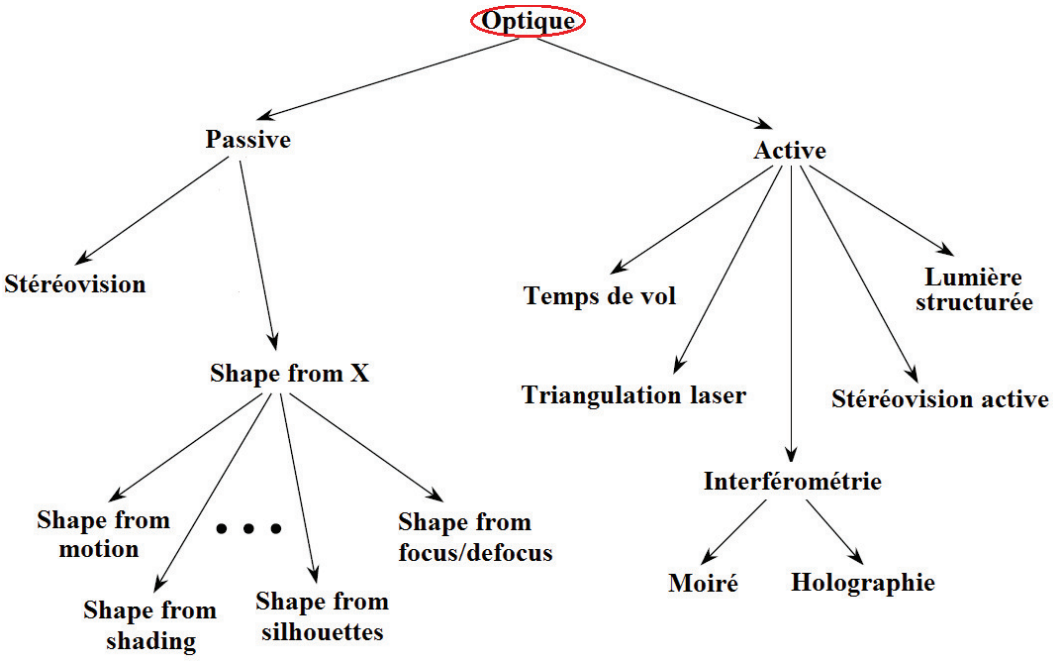


FIGURE 2.2 – Classification non-exhaustive des techniques optiques.

2.3 Numérisation 3D passive

La numérisation passive peut être mono-caméra ou multi-caméras, mono-image ou multi-images. Une seule caméra avec plusieurs images successives permet d'estimer la forme 3D d'un objet par une étude du mouvement de l'objet (shape from motion), par focalisation ou défocalisation (shape from focus/defocus). Les deux approches shape from motion et shape from focus/defocus sont ainsi mono-caméra et multi-images. La profondeur peut aussi être retrouvée par une analyse de l'ombrage (shape from shading) qui utilise une seule caméra et une seule image. Autrement, la numérisation passive peut employer deux ou plusieurs caméras pour retrouver la dimension perdue. Dans ce cas, la reconstruction

3D peut se faire par stéréovision, par photogrammétrie ou par une analyse des silhouettes visuelles de l'objet crées par plusieurs acquisitions 2D prises de différents points de vue (shape from silhouettes). Ces approches sont donc multi-caméras et mono-image.

2.3.1 Photogrammétrie

La photogrammétrie, appelée aussi stéréophotogrammétrie, est une technologie basée sur la géométrie projective. Elle permet la numérisation des objets de grandes tailles comme des bâtiments, des derricks et des entrepôts. Le principe de la photogrammétrie est de prendre des images multiples de l'objet et de localiser manuellement ou automatiquement des points communs à chaque photographie. Les points permettent une mesure 3D de l'objet par une triangulation optique. Les systèmes photogrammétriques sont capables de modéliser des environnements géométriques dans lesquels les objets étudiés comportent des primitives géométriques simples et elle est moins performante pour la numérisation 3D d'une surface courbe [Goulette 1999]. [Uffenkamp 1993] présente un état de l'art sur la photogrammétrie. Cette approche permet de retrouver avec précision la taille et la forme des objets géométriques mais elle reste une approche coûteuse puisque le système est multi-caméras.

2.3.2 Stéréovision

Cette technique consiste à estimer l'information 3D par appariement stéréo dans une configuration multi-caméra [Scharstein & Szeliski 2001, O.Faugeras 1993]. Dans ce cas, une étape d'appariement stéréo permet de calculer la disparité. L'information 3D est ensuite estimée par une triangulation optique. Le procédé d'appariement stéréo utilise l'information colorimétrique comme critère de similarité pour identifier les couples de pixels correspondants entre les vues stéréo. Lorsque l'objet à numériser est faiblement texturé, les algorithmes classiques d'appariement stéréo montrent leurs limites.

Essentiellement, deux classes d'approches sont proposées pour résoudre le problème de mise en correspondance stéréo : les approches locales et les approches globales. Les approches locales utilisent un large voisinage autour des pixels pour augmenter leur discrimination au moment de l'appariement [Hirschmüller & Scharstein 2009]. Les ap-

proches locales ont une difficulté à gérer le problème d'occultation qui constitue une propriété globale de la scène [Scharstein & Szeliski 2001]. Souvent, les approches locales emploient une mesure de similarité calculée à partir de l'information colorimétrique sensible à la variation de la réflectance. Cette catégorie d'approches peut ne pas réussir l'appariement lorsque la scène contient une grande variation de la disparité. Aussi, les approches locales ne sont pas adéquates aux régions faiblement texturées [D.N.Bhat & S.K.Nayar 1998, S.Birchfield & C.Tomasi 1999].

Les approches globales formulent explicitement la régularité de la surface et le processus de la mise en correspondance stéréo comme un problème d'optimisation. Plusieurs approches ont été proposées tels que le recuit simulé [Jones 1997], la programmation dynamique [Ohta & Kanade 1985, Sadeghi *et al.* 2008], l'analyse de la corrélation canonique [Borga & Knutsson 1998], la diffusion non-linéaire [D.Scharstein & R.Szeliski 1998], la propagation de croyances [C.Tomasi & R.Manduchi 1998], et la technique de coupure de graphes [Kolmogorov *et al.* 2003].

Les approches globales sont moins sensibles au problème d'ambiguïté de correspondance et aptes à détecter les régions occultées. Cependant, ces techniques sont plus coûteuses que les approches locales. Les approches globales cherchent une estimation de la disparité qui minimise une fonction de coût globale. La fonction de coût combine les données et les contraintes de régularité. Les fonctions de coût sont souvent basées sur l'information colorimétrique sensible à la variation de l'illumination et de la réflectance. Elles emploient des différences carrées d'intensité pixélique, des différences absolues d'intensité ou une corrélation normalisée [Hirschmüller & Scharstein 2009]. Une alternative consiste à utiliser une fonction de coût basée sur la disparité [Blake & Zisserman 1987, Marr & Poggio 1976]. La valeur de la disparité ou ses dérivées ont été utilisées comme une contrainte pour faire converger le processus d'optimisation [Pollard *et al.* 1985]. Dans [Sadeghi *et al.* 2008], les auteurs proposent de considérer une limite de disparité et une limite de la dérivée directionnelle de la disparité comme des contraintes de régularité pour réduire l'espace de la recherche. Les avantages de la résolution de la mise en correspondance stéréo en imposant une limite sur la magnitude probable du gradient de la disparité ont été examinés dans [Pollard *et al.* 1985]. Dans [Geiger *et al.* 1992], les auteurs proposent que la discontinuité de la disparité dans le système de coordonnées d'un œil correspond

à une occultation dans l'autre œil qui doit être formulé comme une contrainte d'occultation ou de monotonie. Les approches mentionnées utilisent une mesure de similarité basée sur l'intensité pixélique et introduisent l'information de la disparité comme une contrainte limitante.

2.3.3 Shape from X

Cette section décrit les différentes techniques passives de shape from silhouettes (la forme à partir des silhouettes de l'objet), shape from motion (la forme à partir du mouvement), shape from texture (la forme à partir de l'analyse de la texture), shape from focus/defocus (la forme à partir de la focalisation/défocalisation) et shape from shading (la forme à partir de l'ombrage).

2.3.3.1 Shape from silhouettes

La méthode shape from silhouettes (la forme à partir des silhouettes de l'objet) ou encore Visual Hulls exploite les différences entre les images acquises depuis différents points de vue pour procurer une estimation initiale du modèle 3D appelée enveloppe visuelle. La figure 2.3 illustre une approche de numérisation de visages utilisant les silhouettes proposée par [Lee *et al.* 2003].

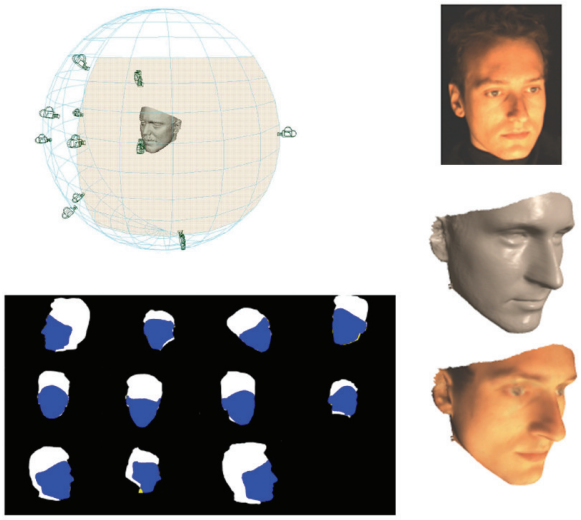


FIGURE 2.3 – Approche de numérisation de visage par les silhouettes [Lee *et al.* 2003].

Les différents points de vue sont obtenus à partir de plusieurs caméras si l'objet à scanner est mobile. Dans ce cas, les paramètres intrinsèques et extrinsèques des caméras doivent être connus [Niem & Wingbermuhle 1997]. Lors de l'acquisition d'un objet fixe, on peut acquérir les images au moyen d'une caméra qui se déplace dans la scène. Les techniques proposées sont rapides et robustes mais restent limitées à des objets de forme simple étant données les limitations induites par les silhouettes[Potmesil 1987, Schmitt & Yemez 1999].

2.3.3.2 Shape from motion

L'approche shape from motion (la forme à partir du mouvement) se ramène à analyser le mouvement d'un objet sur deux ou plusieurs trames successives capturées par une seule caméra pour estimer la profondeur de l'objet. La reconstruction 3D d'un visage à partir d'une séquence vidéo acquise par une seule caméra peut être considérée comme un problème stéréo. Il suffit d'estimer la disparité entre chaque trame et sa suivante. En d'autres termes, le mouvement d'un visage en face d'une seule caméra fixe peut être considéré comme le mouvement d'une caméra en face d'un visage fixe [Fua 2000]. Cette approche ne nécessite pas une étape d'étalonnage préalable de la caméra car les paramètres intrinsèques de la caméra sont estimés en ligne.

Comme en stéréovision, on distingue deux configurations pour le couple de caméras. La première consiste à mettre les deux caméras à une petite distance l'une de l'autre pour avoir une petite ligne de base ce qui génère des images stéréo très similaires et simplifie l'analyse de correspondance. Cependant, puisque l'angle de triangulation est petit, obtenir le point 3D par une intersection spatiale gauche-droite précise devient plus difficile. La deuxième configuration consiste à considérer une distance plus grande entre les deux caméras et donc une ligne de base plus grande ce qui rend l'intersection spatiale plus facile. Par contre, dans ce cas, des distorsions perspectives et des occultations plus sévères rendent l'analyse de la correspondance plus compliquée.

L'approche shape from motion combine les atouts de la stéréovision à faible ligne de base et les atouts de la stéréovision à grande ligne de base. En effet, l'utilisation de trames successives rend l'analyse de la correspondance plus facile. Aussi, les deux trames à apparier présentent une faible occultation. De plus, un suivi de points d'intérêt sur les trames

permet de faciliter l'estimation de l'intersection spatiale (flot optique) [Anke *et al.* 2008]. Cette approche a l'inconvénient d'être très sensible au bruit par apport à une méthode stéréo car le fait que le déplacement entre deux trames successives est très faible engendre une instabilité dans l'estimation de la disparité.

2.3.3.3 Shape from texture

L'approche shape from texture (la forme à partir de la texture) retrouve la forme 3D en analysant uniquement la texture de l'objet ou de la scène sur une seule image. Cette approche se prête bien à la reconstruction 3D d'objets ou de scènes ayant une texture bien particulière [Sansoni *et al.* 2009]. La texture doit disposer des caractéristiques homogènes qui peuvent être par exemple un motif basique qui se répète ou des caractéristiques fréquentielles. Une telle texture est considérée comme un ensemble d'éléments de texture appelés texels définis par le motif basique. Chaque texel décrit le motif en répétition avec une direction et une orientation différente. Ainsi, l'idée est de trouver les transformations possibles des texels pour reproduire l'orientation de la surface de l'objet [Y. Ohta & Sakai 1981, Aliomonos & Swain 1987]. Cette approche est à faible coût mais elle fournit une mesure 3D de faible qualité. De plus, elle ne permet pas de numériser des objets non-texturés comme le visage ou le corps. Un état de l'art sur cette approche est décrit dans [Nitzan 1988].

2.3.3.4 Shape from focus/defocus

Les techniques shape from focus/defocus (focalisation/défocalisation) profitent des propriétés intrinsèques de l'objectif d'une caméra pour extraire des paramètres de profondeur d'une scène, en s'affranchissant des problèmes liés à l'appariement stéréo. Deux méthodes permettent de calculer la profondeur : par focalisation (Shape from Focus) et par défocalisation (Shape from Defocus). Dans le premier cas, la mise au point s'effectue en chacun des points de l'image avec une focale propre à chaque point [Nayar *et al.* 1996, S.Birchfield & C.Tomasi 1999]. Pour la seconde technique, un nombre fini d'acquisitions d'images à différentes focales est réalisé. Ainsi, une fonction de flou est caractérisée pour chacun des points, permettant de déterminer leur profondeur. Les propriété-

tés de l'ouverture de l'objectif et la profondeur de champ du système optique d'acquisition d'image doivent être connues. Si la profondeur de champ est finie, le flou dépend de la distance entre l'objet et la caméra [Hirschmüller & Scharstein 2009].

Les systèmes de reconstruction 3D basés sur la focalisation sont limités en rapidité puisqu'ils nécessitent plusieurs prises de vues avec différentes mises au point. Ils souffrent également d'un manque de précision. En fait, la profondeur de champ est liée à l'agrandissement et au système optique d'observation et l'agrandissement varie avec la mise au point. En plus, il est très difficile d'extraire une information de profondeur suffisamment précise de la défocalisation [Favaro & Soatto 2002]. Une étude comparative des deux techniques de focalisation et de défocalisation a été publiée par Xiong et Shaffer [Xiong & Shafer 1993].

2.3.3.5 Shape from shading

Les méthodes d'estimation de forme à partir de l'ombrage, shape from shading, emploient une seule caméra et utilisent les variations d'intensité dans l'image pour estimer la forme d'un objet [B.Horn & Brooks 1989]. L'estimation de la profondeur en chaque point 3D de l'objet est assurée en analysant la brillance de la scène en ce point qui est définie par le niveau de gris de sa projection pixélique sur le plan image de la caméra. La brillance d'un point 3D de la surface de l'objet dépend de l'éclairage de la scène et de la forme de la surface. Elle dépend aussi des propriétés de réflectance de la surface et de la projection de l'image sur le capteur [Zhang *et al.* 1999].

Le calcul de la normale en tout point de la surface s'effectue grâce à la minimisation d'une fonctionnelle entre la brillance réelle de la surface de l'objet et la brillance obtenue par estimation de la carte de réflectance. Des contraintes sur l'intégrabilité et la forme de la surface peuvent être également prises en compte durant le calcul. Cette technique est de faible coût mais la qualité de la reconstruction n'est pas suffisante surtout en présence de facteurs externes influençant la réflectance de l'objet [B.Horn & Brooks 1989].

Puisque cette technique se base sur les propriétés de réflexion diffuse des surfaces lambertiennes, elle est très dépendante des conditions d'éclairage. Ainsi, les conditions d'éclairage de la scène doivent être parfaitement connues et les surfaces sont considérées comme lambertiennes [B.Horn & Brooks 1989]. En d'autres termes, l'intensité de la lumière réflé-

chie dépend de son incidence. De plus, les sources lumineuses doivent être suffisamment éloignées de telle sorte que l'illumination soit approximativement uniforme sur toute la surface [D.N.Bhat & S.K.Nayar 1998, Hernàndez *et al.* 2008].

2.3.4 Discussion

Le tableau 2.1 propose les différentes caractéristiques des approches passives. Le critère **Nbre de caméras** indique le nombre de caméras utilisées. Le critère **Nbre images/caméra** constitue le nombre d'images utilisées par chaque caméra du système. Le critère **Image de profondeur** permet de distinguer les approches qui fournissent une image de profondeur comme résultat ou non. Le critère **Directe** caractérise les approches qui fournissent directement une information de profondeur ou non. Finalement, le critère **Orientation de la surface** permet de discerner les approches qui fournissent des cartes d'orientation de la surface comme information de profondeur.

Approche passive	Nbre de caméras	Nbre images/caméra	Image de profondeur	Directe	Orientation de la surface
Stéréovision	**2**	**1**	**X**	**X**	
Photogrammétrie	**2**	**1**	**X**	**X**	
Shape from motion	**1**	**2**	**X**	**X**	
Shape from focus/defocus	**1**	**2**			
Shape from shading	**1**	**1**	**X**		**X**
Shape from silhouettes	**N>2**	**1**	**X**	**X**	
Shape from texture	**1**	**1**			**X**

TABLE 2.1 – Les différentes caractéristiques des approches passives.

2.4 Numérisation 3D active

Le principe de la numérisation active consiste à émettre un rayonnement et détecter sa réflexion afin de sonder un objet ou un environnement. Différents types de source de rayonnement sont utilisés : faisceau laser, lumière blanche, ultrason, rayon X ou lumière infrarouge. L'information 3D se calcule soit par l'estimation du temps de vol des rayons lumineux heurtant la surface de l'objet, par le principe de la triangulation optique ou encore par le principe de l'interférométrie ou de moiré.

2.4.1 Temps de vol

La technique de temps de vol utilise un télémètre laser pour mesurer le temps nécessaire au trajet aller-retour de l'impulsion d'un faisceau laser réfléchi [Schuon *et al.* 2009]. Connaissant la vitesse de la lumière c, le temps de retour t détermine la distance parcourue par la lumière est égale à $\frac{(c.t)}{2}$. Évidemment, la précision de la numérisation 3D dépend de celle de la mesure du temps de retour. Cette technique est rapide mais elle fournit un résultat très bruité ce qui nécessite une étape de post-traitement comme le proposent [Cui *et al.* 2010] qui utilisent une caméra TOF Swissranger SR4000 et font recourt à la super-résolution pour débruiter l'information 3D reconstruite.

2.4.2 Triangulation

Les deux approches actives qui se basent sur la triangulation optique et qui sont fréquemment utilisées sont la triangulation laser et l'approche de la lumière structurée.

2.4.2.1 Triangulation laser

Les techniques de triangulation laser projettent un point ou une bande laser pour mesurer la profondeur de l'objet en se basant sur le principe de la triangulation optique 2.4.

Les points O_P et O_C constituent les centres optiques respectivement de la source laser et de la caméra. Leur distance mutuelle est la droite de base d. Les axes optiques z_P et z_C de la source laser et de la caméra forment un angle de α [Sansoni *et al.* 2009]. La source ponctuelle laser génère un rayon lumineux qui heurte l'objet en un point S et sera réfléchi en un point S_i sur le plan image κ de la caméra. La mesure de la position (i_S, j_S) du

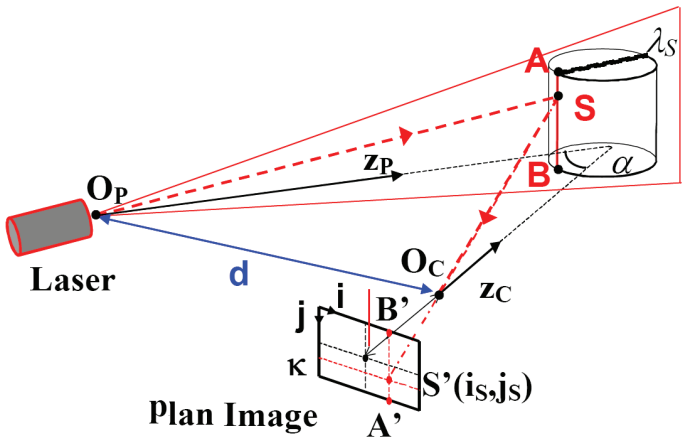


FIGURE 2.4 – Principe de la triangulation laser.

point S_i définit la ligne de mire $\overline{S'O_C}$ et retrouve par une simple géométrie la position de S. La surface complète 3D de l'objet est retrouvée par un balayage laser. Cette technique se caractérise par une haute précision mais le principe de balayage ne permet pas une numérisation en temps réel de l'objet surtout s'il est de grande dimension.

2.4.2.2 Lumière structurée

Cette famille d'approches utilise un couple caméra-projecteur et partage la même approche de triangulation optique que la stéréovision et la triangulation laser. Au lieu de balayer la surface, un patron bidimensionnel est projeté sur l'objet. L'information 3D est calculée pour tous les points simultanément en décodant la distorsion du patron projeté sur l'objet [Zhang 2010].

La lumière structurée peut comprendre un ou plusieurs patrons conçus par un encodage direct, ou un encodage spatial ou encore un multiplexage temporel [Salvi *et al.* 2004]. L'encodage direct consiste à projeter un patron ou un nombre limité de patrons sur le visage. Le décodage s'effectue pour chaque patron séparément dont chaque pixel est identifié uniquement par lui-même, c'est-à-dire par son intensité ou par sa couleur. Cette stratégie d'encodage présente une sensibilité très élevée au bruit et la qualité de la reconstruction dépend de la propriété de la réflexion de la surface faciale. L'encodage spatial consiste à projeter un seul patron. Chaque pixel du patron est encodé avec l'information contenue dans son voisinage en utilisant une codification non-formelle ou une codification mathématique en

utilisant les séquences de deBruijn par exemple. La figure 2.5 présente une reconstruction 3D d'une statue d'Einstein en utilisant une lumière structurée codée par les séquences de deBruijn [Zhang *et al.* 2003].

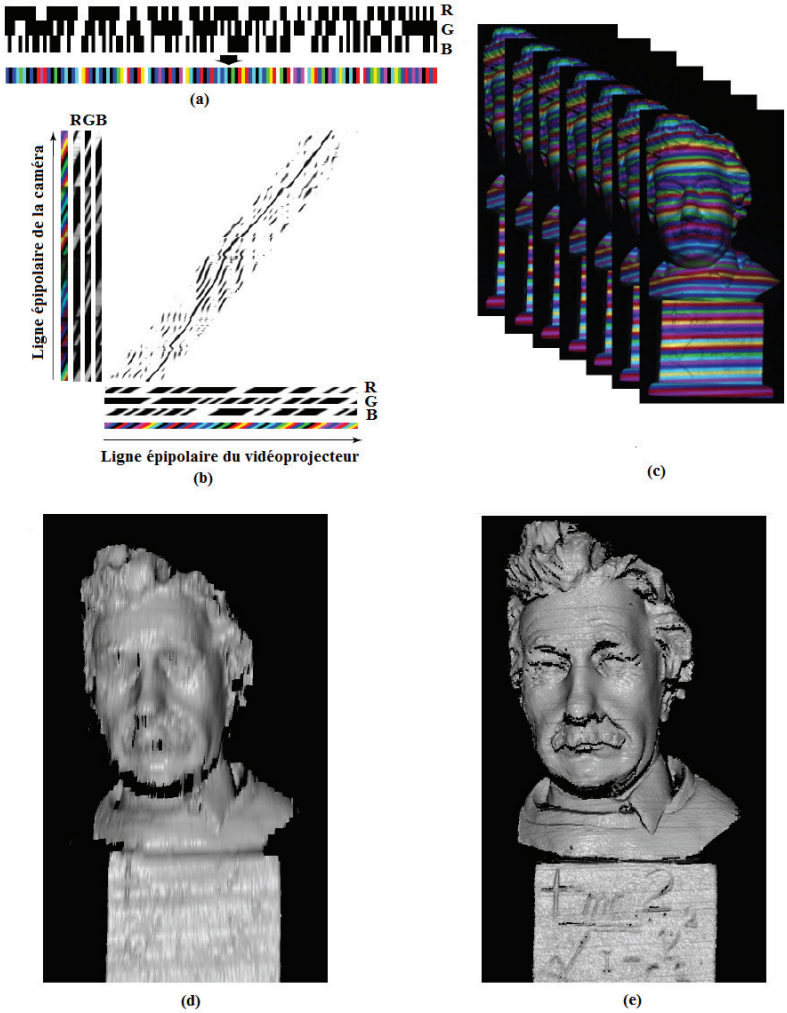


FIGURE 2.5 – Approche de reconstruction 3D par une lumière codée [Zhang *et al.* 2003].

Les auteurs proposent de projeter sur l'objet un patron couleur formé de trois patrons binaires rouge, vert et Bleu. Les trois patrons sont codés par une série de pixels noirs et blancs définie par la séquence de deBruijn [Fredricksen 1970] comme le montre la figure 2.5.a. La séquence de de Bruijn garantie que chaque trois transitions de couleur consécutives soient uniques. La figure 2.5.b affiche la matrice de score utilisée par la programmation dynamique pour trouver le chemin minimum et estimer la correspondance stéréo entre

le patron émis par le projecteur et le patron reçu distordu sur l'objet comme l'affiche la figure 2.5.c. La figure 2.5.d montre un modèle reconstruit en utilisant une seule capture 2D et la figure 2.5.e affiche le résultat final obtenu en faisant intervenir plusieurs captures 2D et en fusionnant plusieurs acquisitions 3D de faible résolution.

Les techniques qui utilisent l'encodage spatial présentent une plus grande difficulté dans l'étape du décodage du patron. En effet, puisque l'encodage tient en compte le voisinage de chaque pixel en plus du pixel lui-même, il est possible qu'en raison d'occultations ou/et ombrages, ce voisinage ne puisse être entièrement récupéré. Le multiplexage temporel consiste à projeter successivement une séquence de patrons sur le visage. Les patrons peuvent être binaires ou n-aires. Ils peuvent aussi suivre la codification de Gray ou la codification sinusoïdale par décalage de phase [Salvi *et al.* 2004]. Cette stratégie affecte ainsi une séquence de valeurs d'illuminations à chaque pixel du patron dans le temps. Elle permet d'estimer la profondeur pour chaque pixel en assurant à la fois une haute résolution et une bonne précision. Le multiplexage temporel se caractérise surtout par son implémentation facile.

Une des techniques les plus utilisées de multiplexage temporel est la codification sinusoïdale par décalage de phase [S.Zhang & Huang 2006, Huang & Zhang 2006]. En effet, trois patrons sinusoïdaux déphasés suffisent pour identifier chaque pixel à la différence des autres codifications de multiplexage temporel. Ainsi, la codification sinusoïdale permet une numérisation 3D avec une résolution pixélique. La codification sinusoïdale par décalage de phase exige une étape de déroulement de phase, sensible à l'éclairage ambiant surtout quand le nombre de patrons projetés est limité [Zhang & Yau 2008, Zhang 2010]. Finalement, la projection de plusieurs patrons impacte le délai d'acquisition et peut générer des artéfacts. Ceci constitue une limite majeure de cette famille d'approches surtout pour la capture d'un visage en mouvement. Autrement, le calcul de la phase peut être effectué en utilisant un seul patron par une analyse de sa distorsion dans l'espace fréquentiel de Fourier [Zhang 2010, Abdul-Rahman *et al.* 2008].

La figure 2.6 illustre la technique de codification sinusoïdale par décalage de phase proposée par [Huang *et al.* 2003] pour la numérisation 3D. Cette technique propose d'utiliser une roue de couleur pour envoyer successivement trois patrons déphasés de $\frac{\pi}{3}$. Les variations de relief produisent localement des étalements ou des resserrements de franges,

soit des variations de phase. En réalisant trois images de la même scène avec un décalage de phase connu, il est possible d'en extraire une mesure de profondeur. Un décodage des trois images obtenues permet de calculer les valeurs de phase locale pour chaque pixel de l'image. Ensuite, [Huang *et al.* 2003] utilisent une étape de déroulement de phase pour estimer les valeurs de phase absolues à partir des valeurs de phases locales en considérant un point ou un axe de référence préalablement défini au moment de l'étalonnage projecteur-caméra.

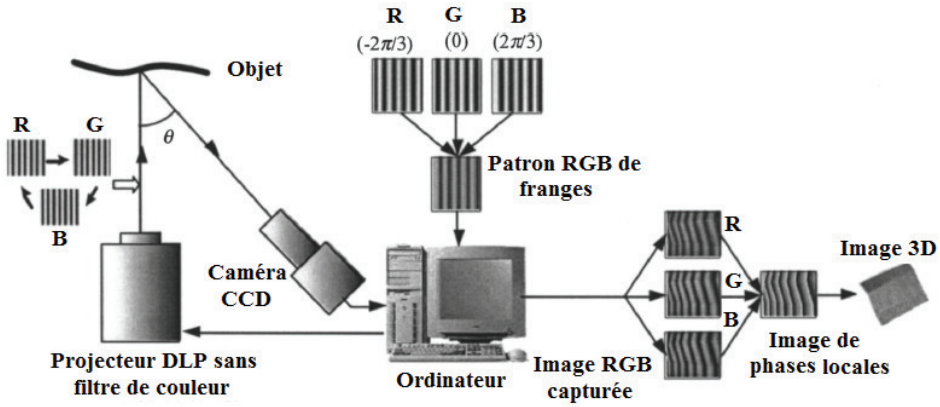


FIGURE 2.6 – Approche de reconstruction 3D par la lumière structurée sinusoïdale et le décalage de phase [Huang *et al.* 2003].

Lors de la projection du patron sinusoïdal par la source de lumière sur l'objet, la distorsion gamma rend des franges sinusoïdales parfaites non sinusoïdales avec une distorsion non-linéaire qui varie avec la phase. L'étape de la correction gamma est cruciale et permet de fournir une composante sinusoïdale efficace. Le déroulement de phase et la correction gamma constituent les deux sources majeures d'erreur. Un survol des différentes techniques de déroulement de phase a été étudié par [Huntley & Saldner 1997]. Aussi, une étude approfondie de l'importance de la correction gamma ainsi qu'une étude comparative des différentes solutions pour la correction gamma sont mises en œuvre dans [Zhang & Yau 2007]. Le principe de triangulation permet ensuite de retrouver la profondeur en utilisant les phases absolues.

Récemment, [Ettl *et al.* 2009] proposent une nouvelle technique appelée Triangulation à la volée (Flying Triangulation) qui utilise une source de projection et une caméra calibrés. Leur technique consiste à projeter des lignes verticales tout en déplaçant le senseur à

main levée pour balayer toute la surface de l'objet. Le principe de la triangulation optique permet de calculer une numérisation 3D non-dense d'une vue partielle de l'objet à partir de chaque image séparément en décodant les lignes distordues sur la surface de l'objet [Willomitzer *et al.* 2010]. Un appariement rigide robuste entre les images successives permet de fusionner les vues 3D partielles non-denses et de calculer un modèle dense complet de l'objet [Arold *et al.* 2009]. La figure 2.7 illustre la technique 'Flying Triangulation'. Pour la numérisation d'un moule dentaire par exemple, 500 images sont capturées dans 17 secondes pour obtenir un modèle 3D de 30 mm avec 3.6 millions de points. Une erreur de précision de 30μm est obtenue pour un volume numérisé de 15mm [Ettl *et al.* 2009]. Une étude de la précision de la technique 'Flying Triangulation' est mise en œuvre dans [Arold *et al.* 2010]. Cette technique permet une numérisation précise, complète et aisée d'un objet complexe en mouvement ou un objet de grande taille en absence d'une déformation élastique [Willomitzer *et al.* 2011].

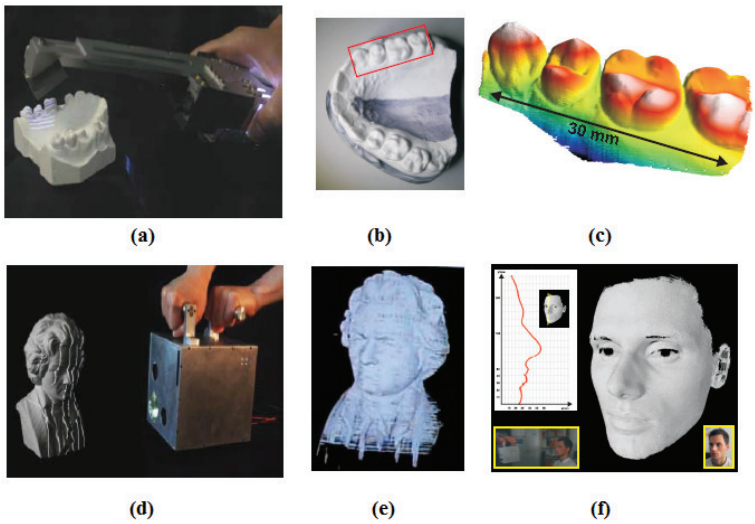


FIGURE 2.7 – Système de numérisation 3D à main levée qui utilise la technique 'Flying Triangulation' [Ettl *et al.* 2009]. a, b, c : numérisation d'un moule dentaire. d, e : numérisation d'un buste. f : numérisation d'un visage par la technique 'Flying Triangulation'.

2.4.3 Interférométrie et Moiré

L'interférométrie est une méthode de mesure 3D de haute précision utilisée surtout en industrie pour assurer un contrôle de qualité non destructif et sans contact. Le principe de

l'interférométrie repose sur la mesure de la différence de chemin optique entre deux ondes lumineuses pour caractériser la profondeur 3D. Cette technique utilise des patrons formés de franges. Ainsi, la mesure se base sur l'information de phase, qui varie de 2π sur des distances de quelques centaines de nanomètres [Surrel 2004].

L'idée derrière la mesure interférométrique de forme, c'est que les franges sont formées par la variation de la matrice de sensibilité qui relie la forme géométrique d'un objet avec les phases optiques mesurées [Zhang 2005]. Cette matrice contient trois variables : la longueur d'ondes, l'illumination et les directions d'observations, à partir desquelles trois méthodes sont essentiellement dérivées. La première est à plusieurs longueurs d'onde, la deuxième se base sur le changement de l'indice de réfraction et la dernière se base sur la variation de la direction de l'illumination ou l'utilisation de deux sources de lumière [Chen *et al.* 2000].

La technique de Moiré projette une lumière à travers deux grilles superposées et légèrement décalées, formées chacune de traits équidistants alternativement opaques et transparents. La distorsion des franges moiré générées permettent ensuite de mesurer l'information 3D [Surrel 2003]. La figure 2.8 affiche un exemple de l'effet moiré. La description mathématique des patrons résultats de la superposition des franges sinusoïdales est la même que celle qui définit les patrons d'interférence formés par des ondes électromagnétiques. L'effet moiré est ainsi souvent qualifié comme une interférence mécanique [Zhang 2005].

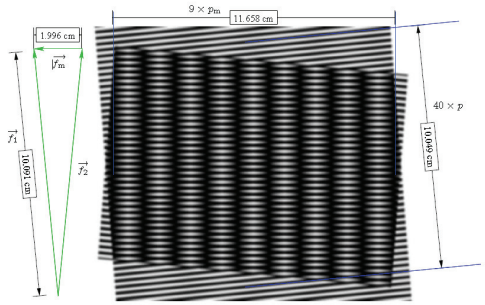


FIGURE 2.8 – Un exemple de l'effet moiré [Surrel 2004].

2.4.4 Discussion

Les approches actives fournissent généralement une meilleure précision que les approches passives. Cependant, une numérisation active d'un visage par exemple est souvent intrusive et nécessite une collaboration explicite du sujet. L'utilisation de la projection de lumière est aussi utilisée pour renforcer les approches passives de stéréovision ou de shape from X. En effet, une solution pour réussir la stéréovision pour les objets mal-texturés consiste à projeter une lumière codée pour texturer l'objet pour garantir la convergence du processus d'appariement. [Zhang & Yau 2008] utilisent un système étalonné formé d'un vidéoprojecteur et de deux caméras. Ils projettent trois patrons sinusoïdaux, calculent les valeurs de phases locales pour chaque pixel. Ensuite, ils proposent d'assurer un appariement stéréo sur les cartes 2D de phase gauche et droite au lieu de l'assurer sur les images elles mêmes. La densité du modèle 3D calculé par la stéréovision active dépend de la richesse du patron texturant et de l'aptitude de la lumière codée à cerner le maximum de détails de la forme. L'utilisation de mesures de similarités spatio-temporelles a été aussi proposée dans [Zhang *et al.* 2004] pour des meilleurs résultats.

Par ailleurs, dans [Nayar *et al.* 1996], une approche active de shape from defocus est proposée par une projection d'une lumière structurée au cours de la numérisation. Aussi, [Hertzmann & Seitz 2003] introduisent une projection de lumières multiples pour optimiser une approche de shape from shading, en utilisant des sources lumineuses préalablement étalonnées. Cette alternative de shape from shading active s'appelle stéréophotométrie. Une étude plus approfondie sur l'état de l'art de la reconstruction 3D a été élaborée par [Blais 2004]. Aussi, une base standard de test ainsi qu'une évaluation des différentes techniques de shape from X sont proposées par [Seitz *et al.* 2006] et [Anke *et al.* 2008].

Le tableau 2.2 propose les différentes caractéristiques des approches actives. Nous rappelons que le critère **Nbre de caméras** indique le nombre de caméras utilisées. Le critère **Nbre images/caméra** constitue le nombre d'images utilisées par chaque caméra du système. Le critère **Image de profondeur** permet de distinguer les approches qui fournissent une image de profondeur comme résultat ou non. Le critère **Directe** signale les approches qui fournissent directement une information de profondeur ou non. Le critère **Orientation de la surface** permet de discerner les approches qui fournissent des cartes d'orientation de

la surface comme information de profondeur.

Approche active	Nbre de caméras	Nbre images/caméra	Image de profondeur	Directe	Orientation de la surface
Triangulation laser	**1**	**1**	**X**	**X**	
Time-Of-Flight	**1**	**1**	**X**	**X**	
Lumière structurée	**1**	**N>=3**	**X**	**X**	
Interférométrie	**1**	**N>=2**	**X**	**X**	
Moiré	**1**	**1**	**X**		
Stéréovision active	**2**	**N>=1**	**X**	**X**	
Active shape from defocus	**1**	**1**	**X**		
Stéréophotométrie	**1**	**N>=2**			**X**

TABLE 2.2 – Les différentes caractéristiques des approches actives.

2.5 Numérisation 3D hybride

Le modèle 3D calculé par une des approches décrites ci-dessus peut comporter des erreurs et des distorsions. Des approches hybrides de numérisation 3D ont été proposées pour débruiter, corriger et densifier les modèles 3D calculés. Nous pouvons citer essentiellement quatre axes de recherche adoptés. Un premier axe fait assister le processus de la numérisation 3D par un modèle générique. Un deuxième axe suggère de fusionner différentes approches de numérisation 3D de l'état de l'art. Ceci permet l'élaboration d'une nouvelle approche hybride qui remédie aux défaillances et minimise la sensibilité des approches fusionnées et surtout de profiter de leurs atouts. Un troisième axe propose de calculer l'information 3D par une approche spatio-temporelle. Il s'agit d'utiliser non-seulement l'information pixélique acquise à un instant t par une ou plusieurs caméras mais plutôt d'utiliser leurs trames précédentes et suivantes pour optimiser la reconstruction 3D. Enfin, le principe de la super-résolution 2D a été adapté pour les données 3D pour améliorer la qualité de la reconstruction et aussi pour la corriger ce qui constitue notre quatrième axe

[Schuon *et al.* 2009, Pan *et al.* 2006].

2.5.1 Approches assistées par un modèle

Souvent, un modèle générique est utilisé pour assister une numérisation 3D. Par exemple, l'utilisation du modèle générique peut assister une approche de shape from motion. Dans ce cas, il permet l'extraction des paramètres intrinsèques et extrinsèques de la caméra ainsi que la géométrie épipolaire du mouvement de la caméra à partir de chaque couple de trames successives [Fua 2000, Fidaleo & Medioni 2007]. La technique de l'ajustement de faisceaux (bundle-adjustment) est souvent utilisée pour optimiser simultanément la trajectoire de la caméra et la structure de la scène [Fua 2000].

En général, il est supposé que le visage n'a pas de variation d'expression entre deux trames successives. Une sélection manuelle de quelques points caractéristiques est nécessaire pour réussir la mise en correspondance entre la trame et le modèle générique. Une utilisation d'un modèle déformable 3D pour la reconstruction 3D d'un visage est proposée par [Blanz & Vetter 2003, Ilic & Fua 2006, Wang *et al.* 2005, Kang & Byun 2008]. Les approches assistées par un modèle déformable 3D permettent d'extraire les paramètres de mouvement de la caméra, les paramètres définissant la forme et la texture du visage en utilisant un principe d'analyse par synthèse [Blanz & Vetter 2003]. [Kang & Byun 2008] développent un algorithme de mise en correspondance hiérarchique assisté par un modèle déformable en échantillonnant le modèle déformable et les séquences d'images 2D par un filtre gaussien.

2.5.2 Fusion d'approches

La fusion de techniques de numérisation 3D a été proposée essentiellement sur des approches passives pour récupérer le maximum de détails de l'objet scanné sans projeter un patron particulier sur l'objet. Klaudiny et Al. proposent une technique de numérisation 3D multicouche et semi-automatique qui réussit à récupérer des détails comme les rides sur un visage [Klaudiny *et al.* 2010]. Selon leur approche, d'abord ils reconstruisent un modèle épars du visage en utilisant des marqueurs préalablement dessinés sur le visage. Un modèle plus dense est ensuite calculé par un appariement stéréo basé sur une optimisation par la

technique Graph-cut. Des détails fins comme les rides et les sourcils sont ensuite estimés par une approche de stéréophotométrie. La technique de stéréophotométrie est une variante de l'approche de shape from shading assistée par une source de lumière. Il s'agit d'estimer les normales à la surface en décryptant la lumière colorée projetée sur le visage lors de l'acquisition comme l'illustre la figure 2.9.

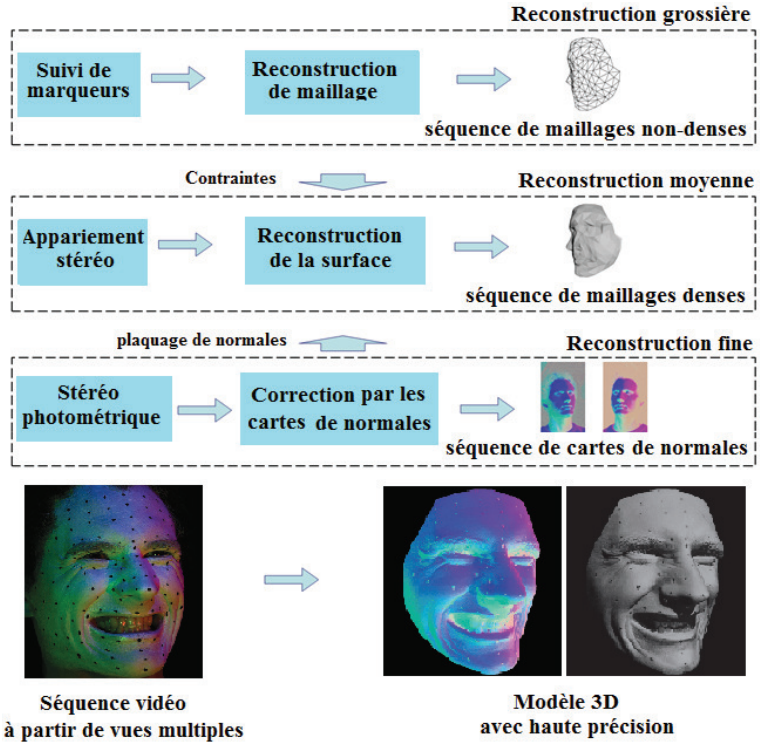


FIGURE 2.9 – Fusion d'une approche de stéréophotométrie et de stéréovision [Klaudiny *et al.* 2010].

Wu et Al. proposent une approche passive hybride qui fusionne la stéréovision multi-caméra et la technique de shape from shading [Wu *et al.* 2011]. La stéréovision fournit une estimation initiale de la forme 3D. Un calcul des normales à la surface est ensuite assuré par la technique de shape from shading ce qui permet de retrouver beaucoup plus de détails et enrichit la forme 3D rendue. Généralement, la technique de shape from shading nécessite un étalonnage préalable de la source de lumière. Wu et Al. montrent que leur technique est robuste en présence d'une illumination quelconque non-étalonnée grâce à la stéréovision. Yoshiyasu et Al. proposent de combiner une approche de silhouettes visuelles avec une approche passive de shape from shading et montrent une aptitude de leur

méthode à fournir une numérisation efficace, précise et fidèle à la forme réelle de l'objet [Yoshiyasu & Yamazaki 2011].

Han et Al. proposent de fusionner la stéréovision et la lumière structurée par décalage de phase en appliquant le processus d'appariement stéréo sur des cartes de phases au lieu de les appliquer sur l'information pixélique directement [Han *et al.* 2009]. Les deux caméras et le vidéoprojecteur utilisés sont étalonnés. Weise et Al. proposent aussi une technique hybride de stéréovision active et de décalage de phase avec une nouvelle approche de déroulement de phases stéréo (Stereo unwrapping) [Weise *et al.* 2007]. Leur technique nécessite aussi un étalonnage préalable des deux caméras et du vidéoprojecteur. Pour retrouver la phase absolue à partir des phases locales estimées par la technique de la lumière structurée, les auteurs estiment le nombre k de périodes à ajouter par un principe d'appariement stéréo. Ainsi, pour chaque pixel p_{gauche} calculé par la caméra gauche et ayant une phase $\phi_{abs} = \phi_{loc} + 2k\pi$, on calcule la mesure de similarité entre son niveau de gris et le niveau de gris des N pixels calculés par la caméra droite ayant une phase locale égale à ϕ_{loc}. La valeur N représente le nombre de périodes qui constituent la lumière structurante envoyée par le vidéoprojecteur. La valeur k retenue pour la phase absolue est celle qui fournit la plus grande valeur de similarité. Aussi, les coordonnées 3D de chaque pixel est estimée par triangulation entre le vidéoprojecteur et la caméra.

2.5.3 Approche spatio-temporelle

[Zhang *et al.* 2003] proposent une approche de stéréovision active et l'utilisation d'une fenêtre spatio-temporelle dans leur algorithme d'appariement stéréo pour une meilleure reconstruction de scènes dynamiques. La figure 2.10 présente l'apport de l'utilisation de la fenêtre spatio-temporelle sur la qualité du modèle reconstruit. Elle compare la reconstruction d'un visage par stéréovision spatiale avec une fenêtre d'une largeur et d'une hauteur égales de 15 pixels avec une reconstruction par stéréovision spatio-temporelle utilisant une fenêtre d'une largeur de 9 pixels, d'une hauteur de 5 pixels sur un voisinage temporel de 5 trames. La reconstruction 3D spatio-temporelle fournit un rendu 3D plus efficace et plus stable qu'une reconstruction 3D spatiale.

Nous citons aussi le travail de Davis et Al. qui analyse l'influence de différentes fe-

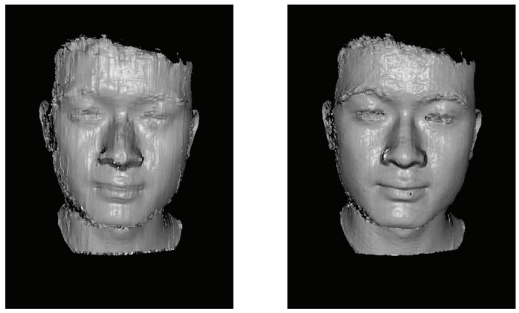


FIGURE 2.10 – Comparaison qualitative d'une reconstruction 3D spatiale et d'une reconstruction 3D spatio-temporelle d'un visage [Zhang *et al.* 2003].

nêtres spatio-temporelles sur la convergence de l'appariement stéréo et sur la qualité de la reconstruction 3D finale [Davis *et al.* 2003]. Aussi, il propose une classification intéressante des différentes approches de numérisation 3D spatiales, temporelles et spatio-temporelles. D'autres applications de la stéréovision spatio-temporelle ont été aussi explorées, par exemple le travail de Shechtman et al. qui suggère un système multi-caméra avec une approche spatio-temporelle pour l'optimisation de la résolution des séquences vidéo [Shechtman *et al.* 2002].

2.5.4 Super-résolution

Les techniques de super-résolution ont été utilisées surtout en 2D. La super-résolution peut être mono-image ou multi-image [Baker & Kanade 2002]. En fait, l'image de haute résolution peut s'estimer en utilisant une seule image de faible résolution et une base d'apprentissage ou en fusionnant une séquence d'images 2D de faible résolution sans apprentissage préalable. La super-résolution mono-image utilise des images naturelles via une base d'apprentissage, pour interpoler les valeurs des pixels. La super-résolution multi-image nécessite le recalage des images de faible résolution, leur fusion, et une étape de débruitage qui se base généralement sur un principe d'optimisation. Nous pouvons distinguer les approches qui effectuent un recalage 2D sur les images de profondeur, des approches qui assurent un recalage 3D sur les données 3D directement.

2.5.4.1 Application sur les images 2D de profondeur

Pour résoudre le problème de super-résolution 3D, plusieurs chercheurs utilisent une technique simple de super-résolution 2D issue de l'état de l'art. Elle combine plusieurs images 2D de faible résolution pour estimer une image de haute résolution. Il en profitent pour résoudre le problème sur les images de profondeur puisque la super-résolution 2D a montré son efficacité comme décrit par [Park *et al.* 2003]. Cette stratégie a été utilisée par Rosenbush et al. pour les systèmes de télédétection par laser LADAR (Laser Detection and Ranging) [Rosenbush *et al.* 2007]. Les systèmes LADAR fournissent des mesures de profondeur en utilisant un détecteur du tableau de plan focal FPA (Focal Plane Array Detector). La technique LADAR consiste à émettre de la lumière infrarouge et collecter ensuite son signal réfléchi sur la scène. En calculant le décalage de phase entre les deux signaux émis et renvoyé, les distances sur la scène sont estimées. Puisque le nombre de pixels disponibles sur un FPA est limité (environ 256x256 pixels), les systèmes LADARs sont incapables d'atteindre la densité pixélique des scanners lasers. Pour remédier à ce problème, les auteurs proposent d'enrichir la résolution spatiale des images de profondeur obtenues. Quatre images de faible résolution et légèrement décalées et déplacées par apport à une image référence sont recalées en utilisant les propriétés de la transformée de Fourier et l'image de haute résolution est ensuite obtenue par une interpolation cubique non-uniforme. Les auteurs montrent que la super-résolution fournit une image de haute résolution fidèle aux données de départ et constitue une solution adéquate pour améliorer la qualité de la numérisation 3D LADAR.

Plus récemment, le principe de la super-résolution à été appliqué aux images de profondeur. Schuon et al. [Schuon *et al.* 2008] ont appliqué la méthode de super-résolution 2D de Farsiu et al. [Farsiu *et al.* 2004] sur les images de profondeurs acquises par une caméra ToF 3DVTM. Comme proposé déjà dans plusieurs approches de super-résolution 2D, ils ramènent le problème de super-résolution à un problème de minimisation d'énergie qui emploie jointement un terme de données et un terme bilatéral de régularisation préservant les contours. Le terme de données permet de renforcer la similarité entre les images de profondeur d'entrée de faible résolution et l'image de profondeur de sortie de haute résolution. Une étape de recalage 2D préalable entre les images de profondeur s'avère nécessaire. Elle

est assurée en utilisant le principe du flot optique.

2.5.4.2 Application directe sur les modèles 3D

Ici, la super-résolution se ramène à une fusion d'un ensemble de modèles 3D de faible qualité qui représentent un objet ou une scène avec une légère variation du point de vue. Les approches existantes de super-résolution 3D s'appliquent sur des scènes statiques. Kil et al ont utilisé la super-résolution pour les scanners de triangulation laser [Kil *et al.* 2006]. Ils ont utilisé 100 modèles 3D numérisés à partir de points de vue similaires pour créer un modèle 3D ayant quatre fois plus de points. L'étape de recalage est assurée par une variante de l'algorithme ICP (Iterative Closest Point). Puisque les données de départ étaient suffisamment denses et que le taux de bruit était faible, ils ont pu obtenir de bons résultats en recalant les 100 modèles et en assurant un rééchantillonnage régulier menu d'une mesure gaussienne d'incertitude de localisation [Kil *et al.* 2006]. Leurs résultats font apparaitre un effet de flou et leur approche ne fonctionne pas sur des données très bruitées ou qui présentent une déformation non-rigide quelconque.

La technique de reconstruction 3D par temps de vol (Time-Of-Flight) produit des modèles 3D très bruités et le post-traitement des modèles résultats est une étape nécessaire [Cui *et al.* 2010]. Ainsi, la super-résolution 3D a été particulièrement étudiée et appliquée pour cette famille de techniques TOF essentiellement pour débruiter leurs résultats comme le montre la figure 2.11.

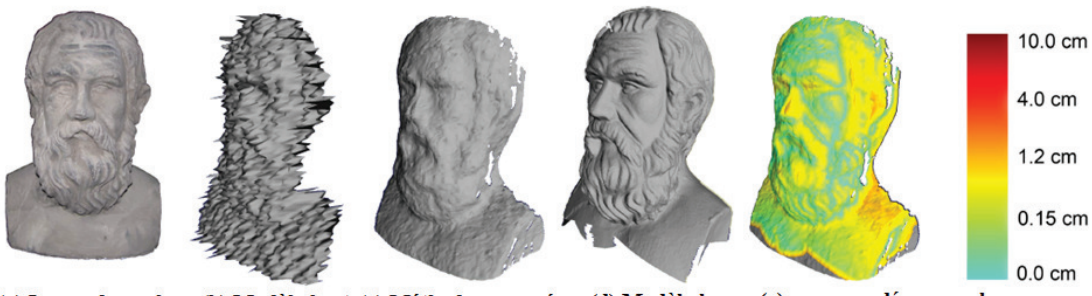


FIGURE 2.11 – Approche de super-résolution 3D proposée par [Cui *et al.* 2010].

L'étape de recalage entre les modèles 3D de faible résolution est effectuée par une

approche d'optimisation probabiliste utilisant les modèles de mélanges gaussiens GMM (Gaussian Mixture Models). Rajagopalan et al. ont proposé d'améliorer la résolution des modèles TOF en utilisant les champs de Markov aléatoires (MRF : Markov-Random-Field). Leur MRF emploie un système de voisinage sur les modèles 3D qui renforce un apriori de régularisation préservant les contours entre chaque couple de points voisins [Rajagopalan *et al.* 2008]. Leur formulation du problème de super-résolution crée deux inconvénients. D'une part, le choix des paramètres est une tâche complexe. D'autre part, la formulation de l'apriori rend le problème non-convexe et des solutions plus sophistiquées sont nécessaires pour résoudre le problème.

2.5.5 Discussion

Les approches de numérisation 3D hybrides permettent de récupérer des détails très fins du visage comme les rides et les sourcils et sont capables de reconstruire des visages 3D en temps réel et de haute qualité. Cependant, les approches assistées par un modèle statistique souffrent généralement de la corrélation non nulle entre le modèle reconstruit et le modèle statistique utilisé [Fidaleo & Medioni 2007]. Aussi, même si la fusion d'approches a réussi à améliorer la qualité de la reconstruction, le coût de la numérisation devient plus élevé. En plus, la déformation non-rigide engendrée par une expression faciale reste encore un défi à surmonter. Ainsi, en présence d'une variation de l'expression faciale, la fenêtre spatio-temporelle utilisée dans les approches de stéréovision spatio-temporelles n'est plus pertinente et risque d'engendrer des artéfacts dans la reconstruction 3D finale. De même, puisque la super-résolution nécessite une étape de recalage, elle ne peut assurer une reconstruction 3D efficace que si le recalage utilisé considère l'aspect déformable du visage. Nous allons dans le paragraphe suivant introduire les approches de numérisation 3D qui considèrent l'aspect déformable du visage et qui sont capables de capturer le mouvement facial lors d'une acquisition 3D.

2.6 Modélisation d'une déformation non-rigide

La modélisation des déformations non rigides d'un objet est une étape cruciale et nécessaire pour réussir une acquisition 3D surtout si l'approche de numérisation emploie des

mesures spatio-temporelles ou applique une super-résolution temporelle. Pour un visage, les déformations non-rigides sont engendrées par une variation de l'expression faciale par exemple. La capture de l'expression faciale trouve aussi son application dans le domaine de l'animation faciale et la synthèse de visages 3D à partir d'une seule vue 2D.

L'utilisation de marqueurs pour la capture des déformations non-rigides et particulièrement les expressions faciales constitue la technique la plus utilisée dans le domaine de l'animation faciale. Une seule caméra est généralement utilisée avec un marquage préalable d'un ensemble de points sur le visage du sujet. Il s'agit de former un maillage grossier en utilisant les marqueurs. Le suivi des marqueurs à travers la vidéo permet de déformer le maillage et caractériser la déformation faciale. Ainsi, plus le nombre de marqueurs augmente, plus la déformation capturée est fidèle à la déformation réelle. Les points du maillage sont ensuite utilisés comme des points d'appui pour déformer le modèle 3D du visage à animer [Guenter *et al.* 1998]. Le travail de Williams en 1990 emploie une seule caméra et assure un suivi des points uniquement en $2D$ [Williams 1990]. Leur résultat est intéréssant mais les expressions générées ne sont pas assez réalistes. Nous pouvons distinguer essentiellement deux axes de recherche dans la modélisation dense des déformations non-rigides. Le premier axe élabore un modèle statistique de déformation. Le second axe propose une modélisation géométrique de la déformation non-rigide.

2.6.1 Approche statistique

Dans cette famille d'approches, nous distinguons les modèles actifs de l'apparence et les modèles déformables. La technique du modèle actif de l'apparence a été proposée par Cootes et Al. [Cootes *et al.* 2001]. En effet, cette approche consiste à créer un modèle statistique combinant un modèle de variation de forme avec celui de variation de l'apparence effectué sur une image normalisée en forme. Le modèle de forme s'obtient par une analyse en composantes principales sur les vecteurs de forme normalisés qui sont constitués par des points marquant la forme du visage sur l'image 2D. Pour construire le modèle d'apparence, les images de la base sont modifiées de façon à faire coïncider les points du visage avec la forme moyenne pour extraire par la suite l'information texture. Une nouvelle analyse en composantes principales s'applique sur les vecteurs concaténant les paramètres

de forme et de texture pour donner naissance à un modèle combiné. La déformation d'un visage se fait par une méthode d'optimisation comme la méthode de descente de gradient [Cootes *et al.* 2001, Edwards *et al.* 1998]. Chuang et al. utilisent quant à eux une ACP combinée à un modèle bilinéaire pour synthétiser une nouvelle expression sur un visage parlant [Chuang *et al.* 2002]. Kang et al. utilisent le modèle actif d'apparence combiné avec une régression linéaire pour annuler l'expression faciale d'un visage dans le but d'améliorer les performances d'un algorithme de reconnaissance de visages [Kang *et al.* 2002].

Introduite par l'équipe de Thomas Vetter de l'université de Basel [Blanz & Vetter 1999], la technique du modèle déformable consiste à construire un modèle de forme et un autre de texture pour tous les visages 3D d'une base d'apprentissage. Ceci est fait après avoir effectué une mise en correspondance dense de tous les visages 3D avec un visage de référence moyennant le flot optique. Le modèle déformable modélise les visages 3D et les expressions 3D comme des combinaisons linéaires de ses vecteurs propres. Une phase d'adaptation se fait par la suite par le biais d'une approche d'analyse par synthèse pour construire un visage 3D à partir d'une seule image 2D. Elle consiste à chercher les paramètres de forme et de texture correspondant à un modèle 3D qui ressemble le plus à l'image 2D introduite [Romdhani & Vetter 2003]. La synthèse d'une nouvelle expression consiste à accorder les poids adéquats aux vecteurs propres du modèle déformable pour changer l'expression du visage 3D reconstruit.

2.6.2 Modélisation géométrique

La capture dense d'une déformation faciale consiste dans un suivi dense de tous les points de la surface faciale à travers la vidéo. Wang et Al. développent une méthode automatique qui assure un suivi dense et non-rigide des points 3D d'un modèle facial parlant [Wang *et al.* 2008]. Ils proposent d'utiliser les cartes harmoniques en imposant des contraintes de correspondance pour assurer une correspondance 3D dense et non-rigide entre les trames successives [Schoen & Yau 1997, O'Neill 2001]. Ainsi, l'appariement de deux modèles 3D ayant une déformation non-rigide se ramène à un appariement de leurs cartes harmoniques respectives. La théorie des cartes harmoniques se base sur la théorie de la géométrie conformale [Gu & Yau 2003, Sharon & Mumford 2004]. Une carte har-

monique entre deux disques topologiques constitue un difféomorphisme avec une énergie d'étirement minimale et une distorsion bornée de l'angle. La carte est stable, insensible à la variation de la résolution et robuste au bruit. Les contraintes de continuité et de régularité implicites et explicites permettent un appariement/déformation régulier et continue. Il fournit aussi une correspondance point-à-point entres les trames. Cependant, leur approche devient moins robuste si les variations non-rigides entre les trames est importante.

Bronstein et al. ont développé une autre approche qui assure un suivi dense des déformations non-rigides en utilisant la géométrie riemannienne [Bronstein *et al.* 2006]. Ils ont proposé un nouvel algorithme appelé GMDS (Generalized Multidimensional Scaling) qui transforme la surface faciale définie dans l'espace réel muni de la distance euclidienne vers une nouvelle représentation surfacique isométrique dans un nouvel espace riemannien muni de la distance géodésique. Bronstein et al. ont utilisé cette approche pour la reconnaissance faciale en utilisant un appariement 3D rigide entre les modèles de la galerie et du modèle à reconnaitre dans le nouvel espace isométrique [Bronstein *et al.* 2005].

2.6.3 Discussion

La capture de la déformation faciale peut servir dans un scénario de reconnaissance faciale, dans l'animation faciale ou aussi pour une annulation de l'expression faciale. L'utilisation de marqueurs sur le visage est une approche intrusive qui nécessite une coopération particulière du sujet scanné. De plus, l'efficacité de cette technique dépend du nombre de marqueurs considérés et ne permet pas d'estimer la déformation faciale pour chaque point du modèle 3D. L'utilisation du modèle statistique permet de synthétiser un visage avec des expressions réalistes à partir d'un visage neutre mais la projection du visage dans l'espace propre génère un visage qui présente une corrélation non nulle avec le modèle statistique utilisé. Les techniques non-rigides d'appariement dense ont montré leur efficacité dans la capture des expressions faciales. Cependant, ces techniques considèrent le visage comme une surface non élastique ce qui n'est pas toujours le cas surtout en présence d'une variation sévère de l'expression faciale.

2.7 Technologies commercialisées

Aujourd'hui, les technologies de numérisation 3D commercialisées peuvent être catégorisées en trois familles : le balayage laser, la lumière structurée par projection de patrons de lumière blanche et la numérisation 3D passive.

2.7.1 Balayage laser

La figure 2.12 présente trois systèmes de balayage laser commercialisés. Une unité de balayage laser est composée par une source ponctuelle laser, un système optique de miroirs pour assurer le balayage et un capteur de lumière. Cette unité est déplacée pour balayer tout le corps humain.

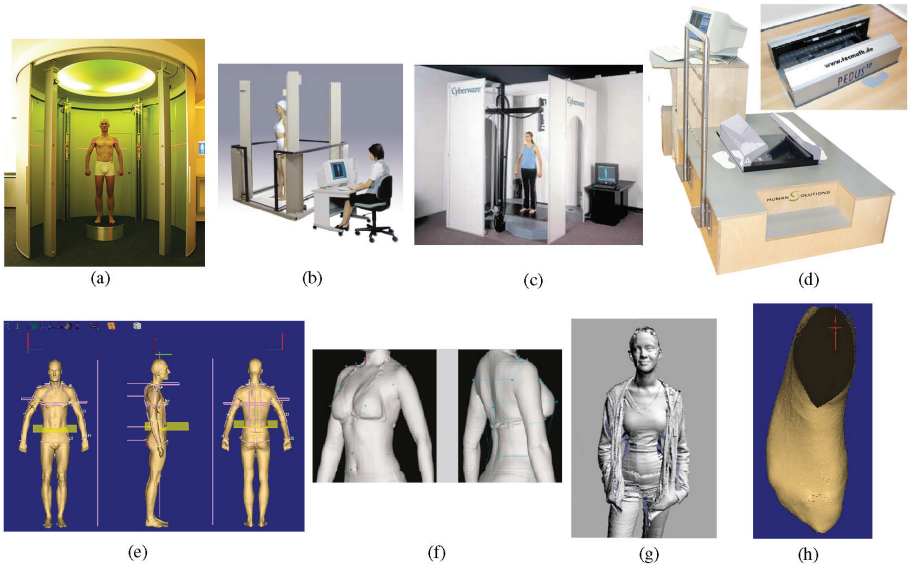


FIGURE 2.12 – Systèmes de balayage laser. a : Le système VITUS Smart XXL de numérisation de corps humain de la compagnie Human Solutions , b : Scanner du corps humain BodyLine de Hamamatsu. c : Scanner de corps humain WBX de la compagnie Cyberware. d : Scanner de pieds Pedus de Human Solutions. e : Un exemple de numérisation par le système VITUS Smart XXL, f : Une numérisation effectuée par le scanner BodyLine. g : Une numérisation par le scanner WBX. h : Un pied numérisé par le scanner Pedus.

La figure 2.12.a montre le système VITUS Smart XXL de numérisation de corps humain de la compagnie Human Solutions (www.human-solutions.com) qui dispose de huit unités laser. La figure 2.12.b présente un scanner du corps humain BodyLine proposé par la compagnie Hamamatsu Photonics (www.hamamatsu.com). Leur système dispose de quatre

unités laser. Le système de la figure 2.12.c est le scanner de corps humain WBX proposé par la compagnie Cyberware (www.cyberware.com) aux états unis. Ce système dispose de quatre unités laser. Le système de la figure 2.12.d est un scanner de pieds Pedus proposé par la société Human Solutions. Ce dernier système dispose de trois unités laser. Les systèmes de numérisation laser sont coûteux et lents, le temps d'acquisition pour le visage ou le corps peut varier d'une à plusieurs secondes. Ceci les rend inadéquats pour l'acquisition d'un visage ou corps animé [D'apuzzo 2006].

2.7.2 Lumière structurée

La deuxième famille de systèmes 3D commercialisés emploie une lumière structurante et estime la forme 3D par décodage du patron distordu sur l'objet. Elle constitue la technique la plus utilisée pour la numérisation 3D de visages et de corps. Au lieu de déplacer l'unité laser comme proposé par le système laser, une projection de patrons lumineux souvent sous forme de franges permet de récupérer l'information 3D de toute la surface en moins d'une seconde. Une ou plusieurs caméras numériques et une source de projection lumineuse permettent de numériser la scène. Cependant le champ de couverture du système de numérisation est limité.

Par exemple, le système de numérisation InSpeck acquis en 2010 par la société Creaform (www.creaform3d.com) et présenté sur la figure 2.13.a, peut mesurer au maximum la moitié de la surface du corps humain. La figure 2.13.a illustre un scénario de numérisation d'un visage par une projection de franges horizontales par le scanner InSpeck. Pour mesurer des parties plus larges du corps humain, plusieurs couples projecteur + caméra doivent être utilisés. Cette procédure a l'inconvénient de ne pas pouvoir être utilisés en même temps pour éviter l'interférence éventuelle entre les patrons projetés par les six unités ce qui implique encore une extension du temps d'acquisition dans le cas des surfaces larges [D'apuzzo 2006].

Le système FACESCAN^{3D} (www.3d-shape.com) de la compagnie 3D-Shape propose de mesurer l'ensemble du visage d'une oreille à l'autre (> 180) avec un seul coup grâce à un dispositif de miroirs comme le montre les deux figures 2.13.c, 2.13.d et 2.13.e. La figure 2.13.f présente le scanner Mephisto EX de la compagnie 4DDynamics et les deux figures

2.13.g et 2.13.h illustrent deux acquisitions 3D calculées par le scanner Mephisto EX de la compagnie 4DDynamics (www.4ddynamics.com).

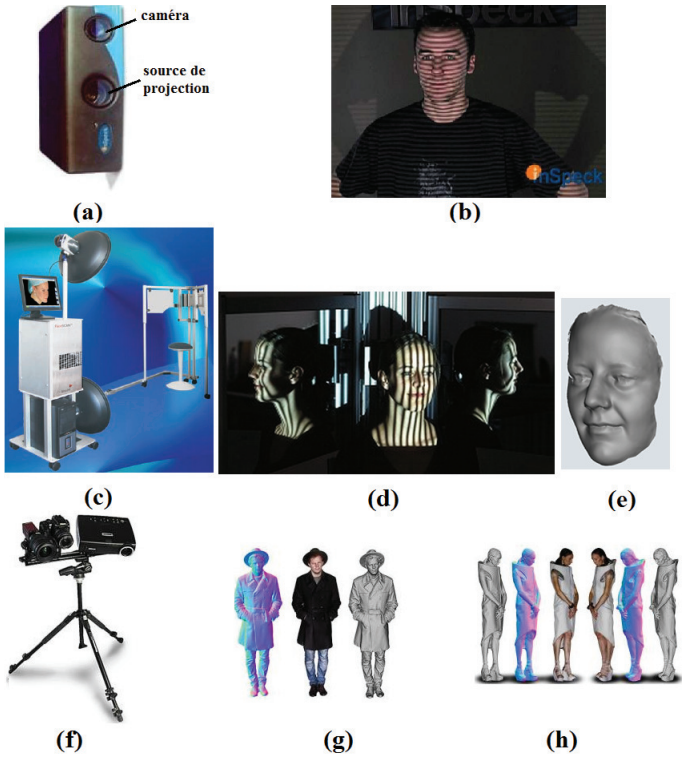
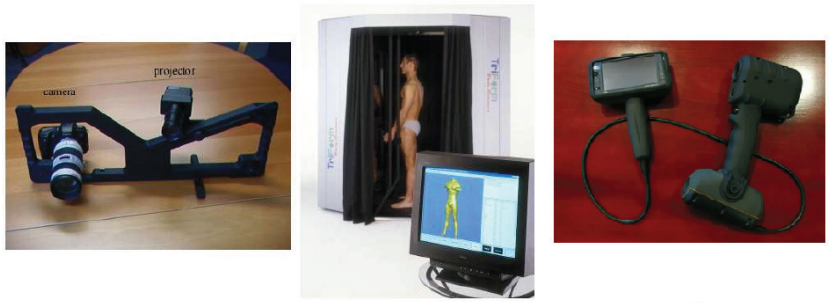


FIGURE 2.13 – a : projection d'un patron de franges sur un visage. b : le scanner canadien InSpeck. c : Le scanner FACESCAN^{3D} de la compagnie 3D-Shape. d et e : Numérisation d'un visage par le système FACESCAN^{3D}. f : Le scanner Mephisto EX de la compagnie 4DDynamics. g et h : acquisitions 3D calculées par le scanner Mephisto EX de la compagnie 4DDynamics.

L'avantage majeur de cette famille de systèmes est leur faible coût par comparaison aux systèmes de balayage laser. Une simple solution Shapesnatcher proposée par la compagnie Eyetronics (www.eyetronics.com) formée d'un couple projecteur + caméra s'affiche sur la figure 2.14.a. Une autre solution à faible coût est le scanner TriForm de corps humains complet. Il est développé par Wicks and Wilson (www.wwl.co.uk) et apparait sur la figure 2.14.b. Le système F5 qui apparait sur la figure 2.14.c est proposé par la compagnie Mantis Vision (www.mantis-vision.com). Il est formé par un projecteur et une caméra infrarouges. Ce système est léger et permet de reconstruire en temps réel une scène en projetant un seul patron infrarouge. Le système actuel ne permet pas de retrouver la texture de la scène.

En médecine, la compagnie 3dMD (www.3dmd.com) est l'un des pionniers des tech-



(a) (b) (c)

FIGURE 2.14 – Systèmes de numérisation par projection de lumière. a : Shapesnatcher de Eyetronics (caméra + projecteur). b : Scanner TriForm de corps humains de Wicks and Wilson. c : Système F5 proposé par Mantis Vision

nologies d'imagerie 3D et propose depuis 1997 ses produits au domaine de la santé. Leur système de numérisation de visage 3dMDface apparait sur la figure 2.15. Il utilise une approche de stéréphotogrammétrie avec quatre caméras assistées par une lumière structurée non invasive.

FIGURE 2.15 – Le système 3dMDface de la compagnie 3dMD.

Le système Kinect (www.xbox.com) allie une caméra couleur, une caméra infrarouge et un projecteur infrarouge et des petits moteurs. La caméra couleur permet de capturer les mouvements du joueur. Le projecteur et la caméra infrarouge sondent la profondeur en criblant l'espace d'une grille de points. L'enregistrement des déformations de cette grille permet au système de déterminer la distance à laquelle se trouvent les objets de la scène. La numérisation d'un visage par exemple par un système Kinect fournit un modèle de faible

qualité. Une étape de morphing du modèle obtenu sur un visage 3D de vérité terrain permet de le corriger et d'obtenir un résultat 3D plus réaliste. Ainsi, le système Kinect peut servir dans le domaine d'animation faciale et de suivi de mouvement mais il ne fournit pas des mesures suffisamment précises qui peuvent servir dans le domaine médical par exemple.

FIGURE 2.16 – Le système Kinect.

2.7.3 Reconstruction passive

La dernière famille de systèmes disponibles sur le marché se base sur l'extraction de l'information 3D sans projeter une lumière particulière. Le système Di3D de la compagnie Dimensional Imaging (www.di3d.com) utilise une approche de stéréophotogrammétrie. Il permet de numériser une vidéo 3D d'une personne avec une capture fidèle de ses émotions. Une localisation manuelle d'un ensemble de points saillants est nécessaire uniquement pour créer la première trame 3D et le suivi des points saillants se fait ensuite automatiquement le long des trames 3D suivantes de la séquence vidéo 3D. Une très haute qualité de numérisation 3D peut être obtenue en utilisant 32 caméras comme le montre la figure 2.17.

Le système de modélisation de visage FaceGen de la compagnie Singular Inversions (www.facegen.com) est une solution logicielle dédiée à l'animation faciale. La figure 2.18 montre l'interface logicielle FaceGen ainsi que quelques modèles 3D générés. Le système FaceGen montre la possibilité de générer des visages 3D en utilisant seulement deux images 2D de la personne dont une image frontale et une image de profil ou même en utilisant une seule image frontale. Le modèle 3D est obtenu par une approche semi-automatique basée sur une modélisation statistique de la forme et de l'apparence du visage. Ce système

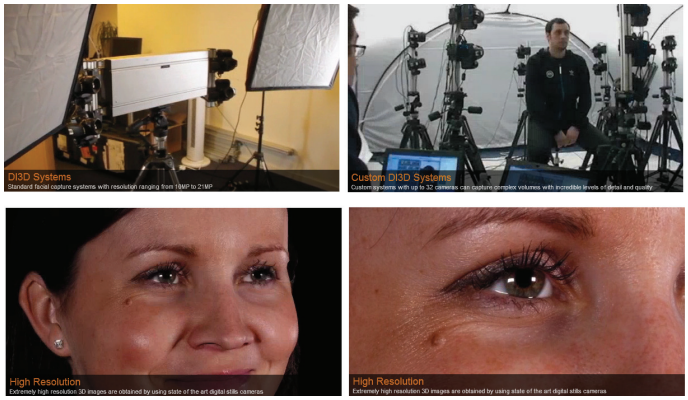


FIGURE 2.17 – Le système de stéréophotogrammétrie Di3D de la compagnie Dimensional Imaging

ne permet pas une mesure réelle de la forme du visage mais il est extrêmement photo-réaliste et adéquat pour des applications de jeux vidéo ou d'animation faciale. La figure 2.18 illustre l'aptitude FaceGen à déformer un visage donné en lui affectant un nouveau genre, une nouvelle couleur de peau ou un maquillage, plus d'âge, une expression réaliste et même une nouvelle coupe de cheveux. Un avantage de cette technique est son très faible coût par comparaison aux systèmes de numérisation de la forme réelle du visage.

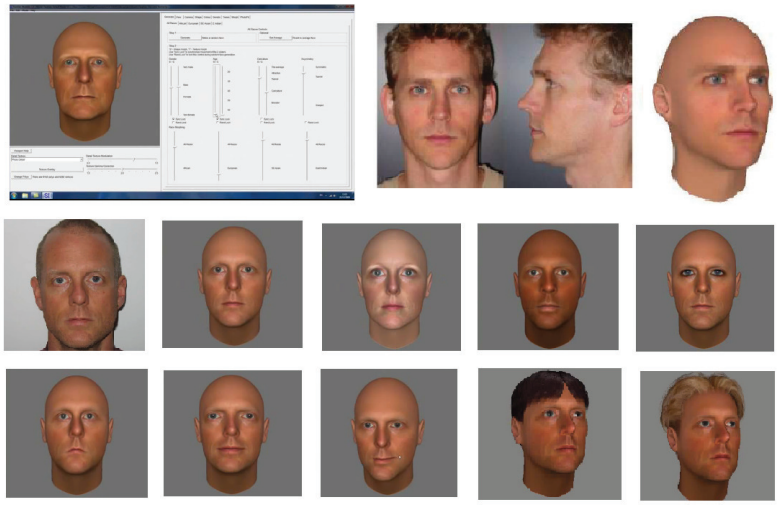


FIGURE 2.18 – Le système de modélisation de visage Facegen de la compagnie Singular Inversions

2.7.4 Classification

La table 2.3 présente une synthèse des systèmes de numérisation 3D disponibles sur le marché. Les critères considérés sont : le principe de la technique, la résolution XY, la précision, la vitesse, la portabilité du système, s'il permet une numérisation 3D du mouvement d'un objet ou pas et si la lumière utilisée est infrarouge ou non.

Compagnie	Principe	Rés. XY	Précision	Vitesse	Portabilité	Mouvement	Infrarouge
Canesta	Tps de vol	320x200	<1cm	30fps		X	X
Swiss Ranger	Tps de vol	176 x 144	<15mm	50fps	X	X	X
Hamamatsu	Laser	1Mpts	5mm	6sec			
Human Solutions	Laser	30pts/cm^2	<1mm	10 sec/scan			
Cyberware	Laser	>500x500	0.01mm	>10sec/scan			
Minolta Range7	Laser	1280x1024	4 μm	2sec/scan	X		
Di3D	Passive stereo	>20Mpts	0.5mm	>25fps		X	
3dMD	Active stéréo	>10Mpts	<=0.2mm	1.5 msec/scan			
XYZRGB	Stereo	>18Mpts	0.01mm	30fps	X		
Kinect	Lum. struct	640x480	<2cm	30fps	X	X	X
Mantis Vision	Lum. struct	50Mpts/frame	0.5mm	20fps	X	X	X
Breuckmann	Lum. struct	2 Mpts	0.2mm	<0.8sec/scan			
4DDynamics	Lum. struct	>10Mpts	<0.15mm	0.5 à 1 sec/scan			
FACESCAN^{3D}	Lum. struct	-	0.1mm	0.8sec			

TABLE 2.3 – Tableau comparatif de quelques systèmes de numérisation 3D disponibles sur le marché.

2.8 Conclusion

Dans ce chapitre, nous avons suggéré une étude des différentes approches de numérisation 3D. Une attention particulière a été accordée aux défis majeurs qui caractérisent le domaine de la numérisation de surfaces 3D animées. La numérisation des surfaces mal-texturées constitue un premier défi. Pour surmonter ce défi, les solutions développées se partagent entre la projection de patrons texturants ou d'une lumière blanche dont la source est préalablement étalonnée [Klaudiny *et al.* 2010]. Un deuxième défi consiste dans le prin-

cipe de la reconstruction lui même qui peut être incapable de fournir un résultat non-bruité ou de haute résolution comme la technique TOF :Time-Of-Flight [Schuon *et al.* 2009] d'où le recourt à la fusion d'approches et aussi à l'utilisation de l'axe spatio-temporel ou du principe de la super-résolution. Un dernier défi que nous pouvons signaler est la variation non-rigide de la forme à travers le temps qui doit être prise en compte au moment de l'acquisition pour garantir une numérisation 3D efficace et précise.

Dans ce travail de thèse, nous proposons une solution active de numérisation 3D de visages fixes ou animés. Nous introduisons tout d'abord une approche hybride qui profite des atouts de la lumière structurée par décalage de phase (Phase-Shifting) et de la stéréovision pour récupérer toute l'information 3D que présentent les deux vues stéréoscopiques. Notre procédé de numérisation ne nécessite pas l'étape hors-ligne d'étalonnage projecteur-caméra qui constitue une tâche fastidieuse exigée par les techniques de la lumière structurée par décalage de phase. Avoir une résolution pixélique n'est pas toujours suffisant pour retrouver tous les détails de la forme 3D comme pour récupérer des rides sur un visage par exemple. Nous proposons donc d'assurer une super résolution spatio-temporelle non-rigide pour considérer une éventuelle variation de l'expression faciale entres les trames 3D successives.

CHAPITRE 3

Stéréovision Active

3.1 Introduction

L'utilisation de la biométrie faciale 3D, dans un scénario de contrôle d'accès par exemple, est privilégiée en raison de son caractère non intrusif et le faible degré de coopération de la part de l'individu. Néanmoins, le visage se caractérise par une texture plutôt lisse. Ceci rend difficile l'appariement stéréo à des échelles de résolution faible ou intermédiaire. Ainsi, pour assurer une numérisation 3D efficace du visage, nous optons pour une méthode basée sur la vision active. Dans cette catégorie, une numérisation 3D de visage par un scanner laser est à écarter, car couteuse et lente. Quant aux capteurs à temps de vol, ils produisent des images de profondeur imprécises et de faible résolution. Dans ce chapitre, nous proposons une approche de numérisation 3D à faible coût qui utilise deux caméras de surveillance réseau. Nous assistons la capture de visages par une lumière structurée pour compenser l'insuffisance de la texture faciale. Nous mesurons l'information 3D par stéréovision.

3.2 Principe de l'approche

Le système de numérisation emploie deux caméras étalonnées et un vidéoprojecteur non-étalonné. Notre système de numérisation est constitué d'une paire de caméras $AXIS$ et un dispositif de projection de lumière structurée. En hors-ligne, un étalonnage des deux caméras est assuré en utilisant une mire commune. Au moment de la numérisation, nous projetons successivement un patron binaire de franges noires et blanches alternées et son patron complémentaire pour pallier la platitude de la texture du visage. En ligne, nous capturons trois images de chaque caméra. Les deux premières images comportent la distorsion

des deux patrons binaires complémentaires sur le visage. La troisième image permet de récupérer la texture du visage. La figure 3.1 définit les différentes étapes nécessaires pour la numérisation 3D d'un visage par notre approche de stéréovision active.

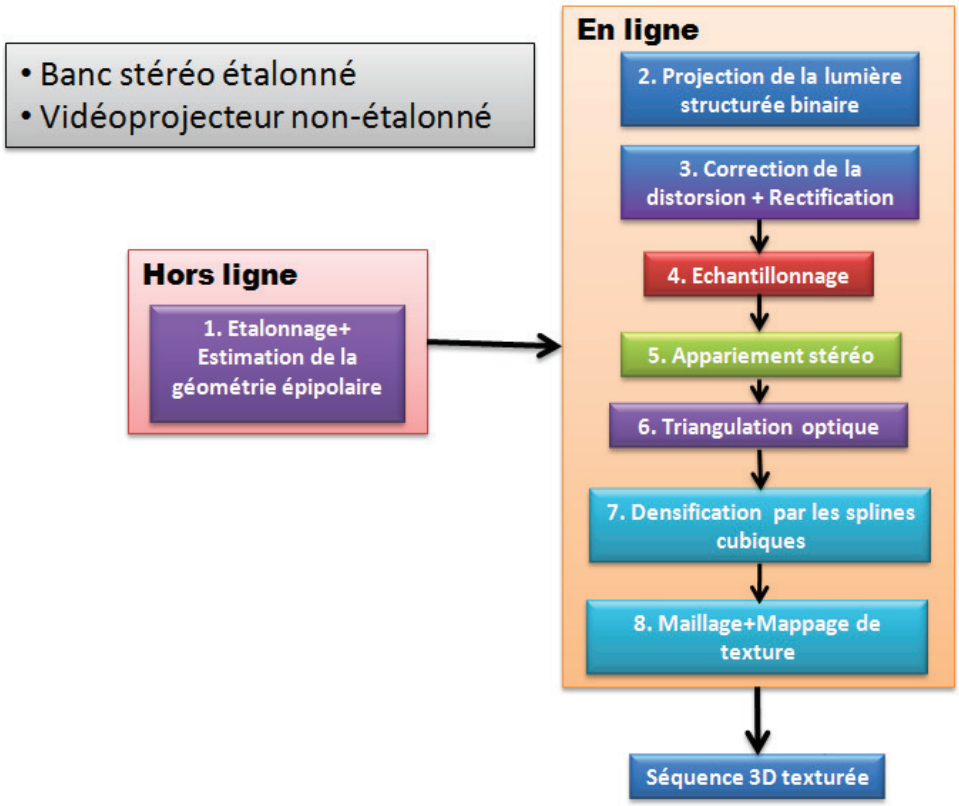


FIGURE 3.1 – Principe de l'approche de stéréovision active proposée.

Les images subissent une correction de la distorsion radiale et tangentielle ainsi qu'une rectification. Ensuite, une localisation des points d'intersection de franges permet d'échantillonner les deux vues gauche et droite du visage. Les points échantillonnés constituent les primitives gauches et droites que nous utilisons pour la reconstruction 3D du visage. Les primitives sont d'abord mises en correspondance par un appariement stéréo. Ensuite, à partir de chaque couple de primitives gauche/droite appariées, nous calculons les coordonnées 3D d'un point du visage par une triangulation optique. Une interpolation par les splines cubiques permet de densifier le nuage de points 3D obtenu. Finalement, un maillage suivi d'un plaquage de texture fournissent un modèle 3D texturé du visage. Le plaquage de la texture sur le modèle facial 3D est assuré en utilisant la correspondance entre la texture

2D d'un pixel et de ses coordonnées 3D calculés. La figure 3.2 illustre le scénario de la numérisation 3D d'un visage.

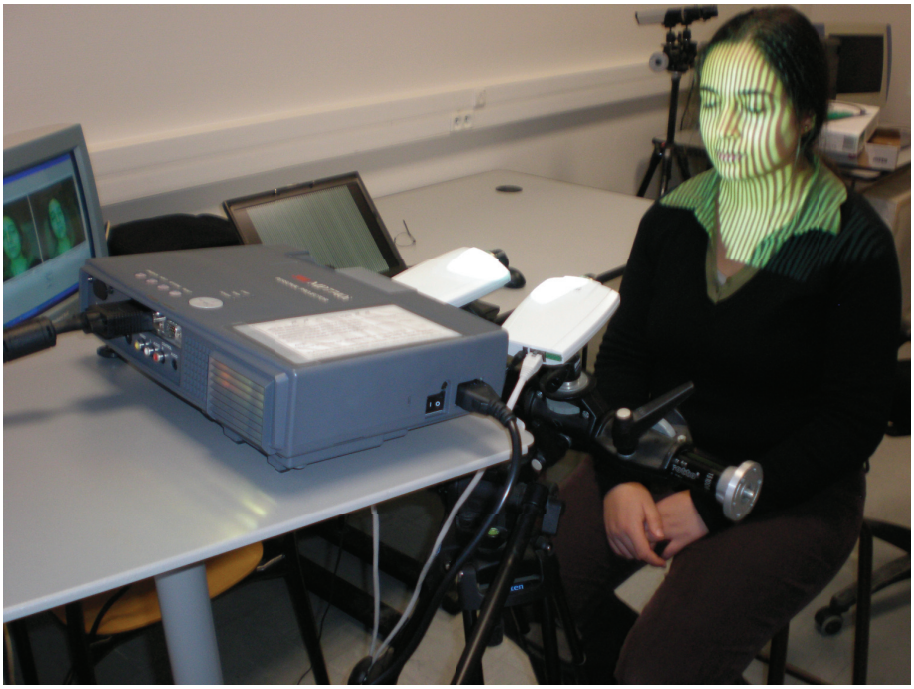


FIGURE 3.2 – Notre système de numérisation 3D.

3.3 Etalonnage Stéréo

La reconstruction 3D d'un objet ou d'une scène à partir de deux vues stéréo nécessite une connaissance de la transformation géométrique entre les deux plans images des deux caméras dans l'espace. Ceci revient à un étalonnage stéréo qui consiste à estimer en une première étape la position de chaque caméra dans l'espace en calculant leurs paramètres intrinsèques et extrinsèques. La deuxième étape consiste à l'estimation de la géométrie épipolaire qui relie les deux plans images.

3.3.1 Paramètres intrinsèques et extrinsèques d'une caméra

La vue d'une scène sur une image acquise par une caméra est obtenue par une projection perspective des points 3D formant la scène réelle sur le plan image de la caméra. Nous adoptons le modèle de sténopé, ou modèle pin-hole dans la littérature anglo-saxonne,

qui constitue une modélisation simple et linéaire du processus de formation des images au sein d'une caméra. Ce modèle est couramment utilisé en vision par ordinateur. Il suppose que la lentille de la caméra respecte les conditions de Gauss c.-à-d. le faisceau doit traverser la lentille au voisinage du centre optique et les rayons incidents doivent faire un angle faible avec l'axe optique de la lentille. Il suffit d'exprimer les relations de passage du repère monde au repère caméra, d'exprimer la projection du repère caméra dans le plan image et d'appliquer la transformation affine qui conduit aux coordonnées de l'image.

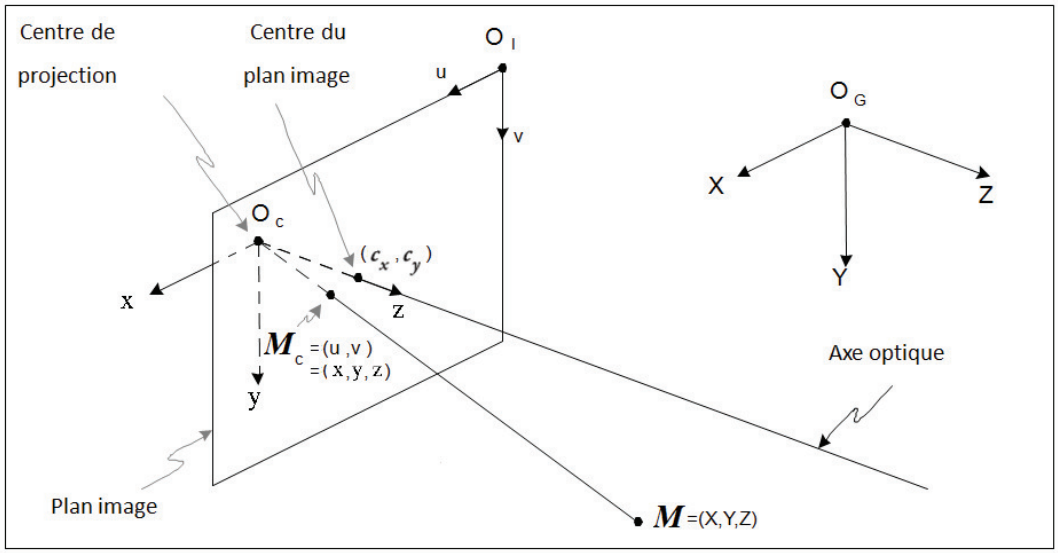


FIGURE 3.3 – Les repères géométriques associés à l'étalonnage d'une caméra.

La figure 3.3 présente les repères géométriques dans lesquels nous allons travailler pour modéliser le fonctionnement de la caméra. Le repère $R_G(O_G, \vec{X}, \vec{Y}, \vec{Z})$ constitue le repère monde,O_G étant son origine. Le repère $R_C(O_C, \vec{x}, \vec{y}, \vec{z})$ est associé à la caméra, C étant le centre optique de la caméra ie. son centre de projection. Le repère $R_I(O_I, \vec{u}, \vec{v})$ est associé au plan image, O_I étant son origine. Soit M un point de coordonnées (X, Y, Z) dans le repère monde. Son acquisition par une caméra constitue un point 2D M_C de coordonnées (u, v) dans le plan image. Aussi, le point M_C a les coordonnées (x, y, z) dans le repère associé à la caméra. La projection du point M sur le plan image se traduit par l'équation (3.1).

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = P_{cam} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, P_{cam} = A[R|T] = \begin{bmatrix} f_x & S_{xy} & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.1}$$

La matrice P_{cam} constitue la matrice de projection du point 3D M de la scène vers le point 2D M_C de coordonnées (u, v) sur le plan image. Cette équation utilise les coordonnées homogènes $(X, Y, Z, 1)$ de M et $(u, v, 1)$ de M_C. Dans cette équation, s est un scalaire choisi arbitrairement. A constitue la matrice intrinsèque qui rassemble les propriétés optiques et géométriques de la caméra. Ces paramètres intrinsèques sont f_x, f_y, c_x, c_y et S_{xy}. f_x et f_y correspondent à la distance focale exprimée en largeurs et en hauteurs de pixels. c_x et c_y définissent les coordonnées du point principal défini par la projection du centre optique de la caméra sur le plan image. Ainsi, il constitue le centre du plan image. La quantité S_{xy} traduit la non-orthogonalité potentielle des lignes et des colonnes de cellules électroniques photosensibles qui composent le capteur de la caméra. La plupart du temps, ce paramètre est négligé et prend donc une valeur nulle. Les paramètres caractérisant la position et l'orientation de la caméra constituent les paramètres extrinsèques de la caméra. Ces derniers forment une matrice qui caractérise une rotation R et un vecteur de translation T permettant de passer du repère monde vers le repère lié à la caméra.

Les paramètres f_x et f_y vérifient $f_x = k_x.f$ et $f_y = k_y.f$ avec f la distance focale séparant le plan image du centre optique. Les termes k_x et k_y définissent les facteurs d'agrandissement de l'image. Ainsi la matrice intrinsèque A peut s'écrire par l'équation (3.2).

$$A = \begin{bmatrix} f_x & S_{xy} & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} k_x & s_{xy} & c_x \\ 0 & k_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.2}$$

La transformation entre le repère monde et le repère associé à la caméra se définit par la relation entre les coordonnées du point $M(X, Y, Z)$ dans le repère monde et les coordonnées de $M_C(x, y, z)$ dans le repère associé à la caméra. L'équation (3.3) définit

cette transformation (dans le cas où $z \neq 0$).

$$\begin{aligned}
&\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} + T \\
&x' = x/z \\
&y' = y/z \\
&u = f_x * x' + c_x \\
&v = f_y * y' + c_y
\end{aligned} \tag{3.3}$$

3.3.2 Distorsion radiale et tangentielle

La distorsion optique d'une caméra constitue aussi une propriété intrinsèque de la caméra. Elle représente la déformation de l'image provoquée par la lentille. En optique la distorsion apparait quand les conditions de Gauss ne sont plus respectées. Les lentilles, par leur symétrie sphérique, ont une certaine distorsion essentiellement radiale et une légère distorsion tangentielle. La distorsion radiale a tendance à arrondir les bords de l'image. La distorsion tangentielle se caractérise par une rotation autour du centre de l'image. L'amplitude de cette rotation varie selon la position des points sur l'image. La figure 3.4 illustre les deux phénomènes de la distorsion tangentielle et la distorsion radiale.

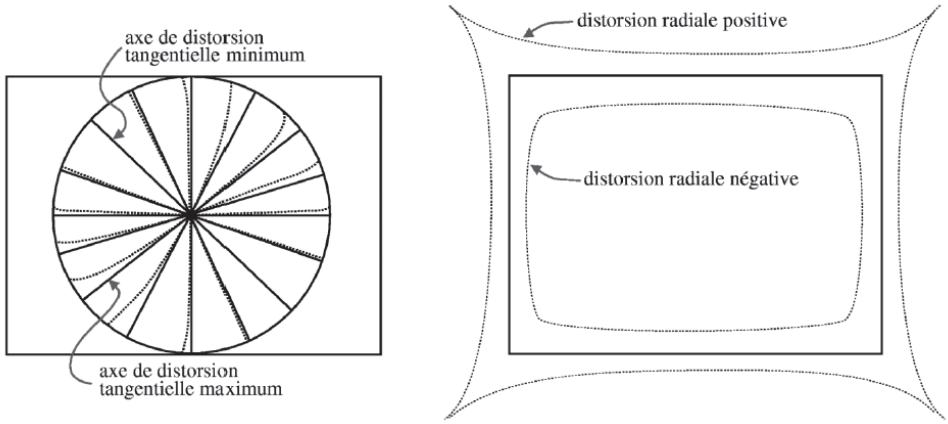


FIGURE 3.4 – Distorsion tangentielle et distorsion radiale.

Il est possible de corriger l'image capturée si l'on possède les coefficients de la distor-

sion de la caméra. L'image peut ainsi être corrigée par une interpolation des pixels ayant subi préalablement un déplacement inverse à celui de la distorsion. La distorsion est définie par cinq coefficients. k_1, k_2 et k_3 sont les coefficients de la distorsion radiale et p_1 et p_2 constituent les coefficients de la distorsion tangentielle. Ainsi, le modèle ci-dessus devient :

$$\begin{aligned}
&\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} + T \\
&x' = x/z \\
&y' = y/z \\
&x'' = x'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x' y' + p_2(r^2 + 2x'^2) \\
&y'' = y'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y'^2) + 2p_2 x' y' \\
&avec \quad r^2 = x'^2 + y'^2 \\
&u = f_x * x'' + c_x \\
&v = f_y * y'' + c_y
\end{aligned} \tag{3.4}$$

3.3.3 Géométrie épipolaire

La géométrie épipolaire se définit par une étude des deux transformations projectives qui caractérisent les deux plans images de deux caméras dans le repère monde. Considérons deux caméras ayant deux centres optiques O_l et O_r comme le décrit la figure 3.5. Un point M de $\mathbb{R}^3$ est transféré sur deux plans projectifs Π_l et Π_r qui représentent respectivement les deux plans images des deux caméras gauche et droite. Ceci donne naissance à deux points M_l et M_r. Connaissant les matrices des projections P_{caml} et P_{camr}, on peut calculer les coordonnées des points M_l et M_r à partir de celles de M. Connaissant la projection de M sur l'un des plans, nous pouvons tirer quelques déductions qui permettent de décrire d'avantage la géométrie épipolaire. le point M_l peut être la projection de n'importe quel point de l'espace situé sur la droite $O_l M_l$. la projection de cette droite $O_l M_l$ dans le plan projectif Π_r est également une droite appelée droite épipolaire de Π_r associée à M_l.

Le centre de projection O_l appartient à toutes les droites joignant un point M de l'espace à sa projection M_l dans Π_l. la projection e_r de O_l dans Π_r appartient donc à toutes les

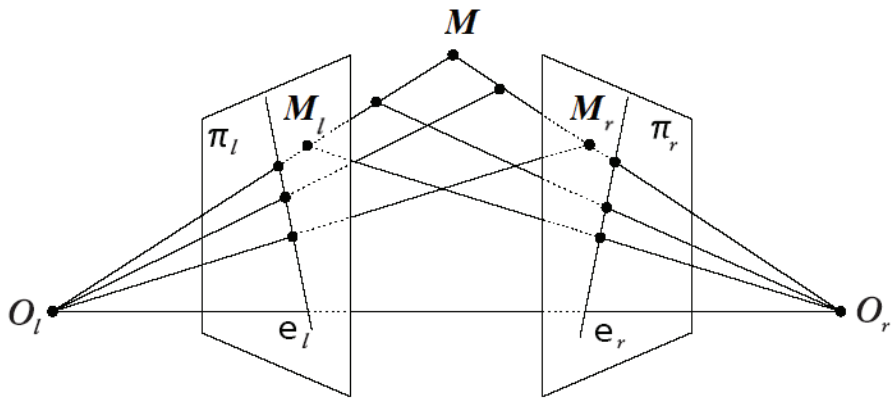


FIGURE 3.5 – La configuration de départ des deux caméras à étalonner.

droites épipolaires d'un point de Π_l. Les points e_l et e_r sont appelés épipoles ou points épipolaires. La droite joignant les centres optiques (donc passant par les épipoles) est appelée droite de base. Le plan épipolaire Π_M associé à M est le plan de l'espace projectif défini par M et les deux centres de projection. Les deux droites épipolaires associées aux deux projections de M sont les intersections de Π_M avec les plans de projection. Le faisceau épipolaire est l'ensemble de tous les plans épipolaires. Deux points image M_l et M_r sont dits homologues s'ils correspondent aux deux projections d'un même point M de la scène sur les deux images.

Soient (R_l, T_l) la transformation qui relie le repère associé à la caméra gauche et le repère monde et (R_r, T_r) la transformation qui relie le repère associé à la caméra droite au repère monde. Soient (R, T) la transformation entre les deux repères des caméras gauche et droite. Ainsi, la transformation (R, T) peut être estimée moyennant l'équation (3.5).

$$R_r = R * R_l T_r = R * T_l + T, \tag{3.5}$$

A partir du modèle de caméra élaboré précédemment, il est possible de dériver une relation linéaire simple pour la mise en correspondance de points de deux images. On désigne ici cette relation par les termes matrice fondamentale et aussi matrice essentielle. Les deux matrices essentielle et fondamentale permettent de définir pour chaque point de la première image la droite épipolaire correspondante sur la deuxième image sur laquelle se trouve le point en correspondance. La matrice essentielle E exprime dans le repère caméra

la relation entre deux points droite et gauche homologues dans le cas de deux caméras préalablement calibrées. Elle est éventuellement calculée en utilisant l'équation (3.6), T_i constituent les composantes de la translation $T = [T_x\ T_y\ T_z]^T$.

$$E = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} * R \tag{3.6}$$

Lorsque les paramètres intrinsèques et extrinsèques de chaque caméra sont inconnues, nous n'avons pas la relation entre le repère caméra et le repère image. Il faut calculer la matrice fondamentale F qui détermine, dans le repère image, la relation entre chaque point de la caméra gauche et la ligne épipolaire qui passe par son homologue de la caméra droite. La matrice fondamentale F est définie par l'équation (3.7). A_l et A_r sont les deux matrices intrinsèques des deux caméras gauche et droite.

$$F = A_r^T E A_l^{-1} \tag{3.7}$$

3.3.4 Approche d'étalonnage

Il y a deux méthodes qui permettent d'étalonner les caméras et calculer la géométrie épipolaire. La première, dite d'étalonnage fort, consiste à utiliser une mire commune pour l'étalonnage des deux caméras. La mire est un objet de géométrie connue capturée par les deux caméras à plusieurs reprises, en variant à chaque fois sa position, à fin de les étalonner. Les mires les plus utilisées sont sous forme d'un échiquier ou d'une sphère. La mire commune permet surtout de définir la géométrie épipolaire en calculant la matrice essentielle ainsi que les deux transformations perspectives à appliquer à chacun des deux plans images. Cette approche s'effectue hors-ligne et permet une estimation précise des paramètres. Elle est facile à mettre en œuvre. Cependant, les deux caméras doivent rester figées au cours de la numérisation.

La deuxième méthode, dite d'auto-étalonnage, consiste à estimer uniquement la géométrie épipolaire en calculant la matrice fondamentale à partir d'un certain nombre d'appariements gauche/droite au moment de la capture de la scène. Elle permet ainsi la mobilité des deux caméras au cours de la numérisation. Cependant, cette approche ne permet

pas une estimation des paramètres intrinsèques et extrinsèques des deux caméras et elle manque de précision. Nous retenons la première approche et nous utilisons un échiquier comme une mire d'étalonnage commune, dont nous extrayons automatiquement les coins, comme proposé dans [Zhang 1999]. Un nombre défini de différentes positions de la mire sont capturées par le système de numérisation comme le décrit la figure 3.6. Chaque coin de l'échiquier est défini par ses coordonnées dans les vues gauches capturées ainsi que les coordonnées de son point homologue dans les vues droites correspondantes. Ceci permet de déterminer les paramètres intrinsèques et extrinsèques du système en estimant l'homographie plane, i.e. la transformation projective, entre le plan de l'échiquier et le plan image de chaque caméra.

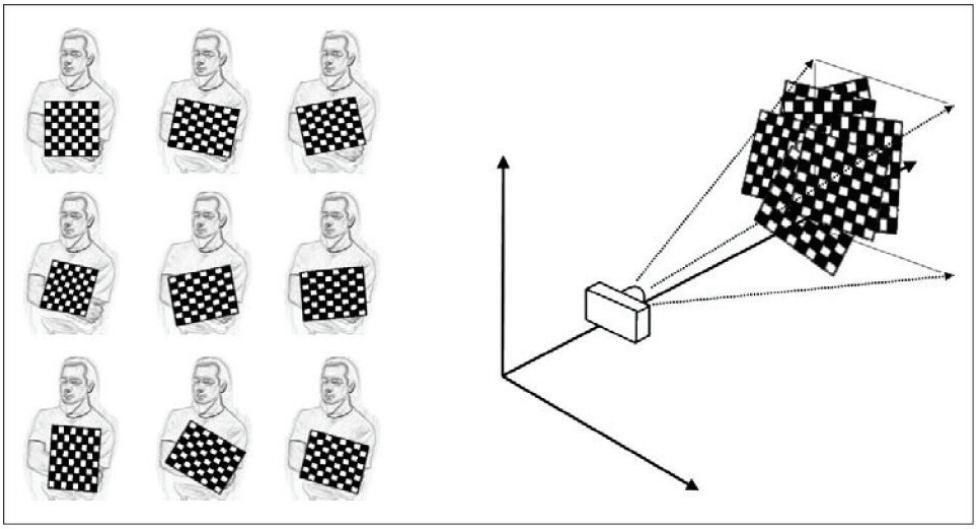


FIGURE 3.6 – Capture de plusieurs positions d'un échiquier.

3.3.5 Rectification

Après l'étalonnage de chaque caméra et l'estimation de la géométrie épipolaire, la rectification permet d'appliquer les deux transformations nécessaires sur les deux plans images gauche et droite comme le décrit la figure 3.7. Ceci permet d'obtenir une nouvelle configuration virtuelle du capteur stéréo dans laquelle les deux épipoles se ramènent vers l'infini. Ainsi, les axes optiques des deux caméras sont parallèles et les lignes épipolaires conjuguées aussi.

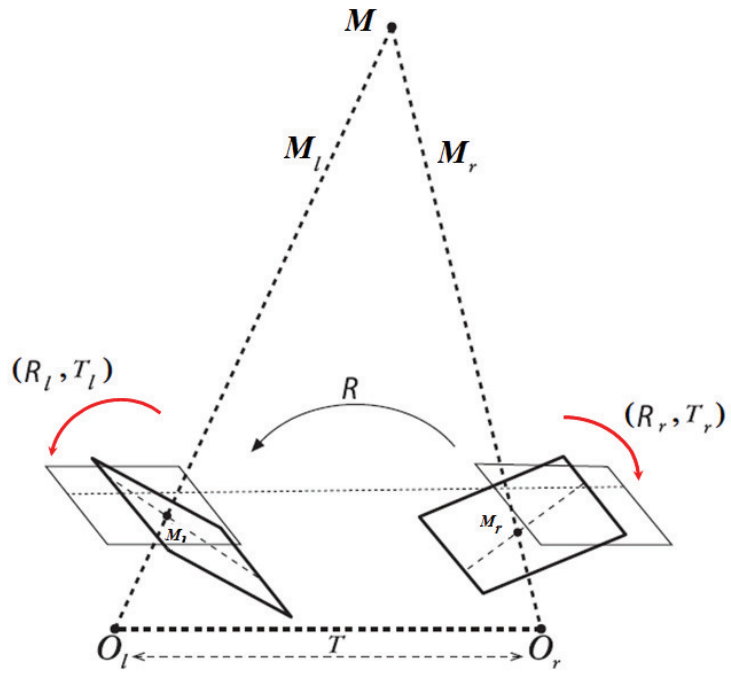


FIGURE 3.7 – Les deux transformations nécessaires pour la rectification du système stéréo.

La figure 3.8 présente la nouvelle configuration de notre système stéréo après la rectification et la figure 3.9 illustre le résultat de la rectification appliquée sur deux vues gauche et droite d'un échiquier ainsi que l'image de la disparité générée. La rectification réduit la transformation entre les deux caméras gauche et droite à une translation T_x sur l'axe X.

Les deux matrices intrinsèques droite et gauche sont mises à jour avec une nouvelle distance focale commune f aux deux caméras gauche et droite et deux nouveaux points principaux gauche et droite appelés aussi centres de projection. Le point principal gauche et le point principal droite ont les mêmes nouveaux coordonnées (cx, cy) respectivement sur les deux nouveaux plans image gauche et droite. Les deux nouvelles matrices de projection gauche et droite P_{caml} et P_{camr} sont ainsi définies par l'équation (3.8).

$$P_{caml} = \begin{bmatrix} f & 0 & cx & 0 \\ 0 & f & cy & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, P_{camr} = \begin{bmatrix} f & 0 & cx & T_x * f \\ 0 & f & cy & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{3.8}$$

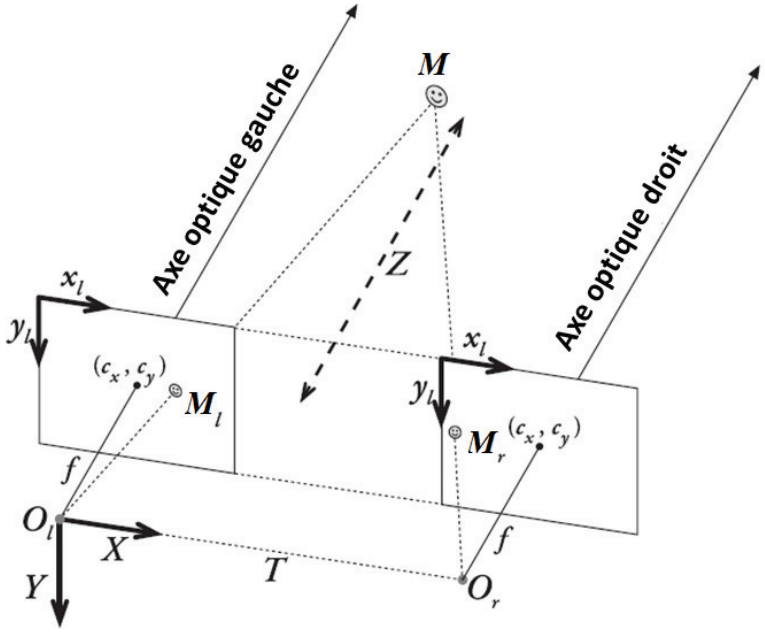


FIGURE 3.8 – La configuration de notre système stéréo après la rectification.

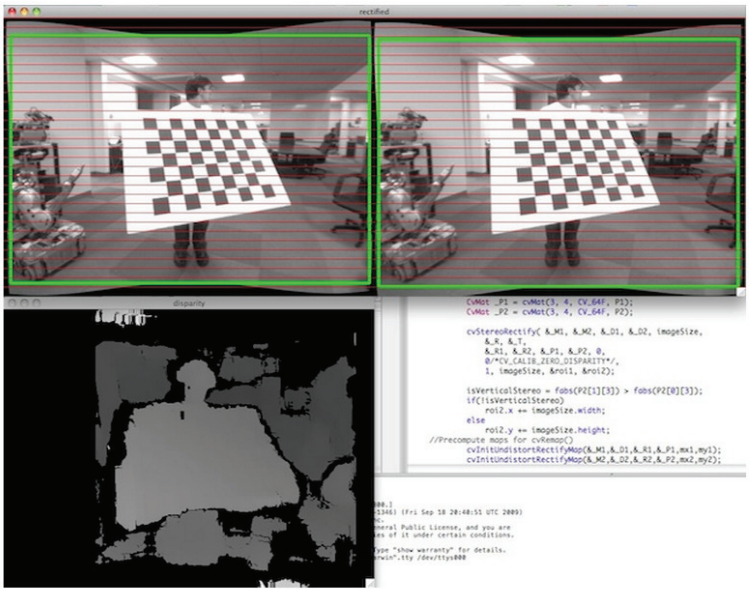


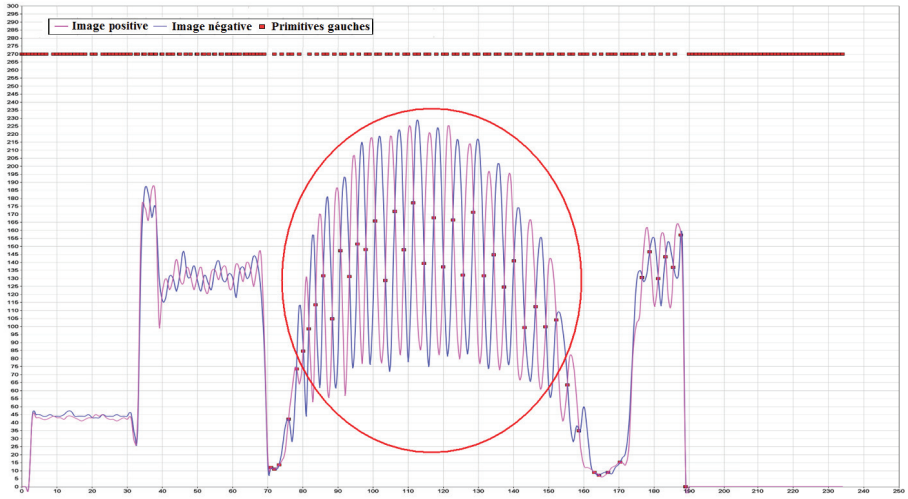
FIGURE 3.9 – Le résultat de la rectification appliquée sur deux vues gauche et droite d'un échiquier ainsi que l'image de la disparité générée.

3.4 Echantillonnage

Pour chaque caméra, la projection des deux patrons de lumières inverses dans un laps de temps réduit génère deux images identiques du visage mais soumis à deux dispositions complémentaires des franges. Ainsi, deux courbes sinusoïdales sont générées sur les profils d'intensité des deux lignes épipolaires homologues correspondantes respectivement aux deux images complémentaires, comme l'illustre la figure 3.10.

(a) Image positive

(b) Image négative



(c) Echantillonnage sous pixélique sur une ligne épipolaire.

FIGURE 3.10 – Principe de l'échantillonnage.

Chaque courbe sinusoïdale est calculée en utilisant la méthode d'interpolation par les splines cubiques. En effet, les splines cubiques sont des courbes d'interpolation lisses et continues définies par des tranches locales de polynômes de troisième ordre continues et de dérivées continues sur les extrémités des intervalles. Les points d'appuis de l'interpolation

sont les pixels de l'image extraites de chaque ligne épipolaire. Le processus d'échantillonnage consiste à localiser pour chaque ligne épipolaire les points d'intersection des deux courbes sinusoïdales.

Puisque les deux courbes sinusoïdales sont en opposition de phase, les points sont localisés sur les contours des franges. Ceci permet d'échantillonner les images de manière dense suivant l'axe vertical Y et semi dense suivant l'axe horizontal X. Les points échantillonnés gauche et droits constituent les primitives gauches et droites que nous utilisons pour la reconstruction 3D du visage. Dans les images gauche et droite, les primitives localisées ont des ordonnées entières définies par les lignes épipolaires et des abscisses réelles issues de l'intersection entre les deux courbes sinusoïdales. Ainsi, cette approche permet une localisation sous-pixélique des primitives gauches et droites.

3.5 Appariement stéréo

Le problème d'appariement stéréo constitue le cœur de la numérisation 3D par stéréovision. En effet, pour identifier les coordonnées 3D d'un point M capturé par deux caméras gauche et droite, nous devons tout d'abord identifier ses deux projections M_l et M_r sur les deux images gauche et droite capturées. M_l et M_r sont appelés des points homologues. Ils permettent d'obtenir les coordonnées 3D en question par une triangulation optique. Ainsi, pour reconstruire un visage 3D, il suffit de trouver pour chaque point du visage sur l'image gauche son point homologue sur l'image droite. L'appariement stéréo consiste à définir tous les couples de points homologues gauche/droite.

3.5.1 Les contraintes

Le processus d'appariement doit respecter deux types de contraintes pour restreindre l'espace de la recherche de correspondance gauche/droite et pour garantir une convergence rapide vers une correspondance gauche/droite fiable. Il s'agit d'une part des contraintes de régularité et d'autre part des contraintes imposées par la nature même de la lumière structurée projetée lors de la numérisation. Les contraintes de régularité sont la contrainte épipolaire, l'occultation, l'unicité, la continuité et l'ordre. Deux contraintes sont imposées par la lumière structurée. Elles sont la contrainte de continuité verticale de la frange et

l'alternance entre les franges noires et blanches sur l'axe horizontal. Les contraintes sont définies comme suit :

- **La contrainte épipolaire** exige que les points gauche et droite correspondants doivent se trouver sur la même ligne épipolaire. La contrainte épipolaire permet de réduire la complexité du problème d'appariement en ramenant un problème de recherche bidimensionnel à un problème de recherche unidimensionnel. En effet, après la rectification, pour chaque point gauche, son point homologue droit est nécessairement sur la même ligne épipolaire.
- **La contrainte d'occultation** impose que la discontinuité perçue sur la première vue stéréo correspond à une occultation dans la seconde vue stéréo et vice versa.
- **La contrainte d'ordre** requiert un même ordre des primitives et de leurs correspondants sur la même ligne épipolaire.
- **La contrainte de continuité** nécessite que la profondeur varie peu sur une surface lisse et régulière. Ainsi, deux points physiquement voisins sur le visage ont nécessairement des valeurs de disparités correspondantes proches. Cette contrainte s'appelle aussi la contrainte de la différence de la disparité.
- **La contrainte d'unicité** impose que pour chaque point dans l'image gauche, existe un seul correspondant dans l'image droite.
- **La continuité verticale de la frange** constitue une contrainte interligne. Elle exige que deux primitives gauche et droite ne peuvent être homologues que si les deux primitives de la ligne précédente et qui se situent sur le bord de la même frange soient homologues aussi.
- **L'alternance horizontale entre les franges noires et blanches** nécessite que deux primitives gauche et droite ne peuvent être homologues que si toutes les deux sont situées sur un même type de passage de franges blanc-noir ou noir-blanc puisque les franges sont alternées en noir et blanc. Ainsi, toutes les primitives sont situées sur les bords des franges soit sur un bord de passage de noir en blanc ou de blanc en noir.

3.5.2 Modélisation du problème

Nous modélisons le problème d'appariement comme un problème de recherche du plus court chemin dans un graphe formé par deux séquences de primitives gauches et droites sur les lignes épipolaires homologues. Le problème d'appariement reste encore difficile à résoudre sur les lignes épipolaires homologues. En effet, outre la ressemblance des primitives au sein d'une même image, des primitives observées dans une image peuvent être occultées dans la seconde. De plus, la détection des primitives n'est pas parfaite. En effet, de fausses primitives peuvent être détectées au moment de l'échantillonnage et de vraies primitives peuvent être ignorées.

Nous proposons un algorithme basé sur le principe de la programmation dynamique pour la résolution de ce problème. Nous justifions notre choix par la robustesse de la programmation dynamique dans la localisation des occultations. Aussi, son implémentation respecte les contraintes d'ordre, d'unicité et de la continuité ce qui renforce la convergence vers une solution fiable. La programmation dynamique est une technique d'optimisation qui cherche à apparier deux ensembles de points en calculant en une première étape une fonction de similarité gauche/droite. Dans une deuxième étape, une fonction de coût cumulatif est définie. Finalement, nous cherchons les couples de primitives gauches/droites homologues qui correspondent à un coût cumulatif minimal.

Nous définissons la fonction de similarité $Similarity(l_i, r_j, n)$ et la fonction de coût cumulatif $\Psi(\phi^*_{i,j}, n)$ pour une ligne épipolaire n comme des matrices. Les lignes et les colonnes des deux matrices sont indexées respectivement par les primitives gauches l_i et droites r_j comme proposé par Ohta et Kanade, [Ohta & Kanade 1985], sur les contours simples. Les occultations sont modélisées par une affectation d'un groupe de pixels dans une vue stéréo à un seul pixel dans l'autre vue stéréo et pénalisant la solution par le coût d'occultation $occlusion$. Soit $\phi^*_{i,j}$ le plus court chemin à trouver qui passe par le nœud (i, j) du couple de primitives gauche l_i et droite r_j. La matrice $\Psi(\phi^*_{i,j}, n)$ est ainsi construite par l'équation (3.9).

$$\Psi(\phi^*_{i,j},n)=\begin{cases} 0, \quad si \quad i \quad = \quad 0 \quad et \quad j \quad = \quad 0 \\ \\ max\begin{cases} \Psi(\phi^*_{i-1,j-1,n})+Similarity(l_i,r_j,n) \\ \Psi(\phi^*_{i-1,j,n})+occlusion \\ \Psi(\phi^*_{i,j-1,n})+occlusion \end{cases} \\ \\ sinon \end{cases} \tag{3.9}$$

Le principe de la recherche consiste à parcourir la matrice de coût cumulatif en se déplaçant verticalement, horizontalement ou sur la diagonale pour extraire la séquence de N couples de primitives gauche/droite ayant la plus faible somme de coûts cumulatifs. Ainsi, la décision de correspondance de la séquence est le résultat de N décisions séparées de l'élection des plus courts chemins partiels. La contrainte d'occultation est prise en compte par le terme $occlusion$. La représentation matricielle de la séquence de primitives gauches et la séquence de primitives droites permet un parcours ordonné sur la matrice de coût. Ainsi, le résultat du parcours est une séquence ordonnée de couples de primitives gauche/droite ce qui répond à la contrainte de l'ordre. En plus, nous analysons automatiquement les couples de primitives gauche/droite homologues trouvés pour que chaque primitive gauche ou droite se présente dans un seul couple de primitives gauche/droite ce qui constitue la contrainte d'unicité. Dans le parcours de la matrice de coût, nous rejetons les cases lointaines de la diagonale de la matrice pour respecter la contrainte de continuité. En effet, les valeurs de disparités augmentent d'une manière exponentielle en s'éloignant des correspondances diagonales.

Nous proposons de construire une mesure de similarité pour mesurer la corrélation entre deux primitives gauche et droite. Elle est définie à fin de renforcer le respect des différentes contraintes et assurer surtout la considération des deux dernières contraintes de la continuité verticale de la frange et de l'alternance horizontale entre les franges noires et blanches.

3.5.3 La mesure de similarité

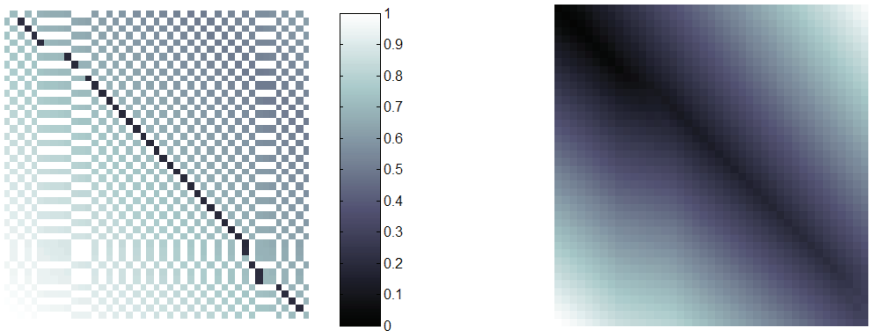
Nous considérons trois critères dans la fonction $Similarity(l_i, r_j, n)$ pour mesurer la corrélation entre deux primitives gauche et droite. Sur les images gauche et droite, les primitives se trouvent soit sur un passage d'une frange noire à une frange blanche ou sur un passage d'une frange blanche à une frange noire. Une primitive gauche l_i ne peut correspondre qu'à une primitive droite r_j que si toutes les deux sont localisées sur un même type de passage de franges. Nous proposons le critère $PatternCode(l_i, r_j, n)$ pour traduire cette contrainte de l'alternance horizontale entre les franges noires et blanches. Il constitue une fonction binaire égale à 0 si les deux primitives l_i et r_j sont localisées sur le même type de passage de franges, il est égal à 1 dans le cas contraire. En utilisant le critère $PatternCode(l_i, r_j, n)$, nous pénalisons le nœud (l_i, r_j) en affectant 1 à $Similarity(l_i, r_j, n)$ si l_i et r_j ne sont pas sur le même type de passage de franges.

$$Similarity(l_i, r_j, n) = \left\{ \begin{array}{l} 1, \ if \ \ PatternCode(l_i, r_j, n) \ = \ 1 \\ \\ \alpha \cdot DispScore(l_i, r_j, n) + \\ \beta \cdot ContinuityConst(l_i, r_j, n) \\ \quad otherwise \end{array} \right\} \tag{3.10}$$

Le terme $DispScore(l_i, r_j, n)$ est une contrainte basée sur la disparité. Il considère que (l_i, r_j) sont homologues et qu'il n'y a pas une occultation majeure à l'intérieur de la région faciale. Ainsi, si (l_i, r_j) sont homologues, la contrainte de la continuité confirme que tous les nœuds qui précèdent le noeud (l_i, r_j) et qui le suivent et qui sont situés sur la première diagonale de la matrice de similarité passante par le nœud (l_i, r_j), correspondent aussi. Le terme $DispScore(l_i, r_j, n)$ est ensuite défini comme la valeur moyenne des termes de disparité séparant tous les nœuds situés sur cette diagonale $Diag$. $LDiag$ est le nombre de nœuds situés sur $Diag$.

$$DispScore(l_i, r_j, n) = \frac{1}{LDiag} \cdot \sum_{(l_s, r_s) \in Diag} (l_s - r_s) \tag{3.11}$$

Le terme $ContinuityConst(l_i, r_j, n)$ traduit la contrainte de la continuité verticale de la frange. Il augmente la chance de correspondance pour un couple donné de primitives gauche et droite(l_i, r_j) sur une ligne épipolaire n si dans la ligne épipolaire $n-1$, la primitive gauche $l'_{i'}$ située sur le même passage de franges que l_i et la primitive droite $r'_{j'}$ située sur le même passage de franges que r_j ont été estimés comme correspondants par le processus de mise en correspondance stéréo calculé sur la ligne épipolaire $i-1$. La fonction $prec(x, n)$ parcourt la ligne épipolaire $n-1$ pour localiser la primitive x' située sur le même passage de franges que la primitive x située sur la ligne épipolaire i. Deux poids α et β sont associés respectivement aux termes $DispScore$ et $ContinuityConst$, $\{\alpha, \beta\} \in [0..1]$. Dans nos expérimentations, nous utilisons $\alpha = \beta = 0.5$. La figure 3.11 présente les fonctions de similarité et de coût cumulatif calculées sur une ligne épipolaire donnée.



(a) Matrice de similarité (b) Matrice de coût cumulatif

FIGURE 3.11 – Matrices de similarité et de coût cumulatif pour une ligne épipolaire donnée.

3.6 Triangulation optique

Après l'appariement stéréo, nous récupérons un ensemble de couples de primitives homologues de type (M_l, M_r), M_l étant une primitive gauche et M_r une primitive droite. Les deux primitives M_l et M_r constituent la projection d'un point 3D M du visage respectivement sur les deux plans images gauche et droite. La figure 3.12 présente la nouvelle configuration du système stéréo, après la rectification, dans laquelle nous appliquons la triangulation optique pour estimer les coordonnées 3D. Les deux points 3D O_l et O_r

constituent les deux centres optiques des deux caméras gauche et droite après la rectification. Dans la nouvelle configuration, les deux centres optiques ont les mêmes coordonnées dans les deux repères images gauche et droite. Aussi, les deux primitives M_l et M_r ont une même ordonnée et deux abscisses différents x_l et x_r dans le repère de la caméra droite. La reconstruction des coordonnées 3D du point M se ramène à une simple application du théorème d'intersection de Thalès dans le triangle (MO_lO_r).

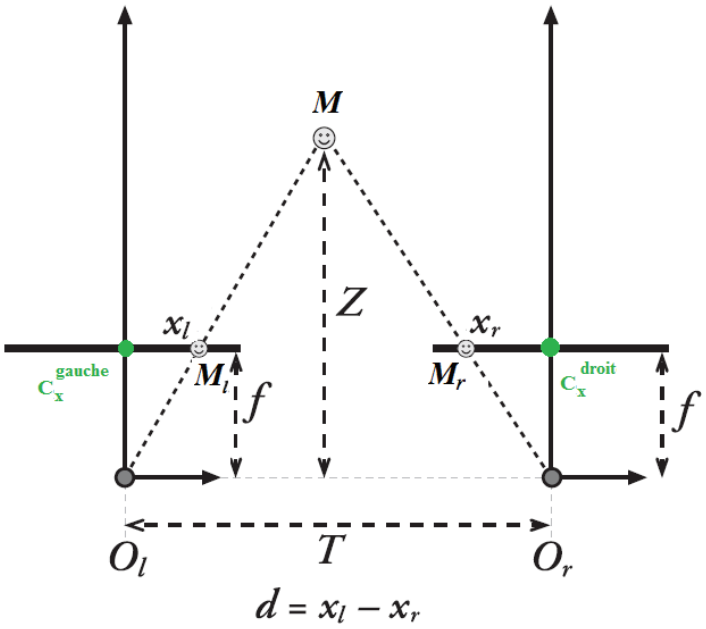


FIGURE 3.12 – La triangulation optique dans la nouvelle configuration du système.

La distance T représente la ligne de base qui constitue la distance séparant les deux centres optiques O_l et O_r. La quantité f n'est autre que la distance focale qui caractérise les deux caméras gauche et droite et qui sépare le centre optique de chaque caméra de son plan image. Le théorème de Thalès dans le triangle (MO_lO_r) se traduit par l'équation (3.12).

$$\frac{T + x_r - x_l}{Z - f} = \frac{T}{Z}, Z = f\frac{T}{x_l - x_r}. \tag{3.12}$$

La disparité d se définit par la distance séparant x_l et x_r. La distance focale f et la ligne de base T ont été calculées au moment de l'étalonnage stéréo, la disparité a été obtenue suite à l'appariement stéréo. Ainsi, l'estimation de la profondeur Z du point 3D P est

obtenue par l'équation (3.13).

$$d = x_l - x_r, Z = f\frac{T}{d}. \tag{3.13}$$

3.7 Densification par les splines cubiques

Les primitives localisées dans les images sont denses dans la direction Y puisque l'échantillonnage se fait pour chaque ligne épipolaire. Néanmoins, les primitives sont non-denses dans la direction X puisque les points élus sont uniquement les points situés sur les bords des franges. Cette étape d'interpolation permet de densifier le nuage de points 3D obtenu par la triangulation optique. L'objectif est d'augmenter le nombre de points représentant la structure géométrique des modèles 3D du visage tout en conservant la précision de chaque point du modèle en ajoutant des points intermédiaires entre les primitives sur l'axe X.

Le principe de l'interpolation consiste à trouver une fonction dont la courbe passe par tous les points d'un ensemble déterminé à l'avance appelés points de contrôle. Nous adoptons un modèle d'interpolation basé sur des splines cubiques. Il s'agit de considérer une série de $(n+1)$ points de coordonnées (x_i, y_i). Nous cherchons une fonction pour chaque intervalle $[x_i, x_{i+1}]$ reliant les points i et $i+1$. Cette fonction constitue un polynôme d'interpolation local de degré 3 ou plus précisément une spline cubique. Il s'agit donc de déterminer n polynômes cubiques $(f_i)_{0\leq i\leq n}$ afin de former la fonction d'interpolation. Les polynômes cubiques sont de la forme :

$$f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i, x_{i-1} \leq x \leq x_i, i \in \{1, 2, ..., n\}. \tag{3.14}$$

Ces polynômes ont leurs dérivées secondes continues. Pour déterminer les coefficients a_i, b_i, c_i et d_i, nous utilisons cette condition de continuité qui s'écrit comme suit :

$$f_i^2(x_i) = f_i^2, f_i^2(x_{i-1}) = f_i^2 - 1, i \in \{1, 2, ..., n\}. \tag{3.15}$$

Les splines cubiques constituent des polynômes d'interpolation locaux de degrés 3 continuellement dérivables. Ces modèles sont les plus adaptés du fait que leurs courbes

d'interpolation sont lisses et continues et passent par les points 3D de contrôle obtenus par la triangulation optique. Aussi, ces polynômes par morceau d'ordre 3 ne présentent pas de fortes oscillations contrairement aux courbes d'interpolation d'ordre supérieur.

f^k constitue la dérivée d'ordre k de la fonction f. Pour garantir une interpolation fidèle du modèle numérique au modèle d'origine, il est indispensable de disposer d'un ensemble significatif de points d'appui (ou de contrôle). En effet, les franges projetées sur le visage doivent être suffisamment serrées pour avoir le maximum de points 3D issus de la triangulation optique.

3.8 Maillage

La première étape consiste à estimer le diagramme de $Voronoi$ qui est une structure géométrique représentant une information de proximité a propos d'un ensemble de points ou d'objets [de Berg *et al.* 2000]. Etant donné un ensemble de sites ou d'objets, nous partitionnons le plan en associant à chaque point son site le plus proche. Les points qui n'ont pas un unique site plus proche, forment le diagramme de $Voronoi$ à savoir que les points du diagramme de $Voronoi$ sont équidistants à deux sites ou plusieurs. Ainsi, pour un ensemble S de n sites, le diagramme de $Voronoi$ $VD(S)$ est obtenu par une partition du plan en blocs de points. Les points formant un même bloc ont le même site le plus proche ou les mêmes sites les plus proches. La figure 3.13 montre un exemple d'un diagramme de $Voronoi$.

Ensuite, nous estimons le diagramme de $Delaunay$ qui a été proposé par Boris Delaunay en 1934 [Delaunay 1934]. La triangulation $Delaunay$ d'un ensemble de points est obtenu en collectionnant les arêtes satisfaisant la propriété du $cercle\ vide$: pour chaque arête, nous pouvons trouver un cercle contenant uniquement les deux points formant l'arête mais aucun autre point de plus [de Berg *et al.* 2008].

La triangulation de $Delaunay$ constitue la structure duale du diagramme de $Voronoi$ dans R_2. En d'autres termes, nous ne pouvons dessiner un segment d'une droite entre deux points du diagramme de $Voronoi$ que si leurs polygones $Voronoi$ ont une arête commune. Dans une terminologie mathématique, il y a une bijection naturelle entre les deux qui inverse les inclusions des facettes. Le cercle circonscrit à un triangle $Delaunay$

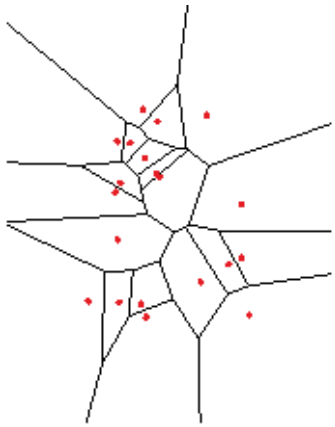


FIGURE 3.13 – Un exemple d'un diagramme de $Voronoi$.

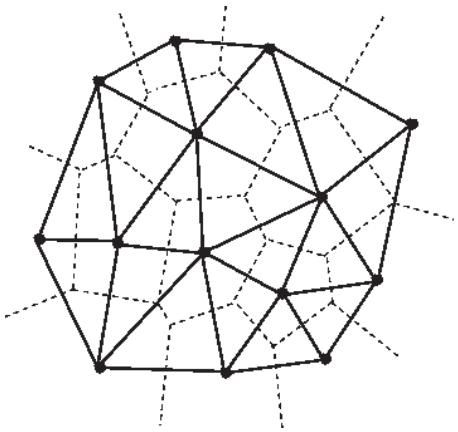


FIGURE 3.14 – Triangulation $Delaunay$ basée sur le diagramme $Voronoi$ affiché en lignes pointillées.

est appelé cercle $Delaunay$.

3.9 Etude expérimentale

Notre système de numérisation est formé de deux caméras et d'une source de projection de lumière. Les caméras utilisées sont reliées à un ordinateur qui constitue la station de numérisation hébergeant notre logiciel. Les connectiques utilisées pour relier les caméras à la station de numérisation sont de type $RJ45$. Les caméras utilisées sont de type $AXIS\ 210$ disposant d'une résolution 640x480 pixels (300 Kilo pixels) et d'une fréquence inférieure ou égale à 30 images progressives couleurs par seconde. Nous utilisons un vidéoprojecteur LCD de type $3M\ MP7740i$ de résolution 1024x768 pixels.

Dans cette section, nous commençons par étalonner le système. Ensuite, nous décrivons les différentes étapes nécessaires pour la numérisation d'un visage. Nous proposons aussi de mesurer la performance du système conçu. Ainsi, nous menons une étude quantitative de la précision et de la régularité de notre reconstruction 3D. Enfin, nous mettons en œuvre une comparaison de notre solution de numérisation avec une vérité terrain.

3.9.1 Etalonnage du système

Nous utilisons la bibliothèque OpenCV pour l'étalonnage de notre système de numérisation. Nous capturons 15 couples d'images gauche et droite d'un échiquier pour l'étalonnage des caméras. Les différentes images d'étalonnage sont présentées sur la figure 3.15.

L'étalonnage stéréo nous fournit les deux nouvelles matrices intrinsèques A_l et A_r des deux caméras gauche et droite après la rectification de leurs plans images respectifs. Dans cette nouvelle configuration, les deux matrices A_l et A_r sont identiques. Elles sont définies par l'équation (3.16) en unités de pixels avec une même distance focale et deux points de projection gauche et droite ayant les mêmes coordonnées respectivement dans le repère gauche et droite. La nouvelle ligne de base T est de $3.5851pixels$ et la transformation entre les deux plans images rectifiés se réduit à une simple translation comme le montre la figure 3.8 de la section 3.3.5.

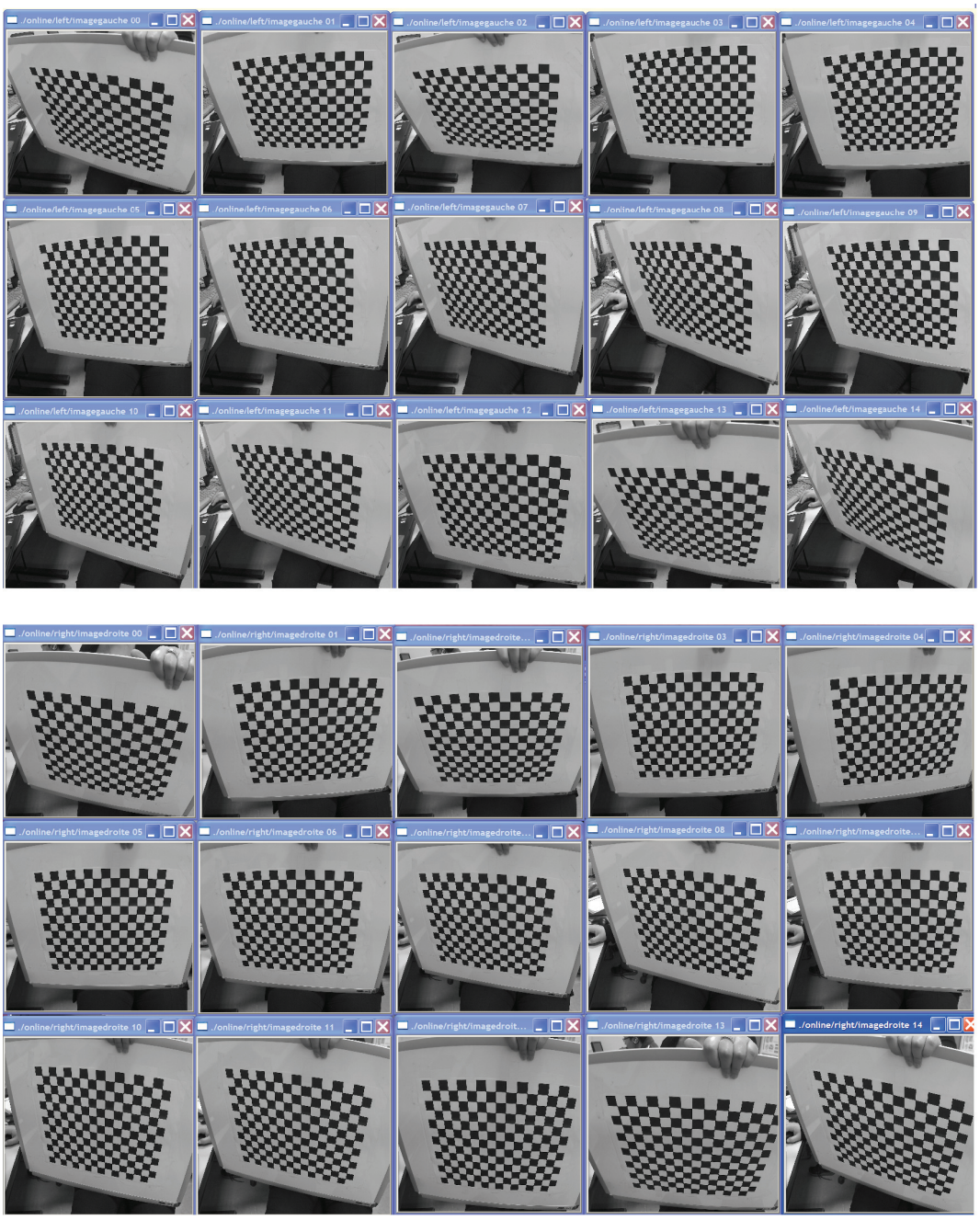


FIGURE 3.15 – Les images gauches et droites utilisées pour l'étalonnage du système.

$$A_l = A_r = \begin{bmatrix} f_x = 717.1575 & 0 & c_x = 317.9088 \\ 0 & f_y = 717.1575 & c_y = 254.4881 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.16}$$

L'étalonnage permet aussi d'estimer les deux vecteurs $dist_l$ et $dist_r$ de la distorsion des deux caméras sous la forme $(k_1, k_2, k_3, p_1, p_2)$. Les trois premiers paramètres (k_1, k_2, k_3) définissent la distorsion radiale les deux paramètres (p_1, p_2) caractérisent la distorsion tangentielle. $dist_l$ et $dist_r$ sont définies par l'équation (3.17) en unités de pixels. Lors de la capture de couples d'images gauche/droite, la première étape consiste ainsi à corriger la distorsion en utilisant les deux vecteurs $dist_l$ et $dist_r$. Ensuite, la rectification des images capturées du visage est assurée en utilisant deux matrices de transformation perpective de taille 640x480. Une des deux matrices est appliquée sur les images de la caméra gauche et la deuxième matrice est appliquée sur les images de la caméra droite.

$$\begin{aligned} dist_l &= \begin{bmatrix} 0.1021 & -0.0776 & 0.0109 & -0.0682 & -1.3950 \end{bmatrix}, \\ dist_r &= \begin{bmatrix} -0.0738 & -1.6447 & -0.0003 & 0.0481 & 13.2580 \end{bmatrix}. \end{aligned} \tag{3.17}$$

3.9.2 Numérisation 3D d'un visage

Pour numériser un visage, nous projetons successivement les deux patrons complémentaires de franges blanches et noires. Nous capturons trois couples d'images gauche/droite. Les deux premiers couples d'images capturent la distorsion des deux patrons complémentaires sur le visage. Le dernier couple d'images gauche/droite permet de récupérer la texture du visage. La première étape est la correction de la distorsion radiale et tangentielle en utilisant les deux vecteurs $dist_l$ et $dist_r$. La figure 3.16 affiche le résultat de la correction sur un couple d'images gauche/droite. La rectification permet ensuite de rendre les lignes épipolaires parallèles. Pour reconstruire le visage, nous proposons de le localiser dans un rectangle en utilisant la technique proposée par [Viola & Jones 2001]. La technique se base sur le classifieur de Haar pour la détection de visages. La figure 3.17 présente le résultat de la localisation sur les images rectifiées. Les étapes de l'échantillonnage, de l'appariement stéréo et de la triangulation optique s'effectuent par la suite uniquement sur les images

gauche et droite présentées sur la figure 3.18.

FIGURE 3.16 – Correction de la distorsion radiale et tangentielle.

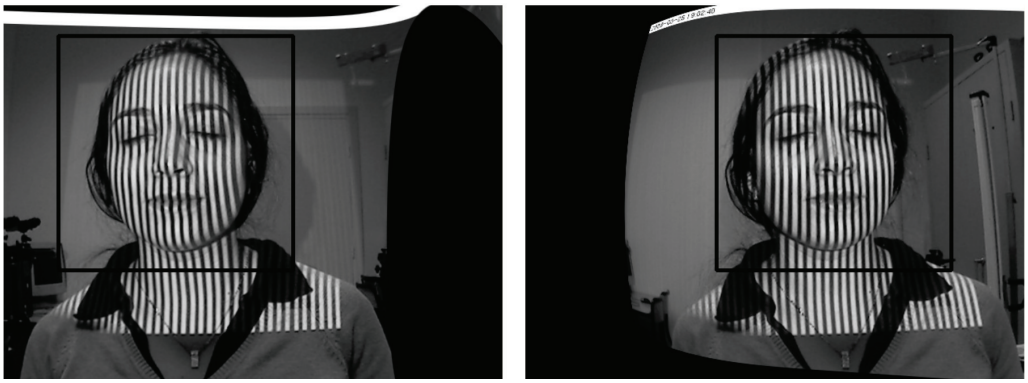


FIGURE 3.17 – Rectification des deux images gauche et droite et détection du visage par le classifieur de Haar.

L'échantillonnage s'effectue pour les deux vues gauche et droite séparément. D'abord, une interpolation par les splines cubiques sur les deux profils complémentaires de chaque ligne épipolaire permet d'obtenir deux courbes sinusoïdales comme le montre la figure 3.19. Les deux courbes définissent ainsi la variation de l'intensité pixélique sur l'axe horizontal pour une ligne épipolaire donnée. L'intersection entre les deux courbes sont les primitives que nous utilisons par la suite pour la numérisation 3D du visage. Les primitives sont des points réels localisés sur les bords des franges avec une précision sous-pixélique. Les primitives échantillonnées sur les deux vues gauche et droite sont colorées en rouge sur la figure 3.20.

La triangulation optique fournit un nuage de points 3D dense sur l'axe Y et non-dense sur l'axe X. Sa densité horizontale dépend de la largeur et du nombre de franges projetées sur la surface faciale. Le nuage de points non-dense reconstruit apparait sur la figure 3.21.a.

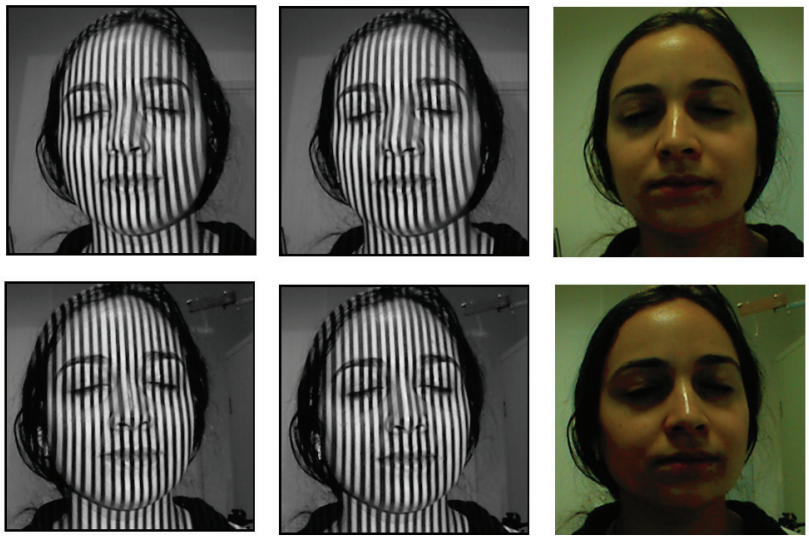


FIGURE 3.18 – Les images gauches et droites rectifiées que nous utilisons pour reconstruire le visage 3D.

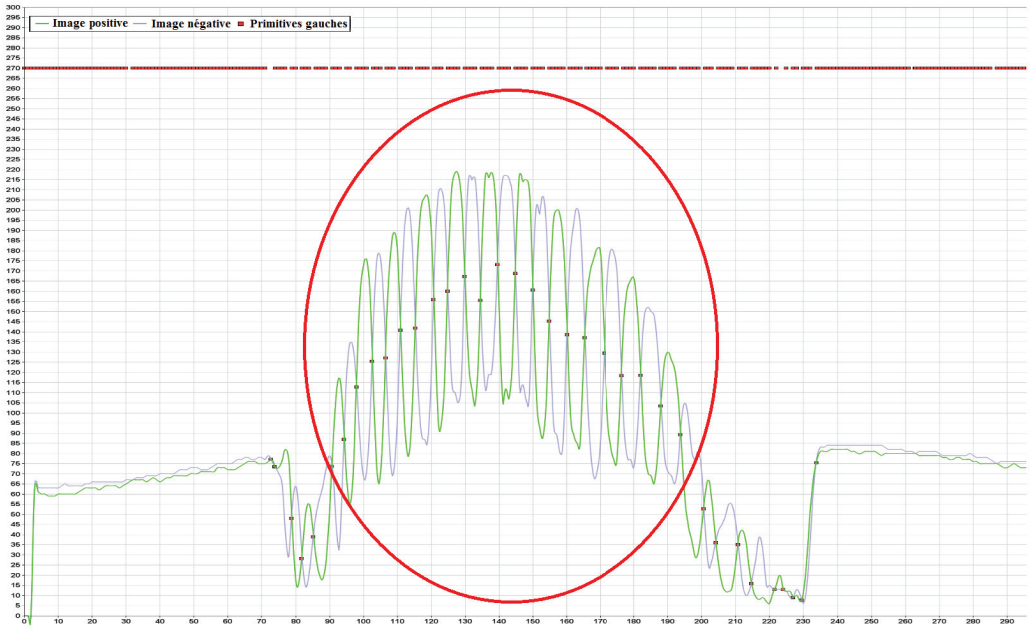


FIGURE 3.19 – Intersection entre les deux profils complémentaires d'une ligne épipolaire gauche.

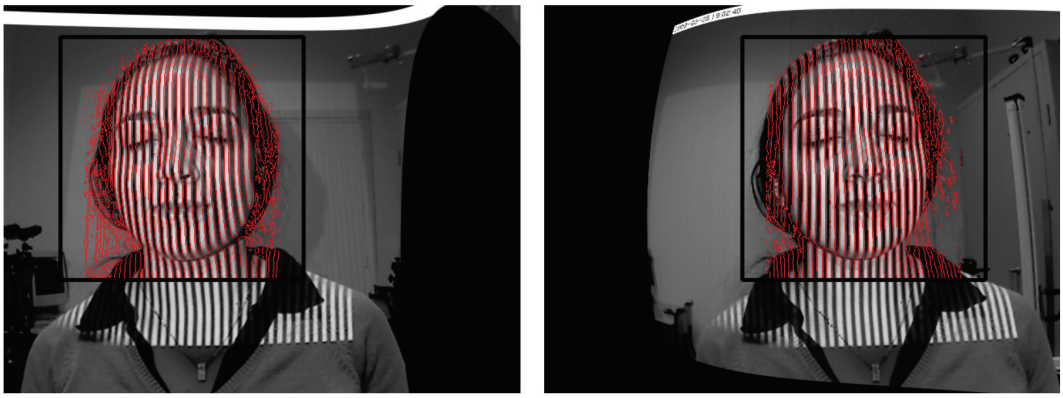
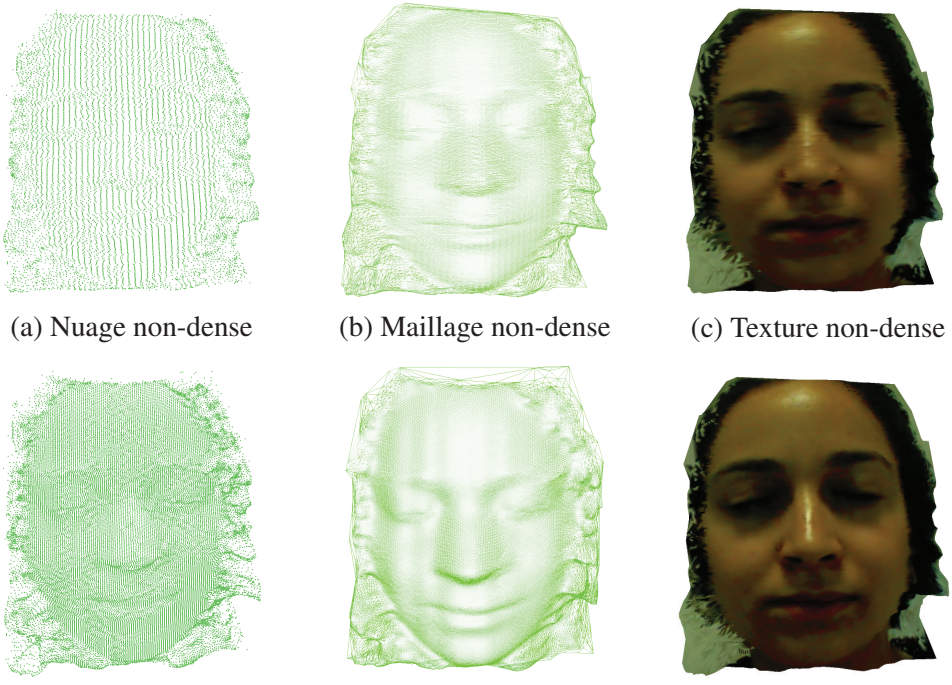


FIGURE 3.20 – Les primitives échantillonnées sur les deux vues gauche et droite.



(a) Nuage non-dense (b) Maillage non-dense (c) Texture non-dense

(d) Nuage dense (e) Maillage dense (f) Texture dense

FIGURE 3.21 – Numérisation 3D par stéréovision active.

Son maillage s'affiche dans la figure 3.21.b et le plaquage de texture associé se présente sur la figure 3.21.c. La densification du nuage de points 3D sur l'axe X est assurée en insérant entre chaque couple de points d'une même ligne épipolaire deux points en utilisant l'interpolation par les splines cubiques comme le montre la figure 3.21.d. Le résultat du maillage du nuage de points dense s'affiche sur la figure 3.21.e. Le résultat du plaquage de texture apparait sur la figure 3.21.f.

3.9.3 Evaluation des performances

Pour estimer la qualité du système de numérisation proposé, nous proposons d'estimer la précision et la régularité de la reconstruction. Nous utilisons un plan uniformément blanc comme objet de référence. L'évaluation de notre système de numérisation est réalisée en étudiant quantitativement et qualitativement la numérisation 3D du plan.

3.9.3.1 Précision du système

La précision d'un système de numérisation est une propriété intrinsèque qui caractérise la déformation engendrée par la technique de numérisation 3D sur la forme réelle de l'objet numérisé. D'abord, il s'agit de reconstruire le modèle 3D du plan, de calculer son équation théorique moyennant 3 de ses points. Ensuite, nous mesurons pour chaque point du plan numérisé la distance qui le sépare du plan défini par l'équation. En analysant la déviation spatiale de tous les points numérisés du plan par rapport à son équation théorique, nous pouvons caractériser la précision de la reconstruction. En effet, l'erreur de précision s'évalue par une analyse qualitative et quantitative du vecteur de déviation spatiale.

En unités de pixel, en étudiant le vecteur de la déviation spatiale du nuage de points du plan reconstruit par rapport à son équation théorique, nous obtenons les informations suivantes. La valeur maximum est de $0.2951pixel$, celle minimum est égale à $6.7731 * 10 - 7pixel$, l'écart type est de $0.0756pixel$ et la valeur moyenne est $0.1207pixel$. L'erreur de précision moyenne en unité de pixel est donc de $0.1207pixel$. Les facteurs d'agrandissement k_x en x et k_y en y sont calculés moyennant 3 points références sur le plan. k_x est de $1.8635mm/pixel$ et k_y est de $1.1209mm/pixel$. k_x et k_y nous permettent de calculer les coordonnées x, y et z en mm du plan reconstruit. Les deux facteurs d'agran-

dissement sont différents uniquement pour les expérimentations effectuées sur le plan parce que nous n'avons pas exigé au moment de l'étalonnage stéréo de garder les mêmes dimensions de l'image après la rectification. En unités de millimètres, l'étude du vecteur de la déviation spatiale du nuage de points du plan reconstruit par rapport à son équation théorique nous fournit les informations suivantes. La valeur maximum est de $0.2949mm$, celle minimum est égale à $1.9878*10-6mm$, l'écart type est de $0.0767mm$ et la valeur moyenne est $0.1221mm$. En d'autres termes, l'erreur de précision moyenne en mm est $0.1221mm$. La résolution spatiale du système est de l'ordre de $15000points$ sur un visage 3D. L'histogramme de la figure 3.22 présente la déviation spatiale en mm des points 3D formant le plan reconstruit par notre système par rapport à son équation théorique.

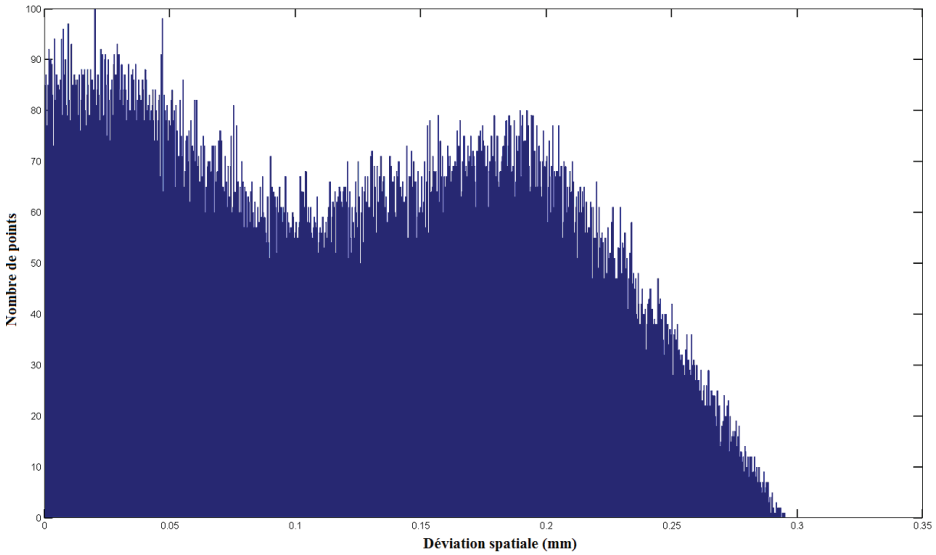


FIGURE 3.22 – Etude de la précision de notre système de numérisation : Histogramme de la déviation spatiale en mm du plan reconstruit par notre technique par rapport à son équation théorique.

3.9.3.2 Régularité de la reconstruction

La régularité de la reconstruction est décrite par la variation des distorsions locales de la reconstruction. Il s'agit de calculer la déviation spatiale entre le nuage de points du plan reconstruit et une équation approximative du plan calculée à partir de la totalité des points par la méthode des moindres carrés. Ensuite, nous mesurons la déviation spatiale

de chaque point du plan par rapport à l'équation approximative du plan. La régularité de la reconstruction du plan s'évalue ensuite par une analyse qualitative et quantitative du vecteur de la déviation spatiale. Ceci permet aussi de mesurer les distorsions locales.

En étudiant le vecteur de la déviation spatiale du nuage de points du plan reconstruit par rapport à son équation approximative, nous obtenons les informations suivantes. La valeur maximum est de $0.0058mm$, celle minimum est égale à $4.4043 * 10 - 8mm$, l'écart type est de $8.5306 * 10 - 4mm$ et la valeur moyenne est $0.0011mm$. La mesure de la régularité moyenne en mm est donc de $0.0011mm$. Nous présentons sur la figure 3.23 l'histogramme de la déviation spatiale en mm des points 3D formant le plan reconstruit par notre système par rapport à son équation approximative.

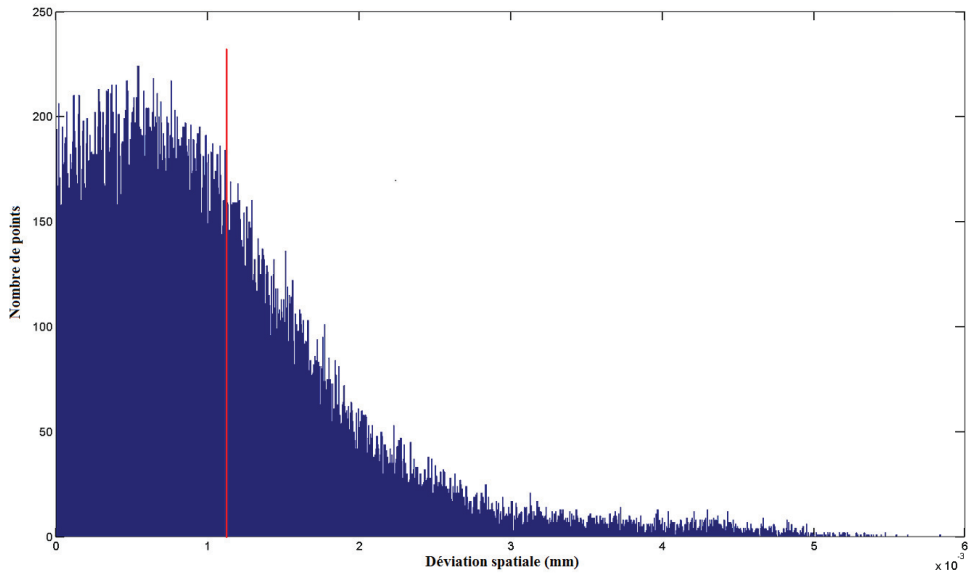


FIGURE 3.23 – Etude de la régularité de notre système de numérisation : Histogramme de la déviation spatiale en mm du plan reconstruit par notre système par rapport à son équation approximative.

3.9.4 Etude comparative avec une vérité terrain

Cette section analyse la qualité de notre reconstruction en la comparant avec une vérité terrain. Nous utilisons le système de balayage laser Minolta VI300 pour sa précision. Ci-dessous, une étude comparative qualitative et quantitative entre le système proposé et le scanner Minolta est décrite. Une reconstruction du plan a été assurée par un scanner Mi-

nolta VI300 de balayage laser pour la mesure de la précision et des distorsions locales du système de numérisation proposé. Les distorsions locales engendrées par la numérisation laser du plan sont aussi estimées. Ensuite, un appariement rigide des deux plans reconstruits permet de mesurer leur déviation spatiale. La figure 3.24 présente les deux plans ainsi que le résultat de leur appariement. Sur l'image la plus à droite, le plan coloré en vert est obtenu par la technique MINOLTA et notre plan est coloré en rouge. Notons que le plan reconstruit n'est pas matte ce qui explique les défauts de reconstruction que nous pouvons remarquer surtout sur le plan reconstruit par la technique laser.

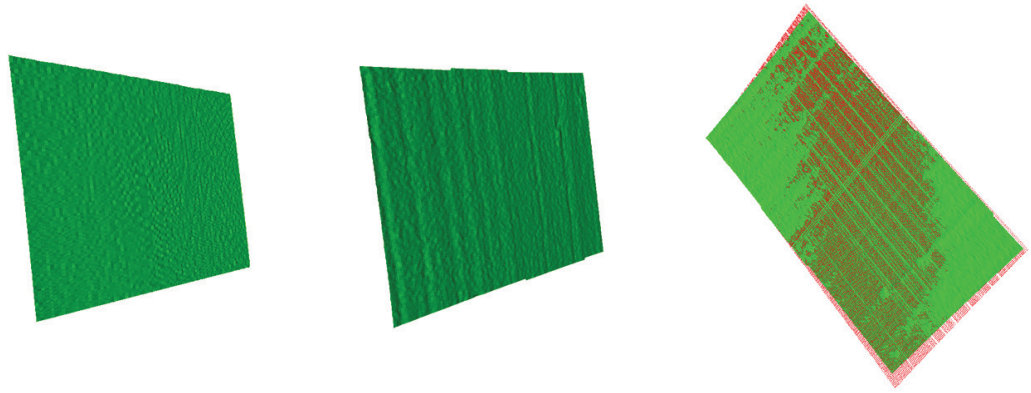


FIGURE 3.24 – Appariement rigide 3D entre le plan reconstruit par notre approche et celui obtenu par un balayage laser : Le premier plan à gauche est reconstruit par la technique de numérisation proposée, le second est reconstruit par le scanner laser Minolta VI300, le résultat de l'appariement est sur l'image la plus à droite.

La figure 3.25 illustre l'histogramme de la déviation spatiale en mm du plan reconstruit par la technique laser par rapport à son équation théorique calculée moyennant 3 de ses points. La valeur maximum est de $0.0201mm$, celle minimum est égale à $0mm$, l'écart type est de $0.0052mm$ et la valeur moyenne est $0.0081mm$. En d'autres termes, l'erreur de précision moyenne en mm est $0.0081mm$. Etant donné que l'erreur de précision moyenne obtenue pour notre système est de $0.1221mm$, le ratio de la précision de reconstruction est $0.1221mm/0.0081mm = 15.0741$. Notre reconstruction est donc 15.0741 fois moins précise que celle du laser.

En étudiant le vecteur de la déviation spatiale du nuage de points du plan reconstruit par la technique laser par rapport à son équation approximative évaluée par la totalité des

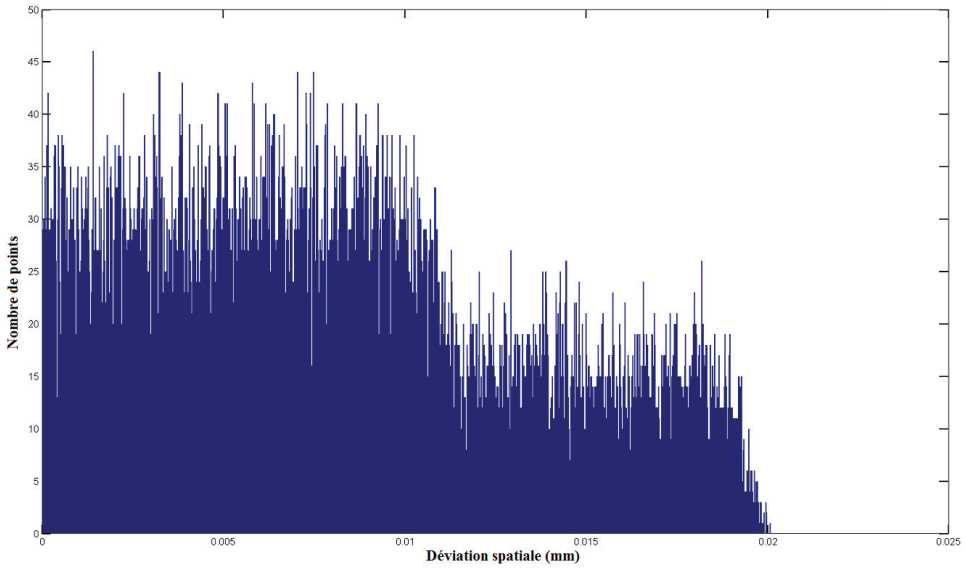


FIGURE 3.25 – Etude de la précision de la numérisation laser : Histogramme de la déviation spatiale en mm du plan reconstruit par la technique laser par rapport à son équation théorique.

points, nous obtenons les informations suivantes. La valeur maximum est de $0.0017mm$, celle minimum est égale à $8.2015 * 10 - 9mm$, l'écart type est de $1.8312 * 10 - 4mm$ et la valeur moyenne est $0.00024mm$. L'erreur de régularité moyenne en mm est donc de $0.00024mm$. Nous présentons sur la figure 3.26 l'histogramme de la déviation spatiale en mm des points 3D formant le plan reconstruit par la technique laser par rapport à l'équation approximative de ce plan. Etant donné que l'erreur de régularité moyenne obtenue pour notre système est de $0.0011mm$, le ratio de la régularité de reconstruction est $0.0011mm/0.00024mm = 4.5833$. Notre reconstruction est 4.5833 fois moins régulière que celle du laser.

Un recalage rigide entre le plan reconstruit par notre système avec celui reconstruit par Minolta est assuré par l'algorithme classique ICP : Iterative Closest Point. Nous récupérons le vecteur de déviation spatiale entre les deux nuages de points des deux plans reconstruits. La valeur maximum est de $64.7590mm$, celle minimum est égale à $0mm$, l'écart type est de $2.3417mm$ et la valeur moyenne est $0.4319mm$. Le nombre de couples de points total est de 57888 couples. Le nombre de couples ayant une déviation spatiale supérieure à $2mm$ est de 1782. Ils constituent 3.08% de la totalité des couples considérés.

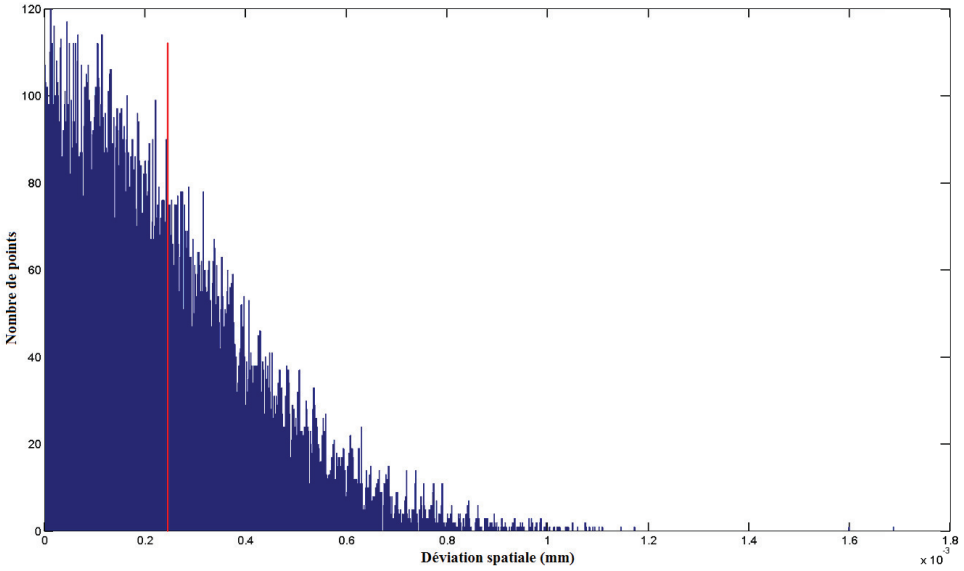


FIGURE 3.26 – Etude de la régularité de la numérisation laser : Histogramme de la déviation spatiale en mm du plan reconstruit par la technique laser par rapport à son équation approximative.

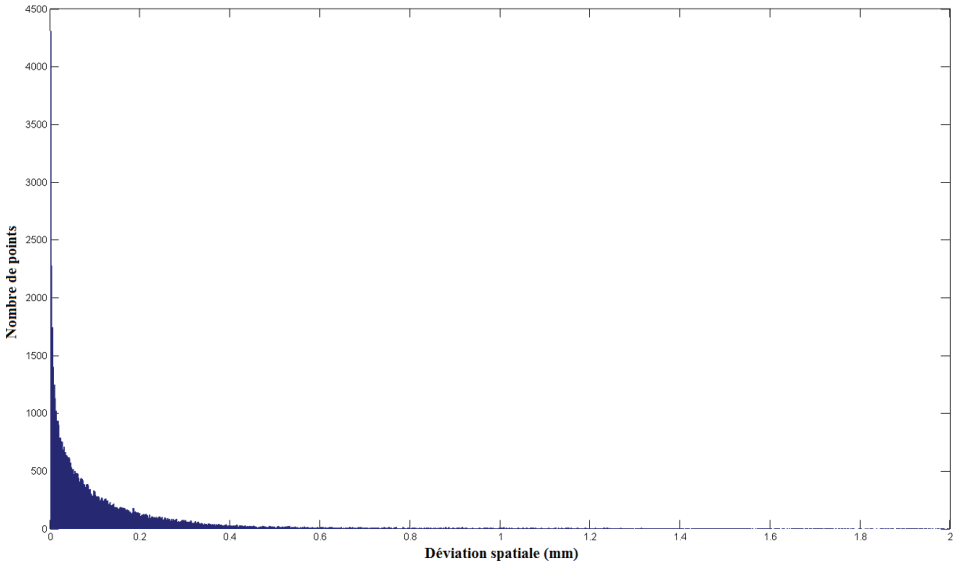


FIGURE 3.27 – Histogramme de la déviation spatiale en mm entre les deux plans reconstruits par la technique de numérisation proposée et par celle de Minolta.

En ne gardant que les lignes ayant une déviation spatiale inférieure à $2mm$, nous obtenons les informations suivantes. La valeur maximum est de $1.990mm$, celle minimum est égale à $0mm$, l'écart type est de $0.2367mm$ et la valeur moyenne est $0.1383mm$. La figure 3.27 illustre l'histogramme de la déviation spatiale en mm entre les deux plans reconstruits par la technique de numérisation proposée et par celle de Minolta.

3.10 Conclusion

Dans ce chapitre, nous avons défini une approche de numérisation 3D particulièrement adaptée aux surfaces mal texturées comme le visage et le corps par une projection d'un patron de franges noires et blanches alternées. Cependant, cette approche fournit un modèle épars dont la résolution sur l'axe des abscisses est proportionnelle au nombre de franges distordues sur le visage.

Pour récupérer plus de détail de la forme faciale 3D, nous proposons dans le chapitre suivant de modifier la lumière structurée en projetant des patrons sinusoïdaux au lieu des patrons binaires. Nous suggérons aussi de décoder l'information pixélique à l'intérieur des franges sinusoïdales par un décodage de la valeur de la phase pour chaque pixel séparément. Ceci nous permet de récupérer l'information 3D réelle pour chaque pixel situé à l'intérieur des franges remplaçant ainsi leur approximation avec l'interpolation par les splines cubiques.

CHAPITRE 4

Numérisation 3D par Décalage de Phase

4.1 Introduction

Certes, dans le domaine de l'animation 3D, seule une reconstruction 3D texturée réaliste est requise. Néanmoins, la précision de la numérisation 3D constitue un facteur critique pour réussir une identification faciale en vidéosurveillance ou pour le suivi post-opératoire d'une chirurgie faciale. En effet, la précision d'un visage numérisé caractérise sa fidélité à sa forme réelle. Pour un système multi-caméra, la précision varie conjointement avec la résolution spatiale des caméras. Ainsi, plus nous utilisons des pixels dans les images, plus nous obtenons des points 3D pour une forme faciale plus fidèle.

La technique de stéréovision active, que nous proposons dans le chapitre 3, fournit une résolution spatiale dépendante de la largeur des franges. En fait, plus les franges sont étroites, plus la surface reconstruite est fidèle à la forme réelle de l'objet. Ainsi, la résolution reste toujours non-dense et inférieure à une résolution pixélique. Aussi, la densification avec les splines cubiques permet une approximation de la surface faciale. Cette interpolation ne se base que sur la recherche d'une fonction qui passe par tous les points du modèle, sur une simple hypothèse de continuité.

Ici, l'idée est de densifier le modèle facial 3D du visage en utilisant le plus de pixels possibles dans les images capturées. Ceci permet une mesure dense 3D du visage et pas une simple approximation. Nous proposons d'abord de projeter une nouvelle lumière structurée sinusoïdale pour différencier les pixels à l'intérieur des franges. Nous suggérons ensuite une nouvelle approche hybride de numérisation 3D pour obtenir une reconstruction dense et précise. Elle emploie la stéréovision et la codification sinusoïdale par décalage de phase

en tirant profit de leurs atouts et en palliant leurs faiblesses.

4.2 Principe

La technique proposée utilise un banc stéréo calibré et un vidéo-projecteur non-calibré. La figure 4.1 définit les différentes étapes nécessaires pour la numérisation 3D d'un visage par notre approche hybride. D'abord, une étape d'étalonnage dans la phase hors-ligne calcule les paramètres intrinsèques et extrinsèques des deux caméras, détermine les distorsions radiale et tangentielle et estime la géométrie épipolaire comme le propose [Zhang 1999]. Aussi, une paramétrisation de la distorsion gamma est aussi assurée en hors-ligne. La distorsion gamma est une erreur non-linéaire qui s'ajoute à la lumière envoyée par le vidéo-projecteur en dégradant le profil sinusoïdal de la lumière structurée projetée.

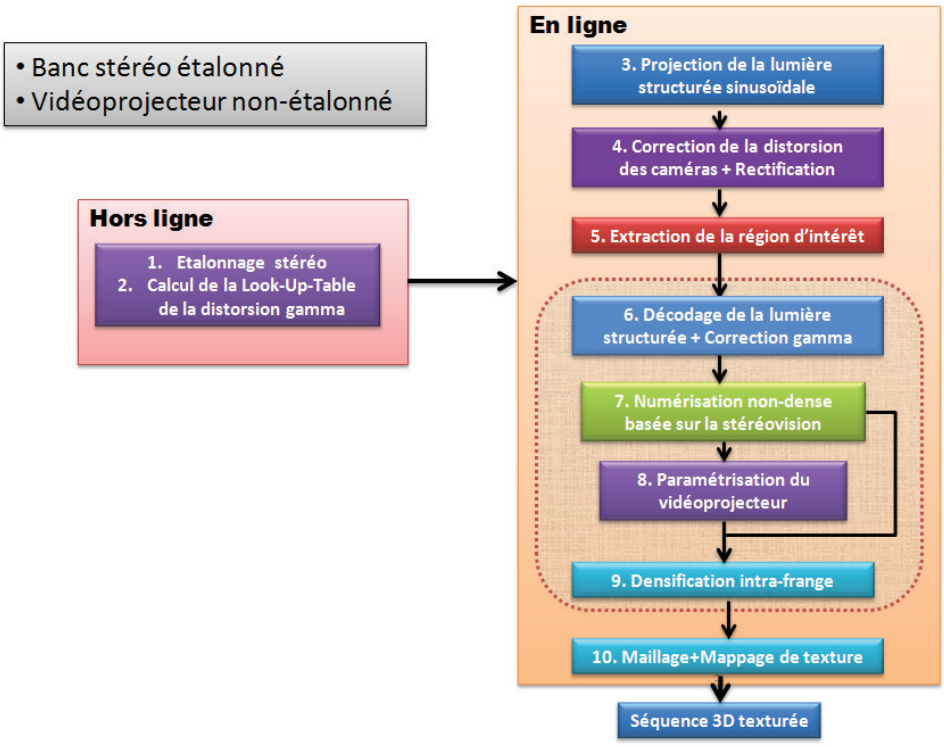


FIGURE 4.1 – Notre approche de numérisation 3D hybride.

Dans la phase en ligne, deux patrons sinusoïdaux en opposition de phase et un troisième patron uniformément blanc sont projetés sur le visage. Trois couples d'images gauches et

droites sont capturées. Nous corrigeons d'abord les distorsions et nous rectifions les images en utilisant les paramètres intrinsèques et extrinsèques ainsi que la géométrie épipolaire calculés en hors-ligne. Ensuite, pour réduire la complexité temporelle de la numérisation, nous proposons une nouvelle approche de segmentation de la région faciale par une analyse de l'amplitude du signal distordu sur le visage dans l'espace fréquentiel.

Un décodage de la lumière structurée et une correction de la distorsion gamma nous permettent de récupérer toutes les valeurs de phase pour chaque pixel sur les deux vues gauche et droite séparément. Une première numérisation non-dense du visage est assurée par appariement stéréo. Les approches classiques de la lumière structurée par décalage de phase nécessitent un étalonnage hors-ligne du vidéoprojecteur avec les caméras et une étape de déroulement de phase. Contrairement à ces approches, pour apporter plus de flexibilité à notre système de numérisation, nous suggérons une paramétrisation enligne du vidéoprojecteur pour le localiser dans l'espace.

Nous proposons de densifier le modèle par reconstruction géométrique, séparément sur les images stéréo fournies par chaque caméra utilisée. Ainsi, nous utilisons les coordonnées du vidéoprojecteur, les coordonnées du plan image et l'information de phase pour calculer les coordonnées 3D de chaque pixel intrafrange. A la différence des approches classiques de codification sinusoïdale, notre méthode ne nécessite pas une étape de déroulement de phase grâce à l'utilisation de l'appariement stéréo dans la première étape de l'approche. Les deux modèles 3D denses gauche et droite sont ensuite fusionnés. L'intérêt d'une telle approche est de densifier le modèle 3D calculé et de renforcer sa précision. Le maillage et le plaquage de la texture sont assurés pour finaliser le modèle facial 3D.

4.3 Localisation de la région faciale

La motivation ici est de réduire le coût de calcul nécessaire pour générer un modèle 3D dense d'un visage. En effet, la localisation de la région faciale dans les deux vues gauche et droite permet un appariement stéréo plus rapide et efficace. Aussi, cette segmentation 2D permet de ne considérer dans la densification que les pixels du visage ce qui diminue le nombre de points aberrants et optimise le temps de densification. Lorsqu'un patron sinusoïdal est projeté sur le visage, le contraste du patron distordu sur le visage est nette-

ment supérieur à celui observé sur son arrière plan. L'idée est de profiter de la variation de contraste et de segmenter la région faciale par une analyse de l'amplitude du signal distordu sur le visage dans l'espace fréquentiel en utilisant la Transformée de Fourier FFT.

Le procédé de la segmentation est assuré en utilisant uniquement les deux composantes sinusoïdales contenues dans les deux images formées lors de la distorsion des deux patrons sinusoïdaux. Puisque les deux patrons sont en opposition de phase, les deux composantes sont obtenues par une soustraction de leur moyenne. Nous les appelons image positive et image négative. Ceci permet d'éviter la contamination de l'information de la texture et de l'illumination lors du calcul de la transformée de Fourier. La localisation d'un visage dans une vue gauche ou droite nécessite essentiellement cinq étapes faisant intervenir les deux images positive et négative. Les quatre premières étapes sont appliquées sur les deux images gauche et droite séparément. La dernière étape fusionne les deux résultats obtenus et permet la construction du masque binaire 2D de segmentation. Les étapes sont décrites ci-dessous :

1. Sur chaque ligne épipolaire et pour chaque pixel P, nous calculons la FFT sur une fenêtre glissante centrée en P et formée de N pixels. Nous calculons l'amplitude de la FFT ce qui permet de caractériser le pixel P par la courbe de variation de l'amplitude de la FFT en fonction de la fréquence.

2. Ensuite, la variation de l'amplitude de la FFT en fonction de la fréquence sur toute la ligne épipolaire permet de construire une surface qui met en valeur les pixels du visage comme l'affiche la figure 4.2. La FFT est calculée sur une fenêtre glissante de taille $N = 128$ pixels. La taille du visage est de 578x578 pixels capturé par une caméra de dimensions 1600x1200.

3. Cette étape consiste à représenter la variation de l'amplitude de FFT pour chaque ligne épipolaire par une courbe. Nous proposons de considérer pour chaque pixel P une somme pondérée des amplitudes des trois coefficients consécutifs $q-1$, q, et $q+1$ de la FFT avec respectivement les poids 0.3, 1, 0.3. q constitue la fréquence caractéristique du patron de lumière projeté. q est aussi le nombre moyen de périodes (franges noires) contenues dans une fenêtre glissante de taille N.

   Sur la figure 4.2, les quatre premières fréquences ne sont pas affichées car la fenêtre

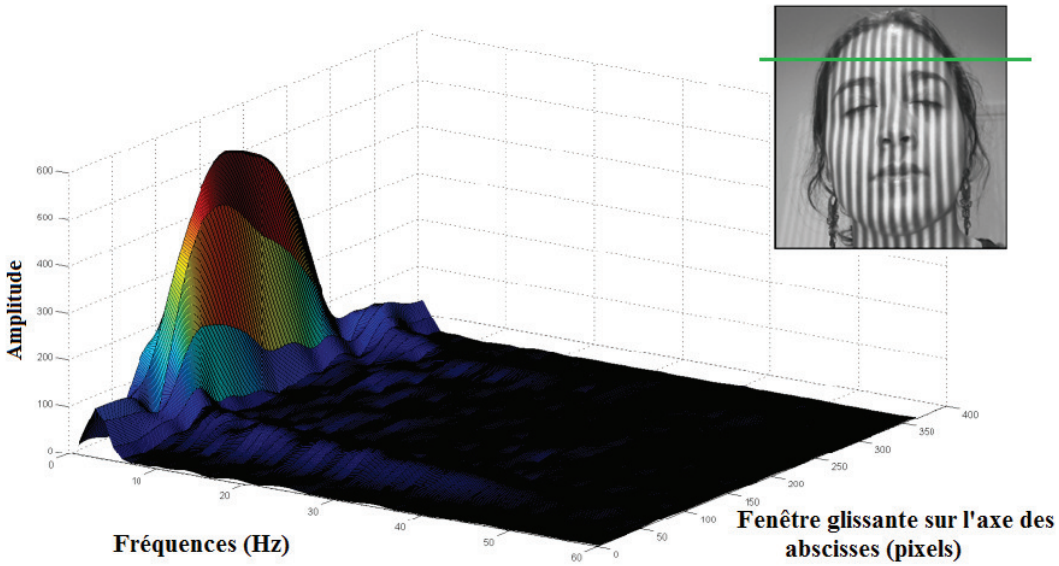


FIGURE 4.2 – Variation de l'amplitude de FFT en fonction de la fréquence. La FFT est calculée sur une fenêtre glissante pour chaque ligne épipolaire séparément.

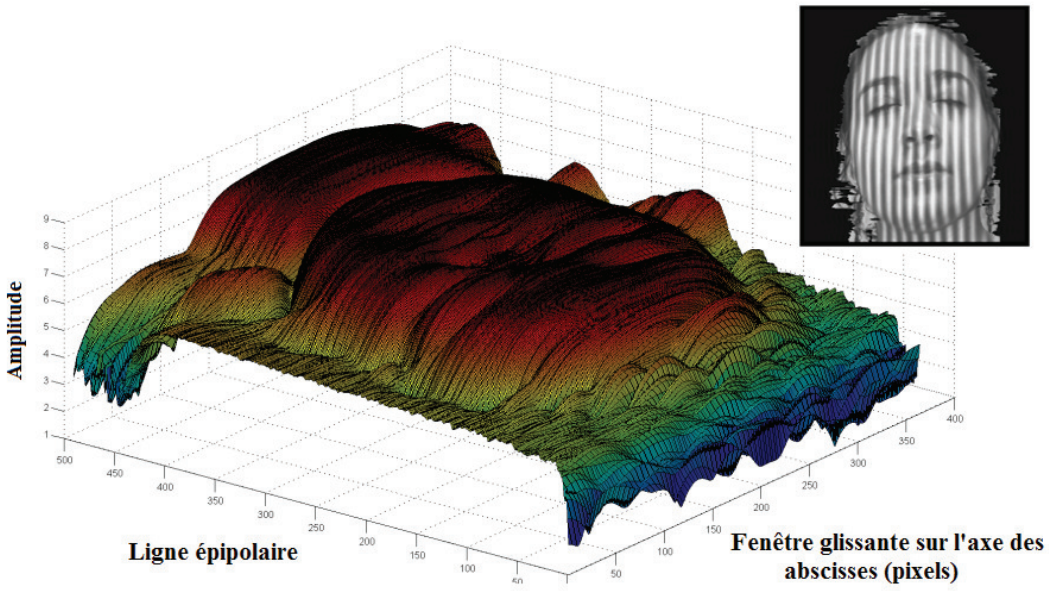


FIGURE 4.3 – Segmentation 2D de la région faciale sur la vue gauche par la technique proposée.

de taille 128 contient en moyenne 6 franges noires. Ainsi, $q = 6$ et nous commencons par la fréquence 5 pour afficher la variation de l'amplitude en fonction de la fréquence sur la figure 4.2. La fréquence caractéristique de notre patron sur la figure est donc la deuxième. La somme pondérée est calculée en considérant uniquement les trois premières fréquences.

4. La représentation de la variation de l'amplitude de la FFT en fonction de la fréquence sur toute l'image est obtenue en utilisant les courbes calculées pour toutes les lignes épipolaires. Ceci permet de construire une surface qui caractérise la région faciale sur toute l'image par une haute valeur d'amplitude de FFT. La figure 4.3 affiche une représentation logarithmique de la surface calculée.
5. Finalement, la localisation du visage se ramène à un seuillage adéquat sur la surface logarithmique pour ne garder que les pixels de haute amplitude et donc situés sur le visage. Une somme des deux surfaces obtenues par l'image positive et celle négative renforce la qualité de la segmentation.

4.4 Décodage de la lumière structurée

Il s'agit d'estimer les valeurs de phase pour chaque pixel sur la vue gauche et celle droite séparément. Ainsi, le décodage de la lumière structurée consiste à résoudre un système de trois équations définis par la distorsion des trois patrons envoyées par le vidéo-projecteur sur le visage. Cette section présente le modèle mathématique associée à notre lumière structurée. De plus, les franges observées sur le visage ne sont pas parfaitement sinusoïdales. Ceci est à cause de la distorsion gamma qui constitue une distorsion non-linéaire qui varie avec la phase. Ainsi, le calcul des valeurs de la phase locale s'effectue en deux étapes. Une première estimation est obtenue par la résolution du modèle mathématique. Une deuxième étape consiste à corriger les valeurs de phases par la correction de la distorsion gamma.

4.4.1 Modèle mathématique

Le modèle proposé est défini par le système d'équations (4.1) et constitue une variante du modèle mathématique proposé par [Zhang 2010]. A l'instant t , $I_p(s,t)$ constitue l'in-

tensité du pixel s sur l'image positive, $I_n(s,t)$ est l'intensité de s sur l'image négative et $I_t(s,t)$ est l'intensité de s dans l'image de texture.

$$\begin{aligned} I_p(s,t) &= I_b(s,t) + I_a(s,t) \cdot \sin(\phi(s,t)), \\ I_n(s,t) &= I_b(s,t) + I_a(s,t) \cdot \sin(\phi(s,t) + \pi), \\ I_t(s,t) &= I_b(s,t) + I_a(s,t). \end{aligned} \tag{4.1}$$

Selon notre approche, un déphasage de π entre les deux patrons sinusoïdaux est optimal pour un scénario de stéréovision ; les échantillons gauches et droites utilisés pour l'appariement stéréo sont localisés avec une précision sous pixélique. Le rôle du dernier patron est double : il permet une normalisation de l'information de phase et aussi, il est utilisé pour texturer le modèle 3D calculé. Ce modèle est défini pour décorréler les signaux sinusoïdaux $I_a(s,t) \cdot \sin(\phi(s,t))$ et $I_a(s,t) \cdot \sin(\phi(s,t) + \pi)$ distordus sur le visage, du terme non-sinusoïdal $I_b(s)$. En fait, $I_b(s,t)$ représente l'information de texture et l'effet de l'illumination et constitue un signal de contamination pour la composante sinusoïdale. $I_a(s,t)$ est l'intensité de modulation. $\phi(s,t)$ est la phase locale définie pour chaque pixel s. $I_b(s,t)$ est calculé comme la moyenne des deux intensités $I_p(s,t)$ et $I_n(s,t)$. $I_a(s,t)$ est ensuite calculée à partir de la troisième équation du système (4.1) et la valeur de la phase $\phi(s,t)$ est estimée par l'équation (4.2).

$$\phi(s,t) = \arcsin\left[\frac{I_p(s,t) - I_n(s,t)}{2 \cdot I_t(s,t) - I_p(s,t) - I_n(s,t)}\right]. \tag{4.2}$$

4.4.2 Correction gamma

La distorsion gamma caractérise le rendu en contraste d'un support photosensible (émulsion photographique ou pellicule, capteur CCD ou CMOS...) ou d'un signal visuel électronique. L'erreur gamma se traduit par une reproduction tonale particulière d'une image plus contrastée et plus sombre que celle réelle.

Ici, l'erreur gamma est introduite par la non-linéarité de la courbe de transfert tension/lumière du projecteur LCD et des caméras CCD que nous utilisons. Cette reproduction tonale particulière provient du fait que la luminosité émise par les cellules photosensibles du vidéoprojecteur et des caméras n'est pas linéairement proportionnelle à la tension électrique appliquée. Ainsi, pour chaque pixel ayant une luminosité réelle l, sa luminosité

observée est l^{γ}. La différence entre le luminosité réelle et celle observée constitue l'erreur gamma. La figure 4.4 présente deux images avant et après la correction gamma et la courbe de la correction gamma qui permet de retrouver les valeurs réelles de la luminosité.

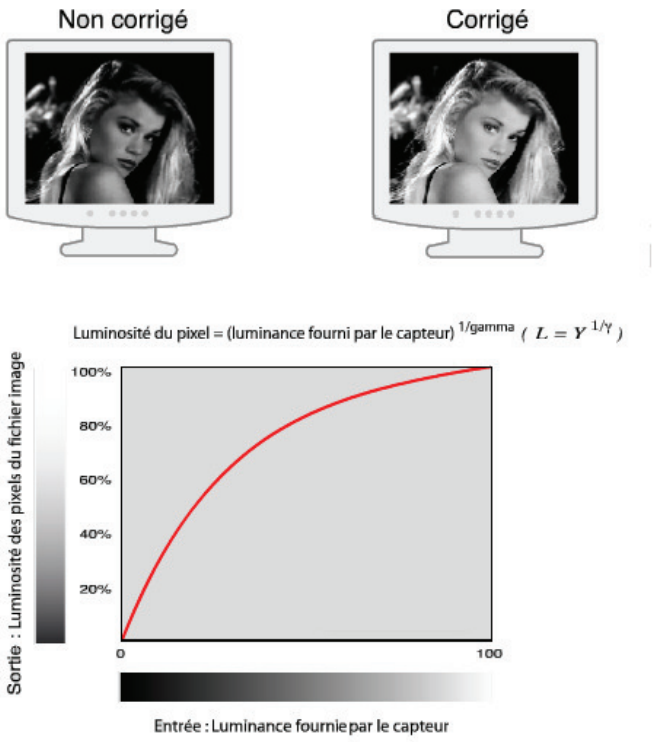


FIGURE 4.4 – Correction gamma.

Lors de la projection du patron sinusoïdal par le projecteur sur l'objet, la distorsion gamma rend des franges sinusoïdales parfaites non sinusoïdales avec une distorsion non-linéaire qui varie avec la phase. La technique la plus utilisée pour la correction gamma est la Look-Up-Table, qui a été proposée par [Zhang & Yau 2007]. Nous faisons recourt à cette technique en hors-ligne pour obtenir une composante sinusoïdale plus régulière et plus fidèle à la sinusoïde envoyée par le vidéoprojecteur. Nous projetons d'abord notre lumière structurée sinusoïdale sur un plan uniformément blanc. Nous capturons les trois couples d'images gauches et droites et nous calculons les valeurs de la phase pour chaque pixel du plan en utilisant notre modèle mathématique.

Idéalement, la variation de la phase est monotone et linéaire pour un plan. Soit M_k un pixel situé entre deux primitives M_0 et M_1 échantillonnées par la stéréovision. Le pixel

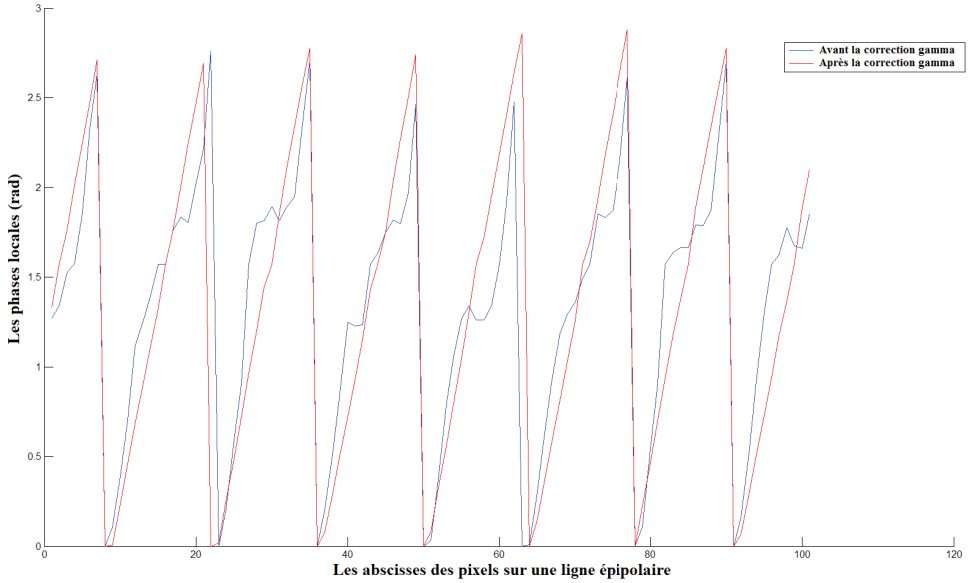
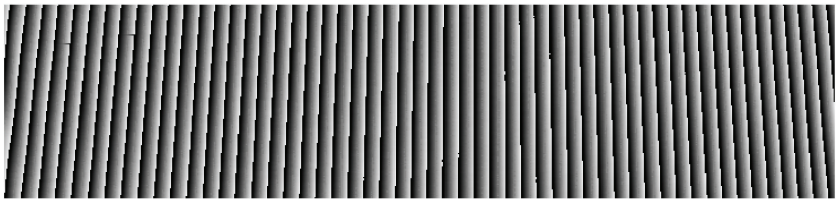


FIGURE 4.5 – Les courbes de la phase locale pour un plan avant et après la correction gamma.

(a) Avant la correction



(b) Après la correction

FIGURE 4.6 – Résultat de l'estimation de la phase locale pour un plan avant et après la correction gamma.

M_k se trouve à une distance de k pixels de M_0. Les phases réelles et parfaites de M_0 et de M_1 sont respectivement $\phi_{ideal}(M_0) = \phi_{real}(M_0) = 0$ et $\phi_{ideal}(M_1) = \phi_{real}(M_1) = \pi$. Soit N le nombre de pixels situés entre M_0 et M_1. Ainsi, la valeur de la phase idéale du pixel M_k est $\phi_{ideal}(M_k) = \frac{k\pi}{N}$. Nous enregistrons ensuite dans la Look-Up-Table pour chaque pixel M_k sa phase réelle $\phi_{real}(M_k)$ et son erreur gamma associée qui constitue la différence entre les deux valeurs de phase réelle $\phi_{real}(M_k)$ et idéale $\phi_{ideal}(M_k)$.

Nous présentons dans la figure 4.5 les courbes de la phase locale pour un plan avant et après la correction gamma. La courbe bleue présente le profil de la phase avant la correction et la courbe rouge présente le profil de la phase suite à la correction gamma. Ensuite, lors de la reconstruction 3D d'un objet quelconque, nous appliquons la Look-Up-Table pour corriger les phases estimées par le modèle mathématique. La figure 4.6 représente le résultat de la correction sur un plan.

4.5 Numérisation 3D hybride

D'abord, nous proposons de calculer un modèle 3D non-dense du visage par une triangulation optique en utilisant uniquement les primitives inter-franges. Ensuite, nous utilisons ce modèle pour paramétriser le vidéoprojecteur et le localiser dans l'espace. Enfin, la densification du modèle facial est assurée par une reconstruction géométrique, séparément sur les images stéréo fournies par les caméras utilisées. Les coordonnées 3D de chaque pixel intrafrange sont calculées en utilisant les coordonnées du vidéoprojecteur, les coordonnées du plan image et son information de phase.

4.5.1 Estimation du modèle 3D non-dense

Le modèle 3D non-dense est généré par stéréovision. Il est formé par les points situés à l'intersection des deux composantes sinusoïdales en opposition de phase de l'image positive et celle négative. Ainsi, la localisation a une précision sous-pixélique. Les primitives gauches et droites appariées, ont nécessairement la même coordonnée Y dans les images rectifiées. Le problème de l'appariement stéréo est résolu pour chaque ligne épipolaire séparément par la programmation dynamique. Le nuage de points 3D non-dense s'obtient par une triangulation optique entre les rayons optiques venant de la paire de points gauche et

droite appariés.

4.5.2 Paramétrisation de la source de projection

Lors de la projection des franges verticales, le vidéo projecteur peut être considéré comme des sources ponctuelles de lumière adjacentes sur l'axe vertical. Une telle considération procure pour chaque ligne épipolaire une source ponctuelle de lumière O_{Prj} située sur le plan épipolaire correspondant. Le modèle 3D non-dense constitue une série de courbes verticales 3D obtenues par l'intersection des franges des images positives et négatives. Chaque courbe 3D décrit le profil d'une frange verticale projetée et distordue sur la surface faciale 3D.

Nous proposons d'estimer le plan 3D contenant chaque courbe 3D d'intersection de franges séparément en utilisant la technique $RANSAC$: $RANdom\ SAmple\ Consensus$. Il s'agit d'un algorithme non-déterministe dans le sens où il produit un résultat correct avec une probabilité qui augmente avec le nombre d'itérations. L'algorithme a été publié pour la première fois par [Fischler & Bolles 1981]. L'axe vertical des sources de lumières adjacentes est ensuite défini par l'intersection de tous les plans 3D calculés comme le décrit la figure 4.7. Cette estimation est effectuée soit en ligne ou hors-ligne à la différence des approches existantes d'analyse de déphasage qui calibrent le vidéoprojecteur hors-ligne et imposent ainsi une position figée du projecteur au moment de la numérisation.

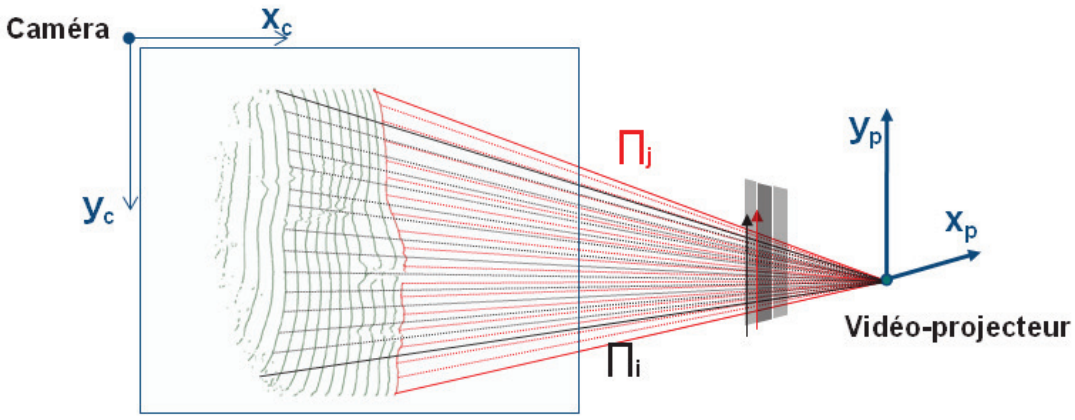


FIGURE 4.7 – Paramétrisation du vidéoprojecteur.

Les données d'entrée de l'algorithme $RANSAC$ sont les points 3D qui forment une

courbe d'intersection de franges, notre modèle paramétré constitue un plan 3D qui doit être ajusté à ce nuage de points observé. A l'entrée de l'algorithme, nous initialisons ce plan 3D en utilisant une approximation de son vecteur normal par la méthode des moindres carrées faisant intervenir tous les points d'intersection de franges qui forment les données observées. Nous utilisons la déviation spatiale moyenne entre le plan 3D et le nuage 3D d'intersection de franges ainsi que l'écart-type comme deux paramètres d'intervalle de confiance pour garantir la convergence de notre algorithme. $RANSAC$ atteint son objectif en sélectionnant itérativement un sous-ensemble aléatoire des données d'origine pour les considérer comme des points 3D aberrants et les négliger lors de l'estimation du plan 3D. A l'entrée de l'algorithme, nous considérons l'hypothèse que tous les points 3D d'intersection de franges constituent des points réguliers et cette hypothèse est ensuite testée comme suit :

- Ajustement du plan aux points réguliers 3D. Il s'agit d'estimer l'équation du plan en utilisant tous les points réguliers.
- Estimation de la déviation spatiale entre tous les autres points 3D et la nouvelle équation du plan pour mettre à jour la liste des candidats réguliers.
- Validation du plan estimé si suffisamment de points 3D ont été classés comme des candidats réguliers. L'équation du plan est ré-estimée en utilisant le nouvel ensemble de candidats réguliers.
- Evaluation de l'équation du plan par une mesure de la déviation spatiale des points réguliers par rapport au plan.

Cette procédure s'effectue itérativement et fournit des statistiques robustes des paramètres du plan. En effet, même si le nuage de points d'entrée comporte un taux important de points aberrants, il peut estimer les paramètres avec une bonne précision. Cependant, RANSAC nécessite plusieurs itérations pour converger vers la solution optimale.

4.5.3 Densification intrafrange

Ici, l'idée est de calculer dans un premier temps les coordonnées 3D des pixels situés à l'intérieur des franges pour la caméra droite et pour la caméra gauche séparément et de les fusionner ensuite pour obtenir le nuage 3D de points final. La première phase est basée sur l'estimation de l'information de profondeur d'un pixel à partir de décalage de phase.

Les techniques existantes d'analyse de déphasage estiment les phases locales dans $[0..2\pi]$ pour chaque pixel dans l'image capturée. Les phases absolues sont obtenues par déroulement de phase en déterminant la multiple intégrale inconnue de 2π à ajouter à chaque pixel pour créer une carte d'évolution de phase absolue et continue par rapport à un axe de référence qui a une phase absolue nulle. Selon l'approche que nous proposons, le modèle 3D non-dense remplace l'axe de référence et nous permet de retrouver l'information intra-frange à partir de la phase locale directement. En effet, chaque point P_i dans le modèle non-dense constitue un point de référence pour tous les pixels situés entre P_i et son voisin P_{i+1} sur la même ligne épipolaire. Pour chaque pixel P_k situé entre $P_i(X_i, Y_i, Z_i)$ et $P_{i+1}(X_{i+1}, Y_{i+1}, Z_{i+1})$, nous calculons sa phase locale ϕ_k en utilisant l'équation (4.2). La valeur de la phase de P_i est $\phi_i = 0$ et celle de P_{i+1} est $\phi_{i+1} = \pi$.

La phase ϕ_k qui appartient à $[0..\pi]$ a une variation monotone si $[P_iP_{i+1}]$ constitue une ligne doite dans le modèle 3D. Lorsque $[P_iP_{i+1}]$ représente une courbe dans le modèle 3D, la fonction ϕ_k décrit une variation de la profondeur dans $[P_iP_{i+1}]$. Ainsi, les coordonnées 3D $(X(\phi_k), Y(\phi_k), Z(\phi_k))$ du point 3D P_k qui correspond au pixel G_k sont calculées par une reconstruction géométrique comme l'illustre la figure 4.8.

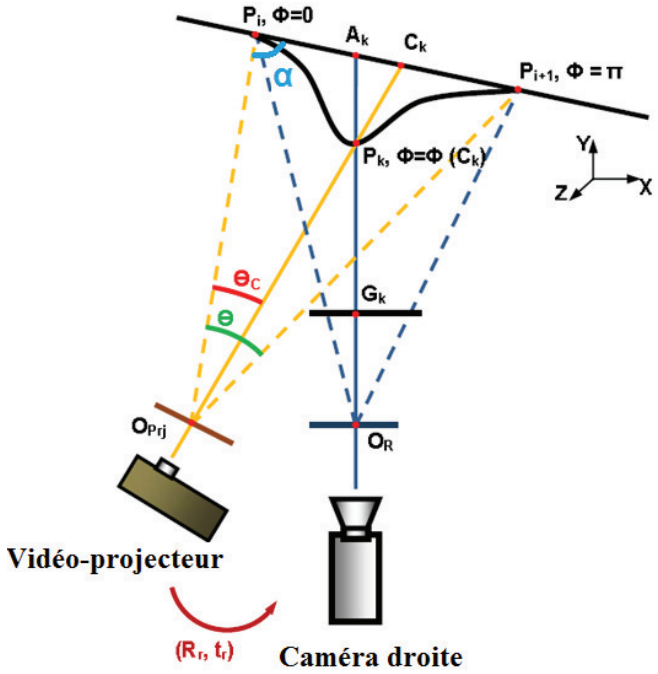


FIGURE 4.8 – Calcul de la profondeur pour les pixels situés à l'intérieur des franges.

Le calcul des coordonnées 3D des points situés dans l'intra-frange est effectué pour chaque ligne épipolaire i séparément. Un plan épipolaire est ainsi défini pour chaque ligne épipolaire. Il contient les deux centres optiques O_L et O_R des deux caméras gauche et droite respectivement et passe aussi par tous les points 3D situés sur la ligne épipolaire courante i. Chaque point 3D P_k se caractérise par sa propre valeur de phase $\phi(P_k)$. Le rayon de lumière venant de la source lumineuse vers le point 3D P_k, intersecte le segment $[P_iP_{i+1}]$ en un point 3D C_k ayant la même valeur de phase $\phi(C_k) = \phi(P_k)$ que P_k. Pour localiser C_k, nous avons besoin d'estimer la distance P_iC_k. Cette distance est calculée en appliquant la loi des sinus dans le triangle $(O_{Prj}P_iC_k)$ comme le décrit l'équation (4.3).

$$\frac{P_iC_k}{sin(\theta_C)} = \frac{O_{Prj}P_i}{sin(\pi - (\theta_C + \alpha))}. \tag{4.3}$$

La distance $O_{Prj}P_i$ et l'angle α entre $(O_{Prj}P_i)$ et (P_iP_{i+1}) sont connus. Aussi, l'angle θ entre $(O_{Prj}P_i)$ et $(O_{Prj}P_{i+1})$ est connu. Ainsi, l'angle θ_C est défini par l'équation (4.4). Après la localisation de C_k, le point 3D P_k est identifié comme l'intersection entre (O_RG_k) et $(O_{Prj}C_k)$.

$$\theta_C = \frac{\pi}{\theta}.\phi(C_k). \tag{4.4}$$

La figure 4.9 illustre l'estimation de la phase locale pour une ligne épipolaire de la vue d'un visage fournie par la caméra gauche. La courbe rouge représente la phase réelle estimée par notre modèle mathématique en utilisant l'équation (4.2). La courbe verte caractérise la variation de la phase locale pour une surface plate. Ainsi, la courbe bleue qui montre la différence en radium entre les deux courbes rouge et verte, traduit la variation de la profondeur que nous estimons par notre approche de conversion de phase en profondeur que nous venons de décrire.

La figure 4.10 affiche le résultat de la conversion de la phase locale en profondeur pour une ligne épipolaire de la vue d'un visage fournie par la caméra gauche. La courbe rouge caractérise la fonction sinusoïdale qui correspond à la phase réelle estimée par notre modèle mathématique en utilisant l'équation (4.2). La courbe verte définit la fonction sinusoïdale de la phase locale pour une surface plate et la courbe bleue montre la différence en radium entre les deux courbes rouge et verte. La courbe cyan représente la profondeur obtenue par

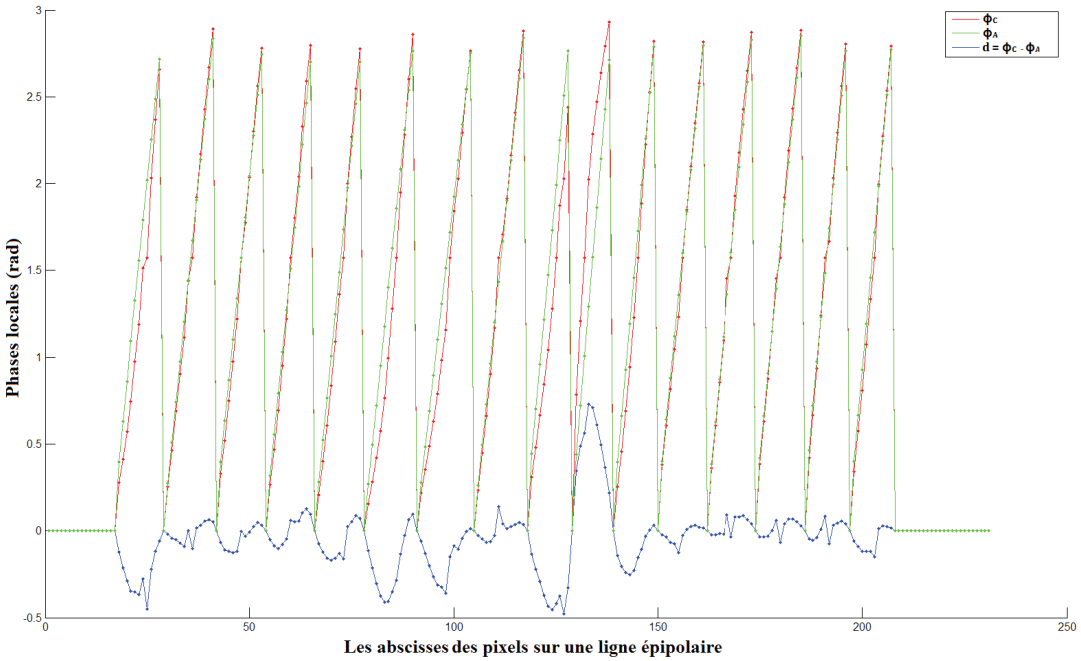


FIGURE 4.9 – Estimation de la phase locale pour une ligne épipolaire sur une image gauche d'un visage.

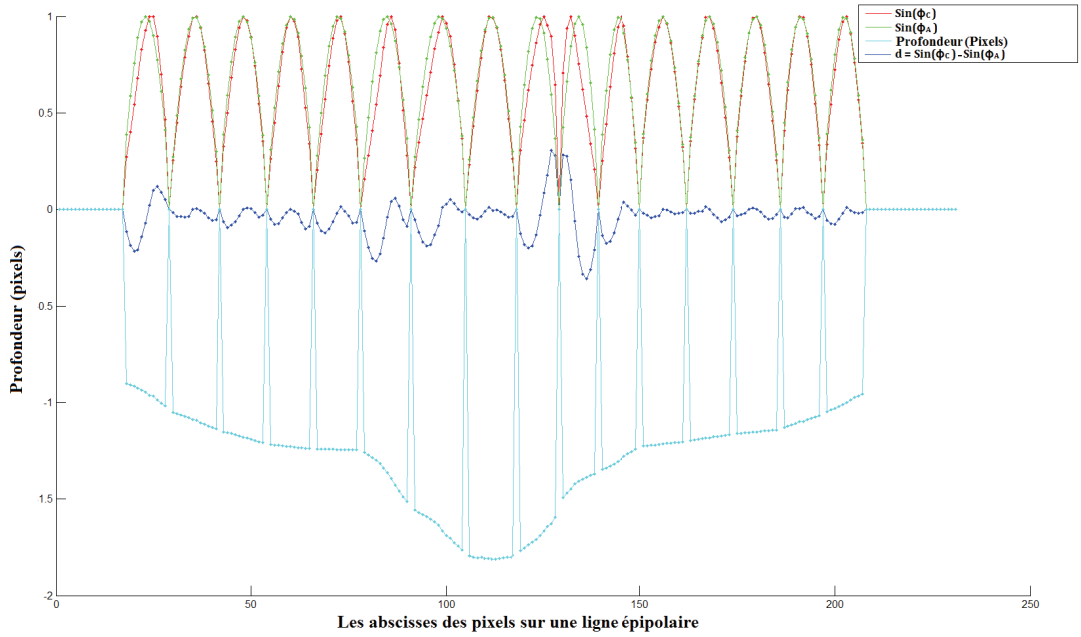


FIGURE 4.10 – Le résultat de la conversion de la phase en profondeur pour une ligne épipolaire.

notre technique de conversion de phase en profondeur.

Les densifications intra-franges gauche et droite fournissent deux nuages de points 3D gauche et droite naturellement recalés puisque leurs coordonnées 3D sont estimés en se basant sur le même modèle 3D non-dense obtenu par appariement stéréo. Aussi, les deux nuages de points 3D gauche et droite présentent une information 3D homogène et n'ont besoin que d'être fusionnés pour retrouver le nuage de points de haute résolution.

4.6 Etude expérimentale

Tout d'abord, nous présentons une étude expérimentale de notre approche de localisation de visages. Nous effectuons aussi une paramétrisation de la source de projection. Ensuite, les différentes étapes de l'élaboration d'un visage dense 3D sont illustrées. Nous proposons une étude comparative de notre technique avec la technique de balayage laser de MINOLTA VI300. Enfin, une évaluation de la performance de notre système est mise en œuvre par une étude de la précision et de la régularité de notre reconstruction 3D. Notre système stéréo comporte un banc stéréo formé par deux caméras réseaux $AXIS$ et un vidéoprojecteur LCD $3M\ MP7740i$. Nous utilisons d'abord un couple de caméras de faible résolution $AXIS\ 210$ de résolution 640x480 capable de capturer jusqu'à 30 images/secondes. Ensuite, nous utilisons un couple de caméras $AXIS\ 223M$ de résolution 1600x1200 capable de capturer jusqu'à 10 images/secondes.

4.6.1 Localisation de la région d'intérêt

Pour la numérisation 3D d'un visage, trois couples d'images gauche/droite sont capturées successivement. Elles comportent la distorsion des trois patrons envoyés par le vidéoprojecteur sur le visage. La figure 4.11 illustre les images d'un visage sans expression capturées par les caméras de faible résolution 640x480. Pour localiser le visage, nous suggérons une analyse fréquentielle 3D de la distorsion des patrons sinusoïdaux sur le visage comme le décrit la section 4.3. Le choix de la taille de la fenêtre glissante dépend de la résolution des caméras. Nous utilisons une fenêtre de taille $N = 64$ pour les caméras de faible résolution 640x480 et une fenêtre de taille $N = 128$ pour les caméras de résolution 1600x1200.

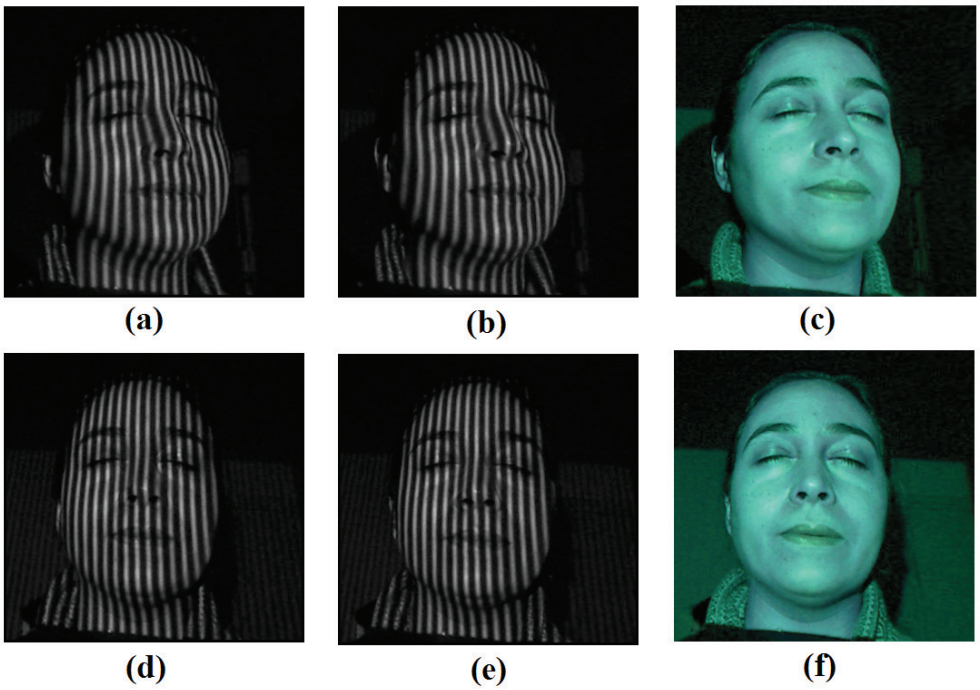


FIGURE 4.11 – Les images d'un visage sans expression capturées par les caméras de faible résolution.

Les deux surfaces 3D dans les figures 4.12.a et 4.12.d illustrent les deux profils 3D gauche et droite de la fonction de variation de l'amplitude de FFT en fonction de la fréquence, calculée sur la fenêtre glissante de taille $N = 64$ pixels, pour un visage de 350x300 pixels capturé par deux caméras de résolution 640x480. Les deux figures 4.12.b et 4.12.e présentent les deux représentations logarithmiques gauche et droite de cette variation d'amplitude. Enfin, les deux images 4.12.c et 4.12.f présentent les deux cartes 2D qui codent cette variation en niveaux de gris. La localisation du visage sur la vue gauche et droite est assurée en utilisant un seuil qui permet d'isoler la région faciale par un masque binaire. La figure 4.13 présente les résultats de la segmentation 2D assurée sur les deux vues gauches et droites en utilisant différents seuils. Nous choisissons la valeur de seuil 0.6 obtenue de manière empirique pour bien localiser la région faciale.

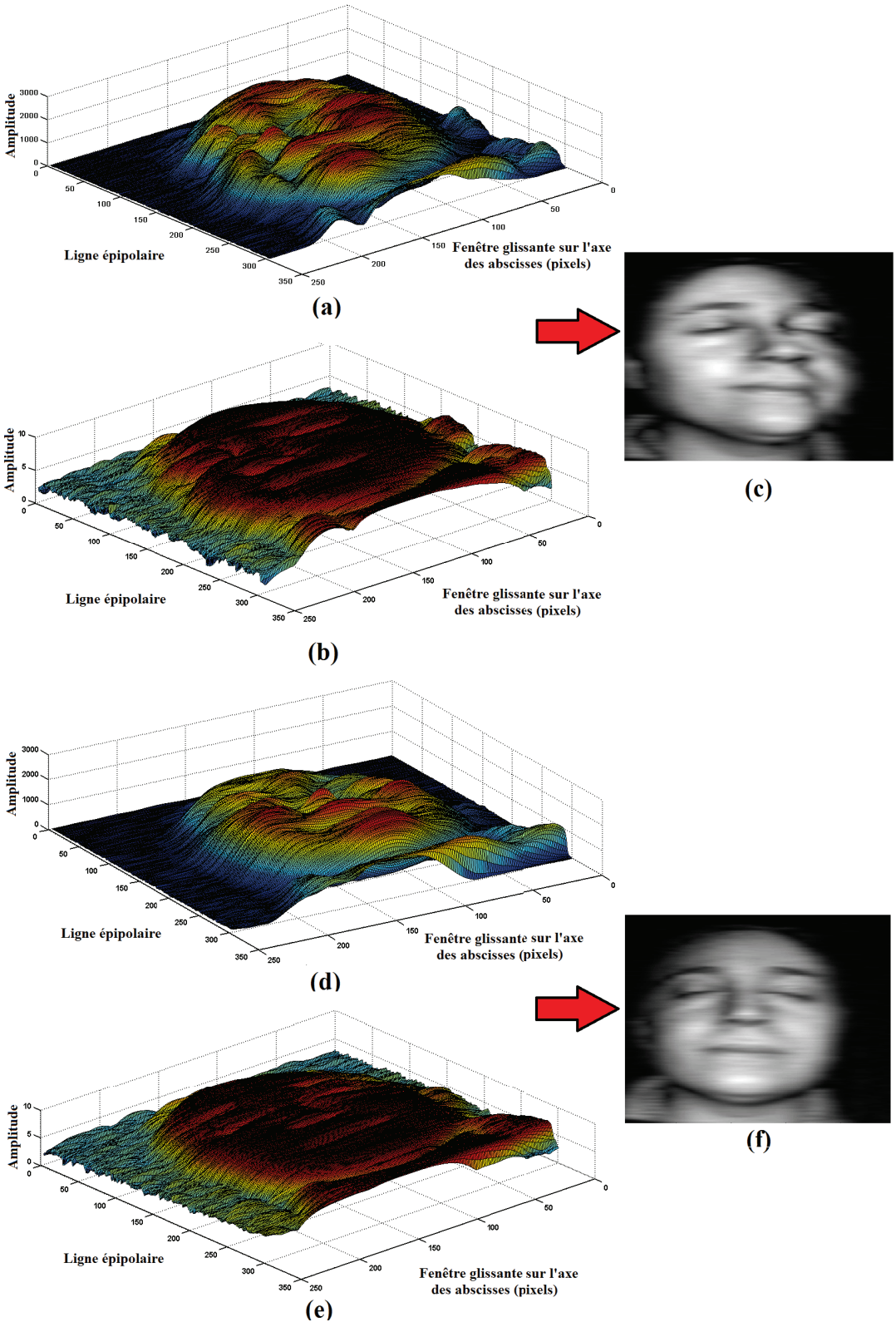


FIGURE 4.12 – Analyse fréquentielle 3D de la distorsion du patron sinusoïdal sur le visage.

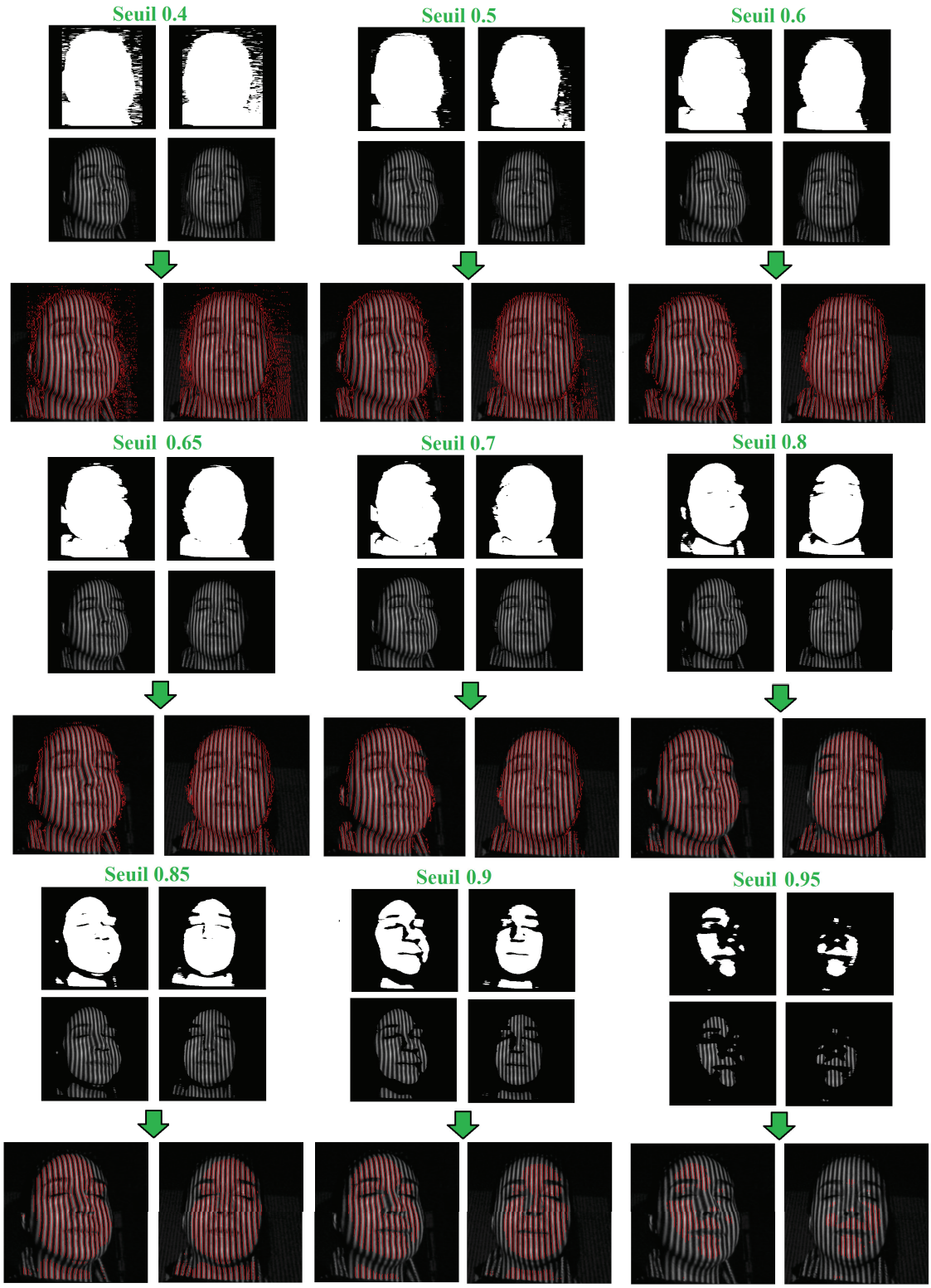


FIGURE 4.13 – Localisation de la région d'intérêt.

4.6.2 Paramétrisation de la source de projection

L'axe vertical du projecteur est défini par son vecteur directeur 3D $\vec{N}_{proj}$ et un point 3D P_{proj}. Ils sont calculés par une analyse de chaque plan 3D défini par tous les points 3D situés sur le profil vertical de chaque intersection de frange (section 4.5.2). Les équations des plans des différentes franges sont estimées par la méthode RANSAC. La figure 4.14 présente l'ensemble des points réguliers (colorés en vert) identifiés par la technique RANSAC et qui se situent sur les plans 3D des franges ainsi que les points aberrants (colorés en rouge) qui ont été écartés par l'algorithme RANSAC lors de l'estimation des équations des plans. Nous considérons 300 itérations pour la convergence de l'approche RANSAC. Sur la figure 4.14, le visage non-dense numérisé présente des erreurs de reconstruction au niveau de la bouche généré par un faux appariement stéréo. L'algorithme RANSAC réussit à identifier les points erronnés et de les écarter lors de l'estimation des équations des plans.

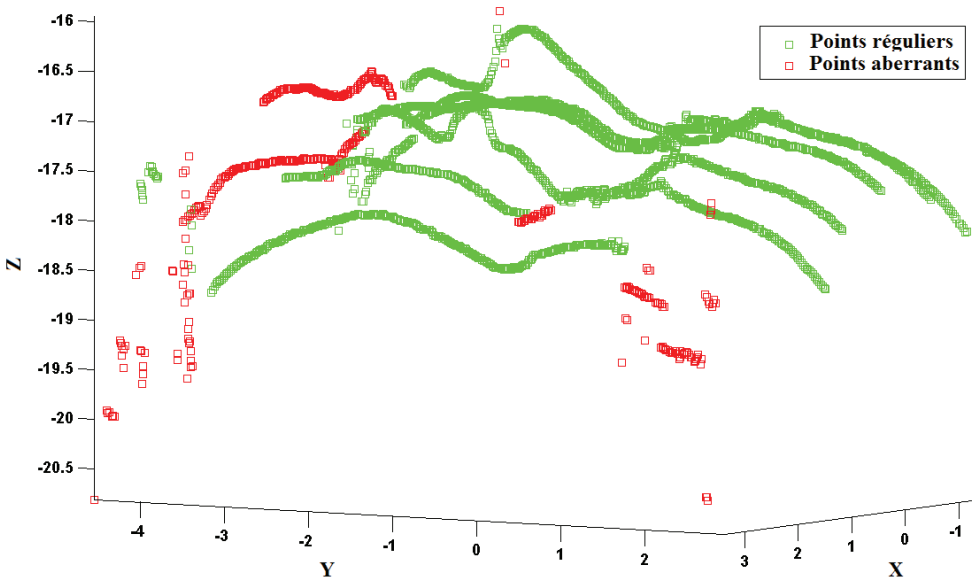


FIGURE 4.14 – Estimation des équations des différents plans 3D des franges verticales distordues sur le visage par la technique RANSAC.

Cette approche permet de localiser les plans 3D avec une précision de $0.002mm/pixel$. Le vecteur directeur $\vec{N}_{proj}$ est ensuite calculé comme étant le vecteur normal à tous les vecteurs normaux caractérisant les plans des franges. Aussi, le point P_{proj} se calcule comme étant l'intersection de tous les plans des franges. $\vec{N}_{proj}$ et P_{proj} sont estimés par une optimi-

sation au sens des moindres carrés. La déviation spatiale moyenne de $\vec{N}_{proj}$ par apport aux plans 3D des franges est de $0.003rad$. Le calcul de la profondeur à l'intérieur des franges utilise le centre du vidéoprojecteur O_{prj} et les deux points d'intersection de franges P_i et P_{i+1} comme le décrit la section 4.5.3. Ainsi, si O_{prj} est légèrement dévié de sa position réelle, les coordonnées 3D des points P situés entre P_i et P_{i+1} sont calculées avec une imprécision qui augmente inversement à la distance qui leur sépare de O_{prj}. Par conséquent, pour chaque ligne épipolaire, tous les points situés sur le segment $[P_iP_{i+1}]$ le plus proche de O_{prj} sont les points les plus aberrants. L'accumulation des erreurs de toutes les lignes épipolaires forme une frange de points 3D aberrants parallèle à l'axe du vidéoprojecteur et engendre une discontinuité visible et considérable sur la surface numérisée.

4.6.3 Numérisation 3D d'un visage

La numérisation d'un visage de 350x300 pixels, en utilisant des caméras 640x480, nécessite environ 0.6 secondes avec une implémentation en C++ .Net et un processeur INTEL Core2Duo (2.20Ghz) et une mémoire vive RAM de 2GB. L'étape de l'appariement stéréo s'effectue pour chaque ligne épipolaire puisque les deux plans images sont rectifiés comme l'illustre la figure 4.15. Le nombre de couples de primitives gauche/droite appariéss est de 4486. Ainsi, la triangulation optique permet de calculer un modèle 3D non-dense du visage formé de 4486 points.

La densification de ce modèle est assurée en utilisant tous les pixels gauches et droits. Ainsi, nous obtenons deux cartes de disparité gauche et droite qui apparaissent sur la figure 4.16. La figure 4.17.a montre le modèle non-dense calculé. La fusion des deux cartes de disparité permet d'obtenir un modèle 3D de haute résolution avec 148398 points qui s'affiche dans la figure 4.17.b. Enfin, la figure 4.17.c illustre le résultat du mappage de texture. La figure 4.18.a montre la densification du nuage 3D épars en utilisant uniquement les pixels gauches. La figure 4.18.a présente le résultat de la densification du nuage 3D épars en utilisant uniquement les pixels droits. La densification gauche et droite du nez s'affiche sur la figure 4.18.c.

Nous proposons aussi d'utiliser les deux caméras de haute résolution pour numériser le visage. En utilisant des caméras 1600x1200, la numérisation d'un visage de 610x570

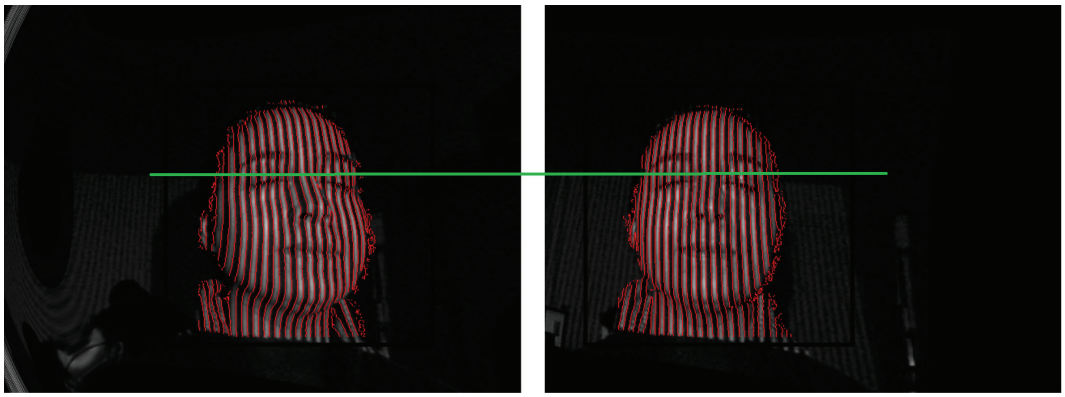


FIGURE 4.15 – Les deux vues gauche et droite à apparier pour les caméras de faible résolution.

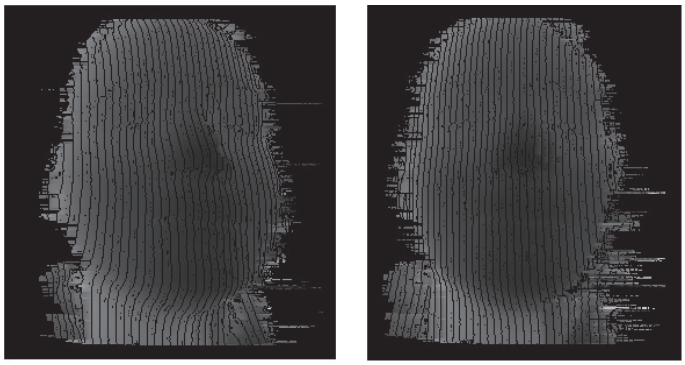
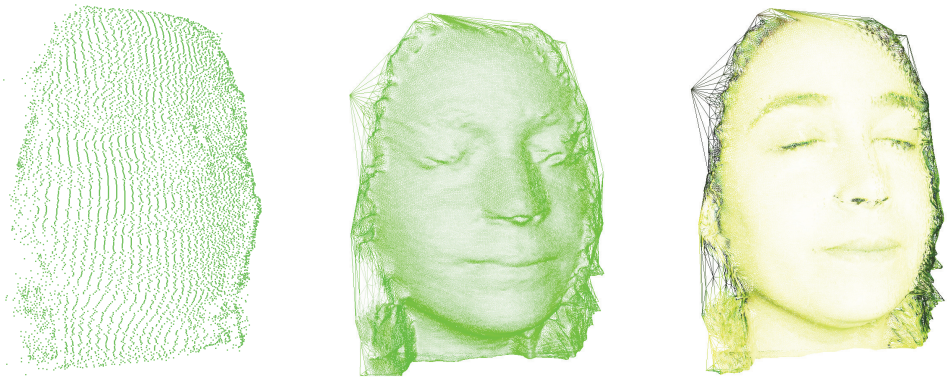


FIGURE 4.16 – Les deux cartes de disparité gauche et droite pour les caméras basse résolution 640x480.



(a) Modèle non-dense (b) Modèle dense (c) Modèle final texturé

FIGURE 4.17 – Numérisation dense 3D avec les caméras de faible résolution.

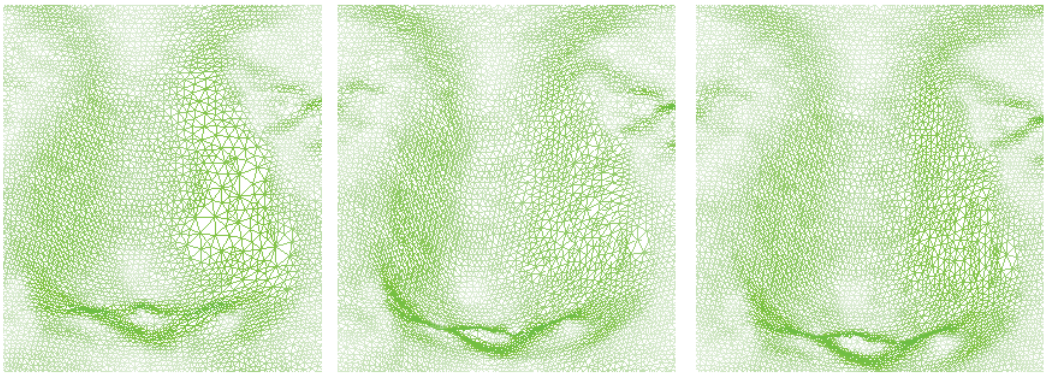


FIGURE 4.18 – Numérisation 3D du nez.

pixels nécessite environ 1.3 secondes. Les figures 4.19 et 4.20 présentent le résultat de la localisation du visage et de l'échantillonnage respectivement sur les deux vues gauche et droite. L'appariement stéréo s'effectue ensuite pour chaque ligne épipolaire puisque les deux plans images sont rectifiés comme le décrit la figure 4.21. Le nombre de couples gauche/droite appariés est de 8886. Ainsi, la triangulation optique permet de calculer un modèle 3D non-dense du visage de 8886 points. La densification de ce modèle s'effectue en utilisant tous les pixels gauches et droites. Ainsi, nous obtenons deux cartes de disparité gauche et droite présentées sur la figure 4.22. La figure 4.23.a présente le modèle non-dense calculé. La fusion des deux cartes de disparité permet d'obtenir un modèle 3D de haute résolution avec 243398 points qui s'affiche dans la figure 4.23.b. Enfin, la figure 4.23.c présente le résultat du mappage de texture.

4.6.4 Etude comparative avec une vérité terrain

Une estimation de l'erreur de reconstruction a été réalisée en comparant la qualité du modèle de visage reconstruit par notre technique et celui reconstruit par un scanner laser MINOLTA VI300. L'utilisation de l'algorithme ICP permet d'estimer une carte de déviation spatiale entre les deux modèles. Le modèle épars du visage contient 4486 points et sa densification fournit 148398 points. La précision de la reconstruction est ainsi estimée en utilisant un modèle 3D fourni par un scanner laser MINOLTA VI-300 comme vérité terrain. Le modèle laser obtenue par le scanner MINOLTA VI300 contient 11350 points.

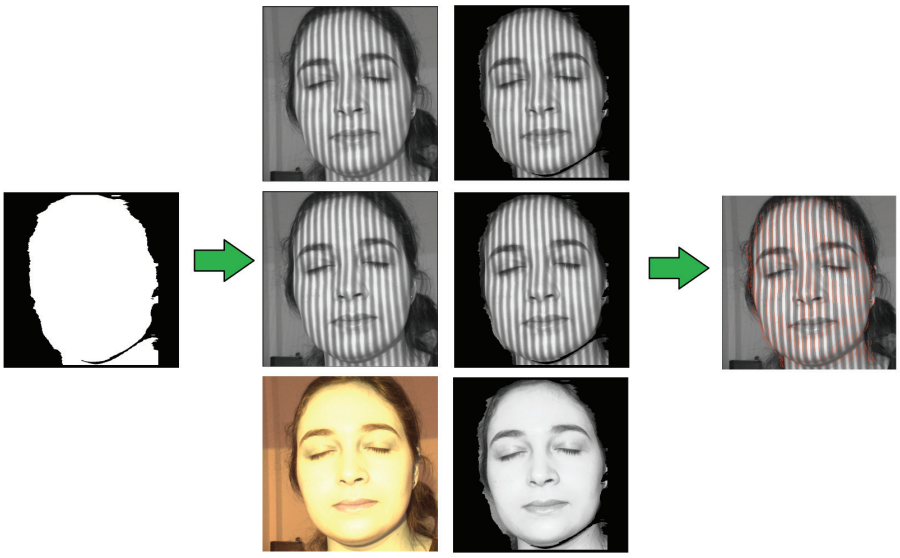


FIGURE 4.19 – Localisation et Echantillonnage du visage gauche avec une caméra de haute résolution 1600x1200.

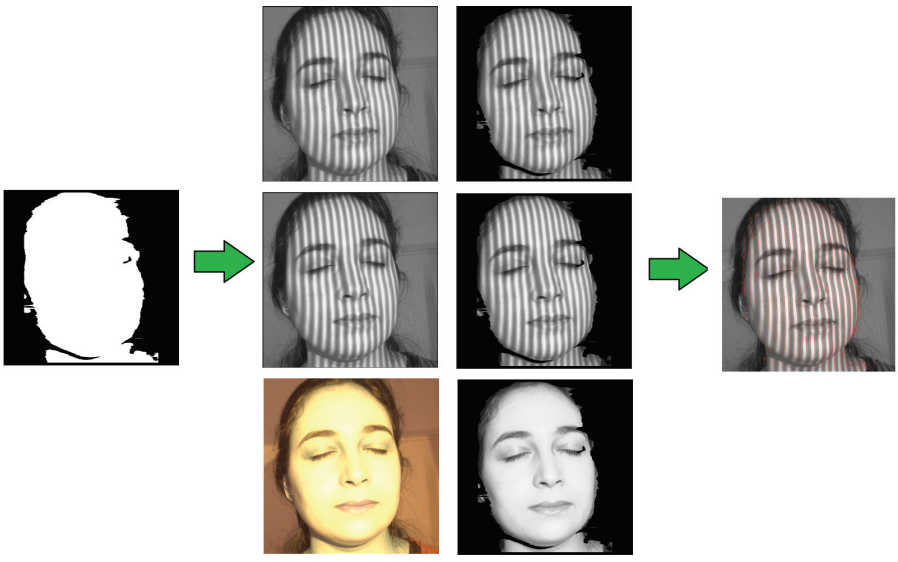


FIGURE 4.20 – Localisation et Echantillonnage du visage droite avec une caméra de haute résolution 1600x1200.

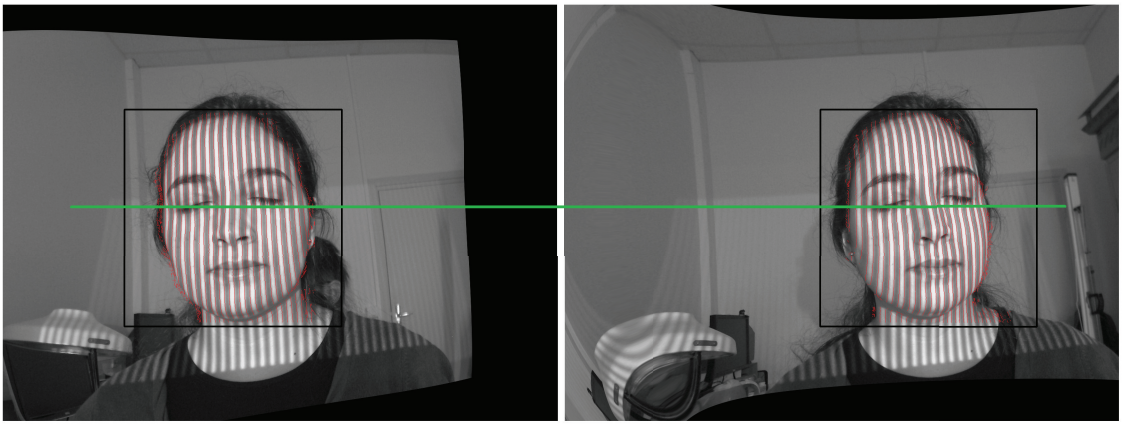


FIGURE 4.21 – Les deux vues gauche et droite à apparier capturées par les caméras de haute résolution 1600x1200.

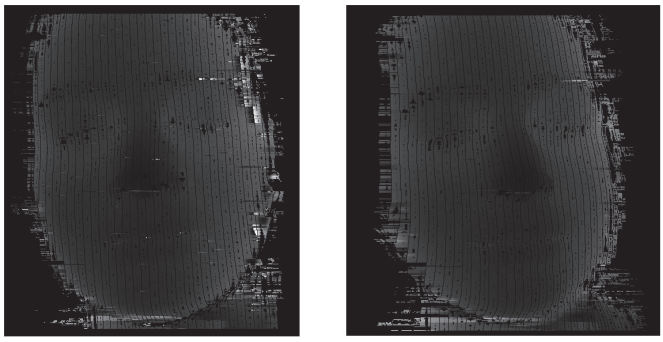


FIGURE 4.22 – Les deux cartes de disparité gauche et droite calculées.

(a) Modèle non-dense (b) Modèle dense (c) Modèle final texturé

FIGURE 4.23 – Numérisation dense 3D avec les caméras de haute résolution 1600x1200.

Nous utilisons une variante de l'algorithme ICP qui recale chaque point du premier modèle à la facette la plus proche du second modèle puisque les deux nuages de points ne sont pas de la même densité. Une déviation moyenne de $0.3146mm$ a été estimée. La déviation maximale est de $3.9922mm$ et la déviation minimale est nulle. L'écart-type est estimé à $0.3005mm$. La figure 4.24 montre la déviation spatiale estimée lors de la comparaison d'un modèle de visage reconstruit par notre technique avec celui calculé par un scanner laser MINOLTA VI300.

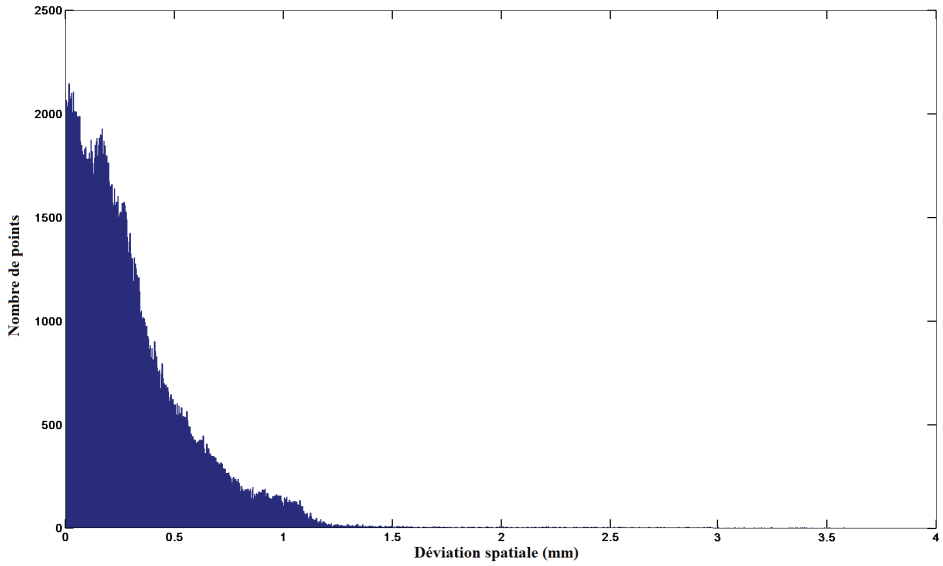


FIGURE 4.24 – Déviation spatiale estimée lors de la comparaison d'un modèle de visage reconstruit par notre technique avec celui calculé par un scanner laser MINOLTA VI300.

4.6.5 Evaluation de la performance de la numérisation

Nous proposons d'estimer la précision et la régularité de la reconstruction en utilisant un plan uniformément blanc comme objet de référence. L'évaluation de notre système de numérisation est réalisée en étudiant quantitativement et qualitativement la numérisation 3D du plan.

4.6.5.1 Précision du système

Il s'agit de numériser un plan et de calculer son équation théorique moyennant trois de ses points. Le modèle épars du plan contient 14344 points et la densification du modèle

épars fournit un nuage de point total de 180018 points. Nous mesurons pour chaque point du plan la distance qui le sépare du plan défini par l'équation théorique. Une déviation moyenne de $0.0092mm$ a été obtenue. Elle constitue l'erreur de précision moyenne. La déviation maximale est de $0.04mm$, celle minimale est de $3.4765e-015mm$. L'écart-type obtenu est de $0.0096mm$. La courbe qui représente la déviation spatiale se présente sur la figure 4.25.

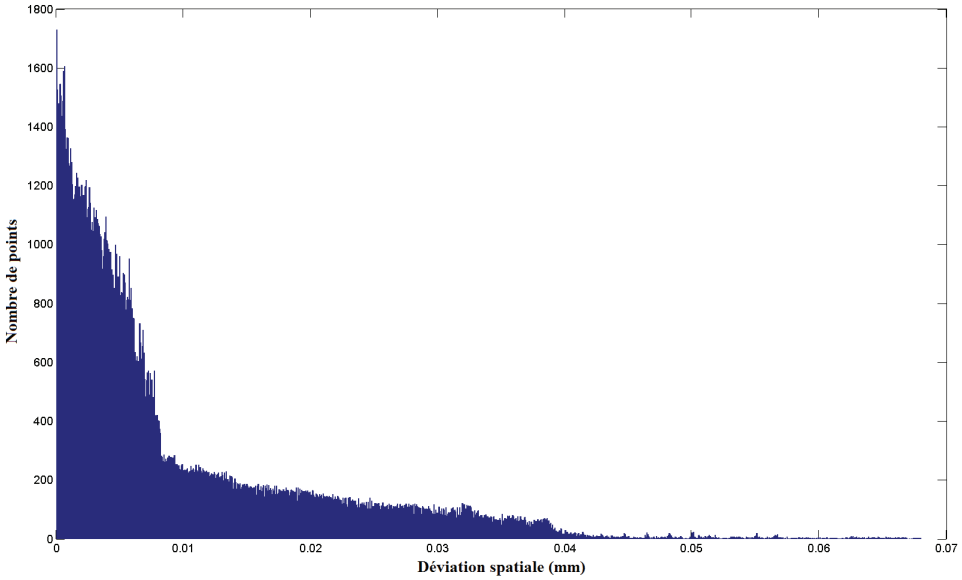


FIGURE 4.25 – Déviation spatiale estimée pour la mesure de la précision sur un plan.

4.6.5.2 Régularité de la reconstruction

Pour mesurer la régularité de la reconstruction, nous calculons une équation approximative du plan en utilisant la totalité des points par la méthode des moindres carrés. Ensuite, nous calculons la déviation spatiale entre le nuage de points du plan reconstruit et le plan défini par son équation approximative. L'erreur de régularité de la reconstruction du plan en question représente la déviation spatiale moyenne estimée à $0.0048mm$, l'écart-type obtenu est de $0.0041mm$, la valeur maximale est de $0.04mm$ et celle minimale est de $2.2158e-007mm$. La déviation spatiale obtenue s'affiche sur la figure 4.26.

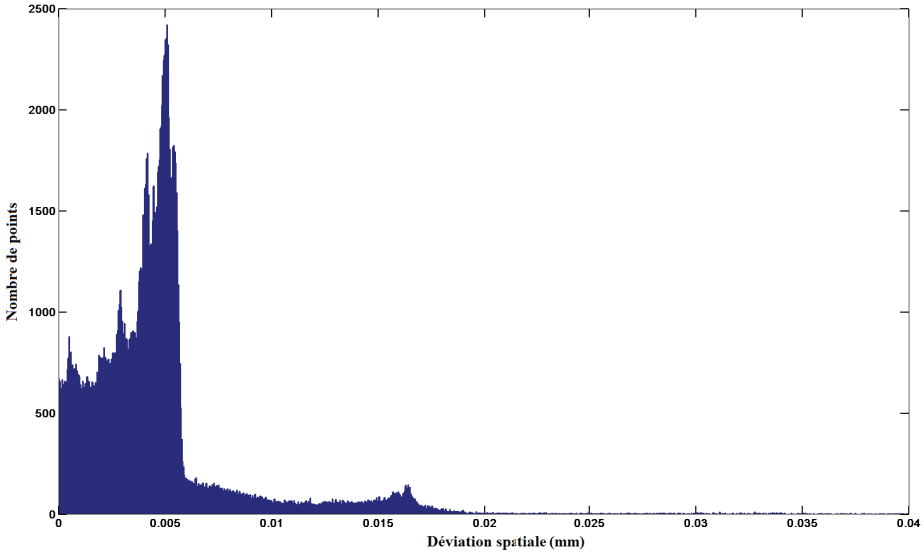


FIGURE 4.26 – Déviation spatiale estimée pour la mesure de la régularité sur un plan.

4.7 Conclusion

Dans ce chapitre, nous avons proposé de calculer la profondeur à l'intérieur des franges par une analyse de décalage de phase et une estimation en ligne des paramètres du vidéo-projecteur. Notre approche hybride de stéréovision et de codification sinusoïdale permet d'accorder plus de flexibilité au moment de la numérisation. En effet, puisque l'estimation des paramètres du vidéoprojecteur se fait en ligne, nous pouvons déplacer le vidéo-projecteur au cours de la numérisation pour couvrir plus de surface. La complexité de l'appariement stéréo, sur l'axe des ordonnées, dépend de la résolution pixélique des caméras. Cependant, sur l'axe des abscisses, cette complexité dépend uniquement du nombre de franges projetées sur la surface faciale. Ceci permet de numériser plus rapidement un visage animé même avec des caméras de très haute résolution.

Une numérisation 3D peut générer des erreurs causés par une occultation, une variation brusque de l'illumination ou par une variation rapide de la pose ou de l'expression de l'individu au cours de la projection des trois patrons nécessaires pour sa numérisation comme l'illustre la figure 4.27. En cas d'occultation, les erreurs générées sur le modèle non-dense, mènent à une estimation erronée de la phase. La figure 4.28 montre un exemple de numérisation erronée engendrée par une occultation au niveau de la bouche et son im-

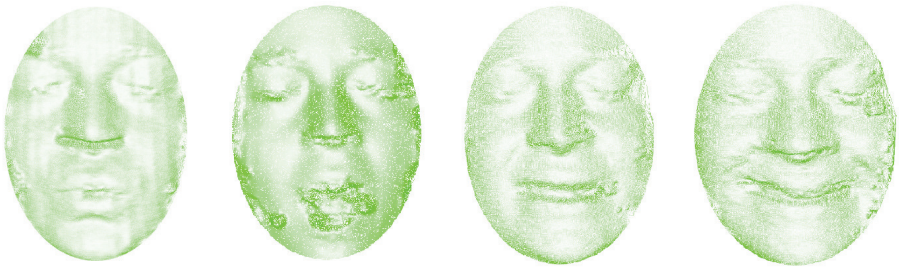


FIGURE 4.27 – Quelques erreurs de numérisation générées par notre système de numérisation 3D.

pact sur le modèle dense final. Nous proposons, dans le chapitre suivant, une approche de correction/super-résolution spatio-temporelle pour permettre une numérisation plus précise d'un visage animé par une déformation non-rigide.

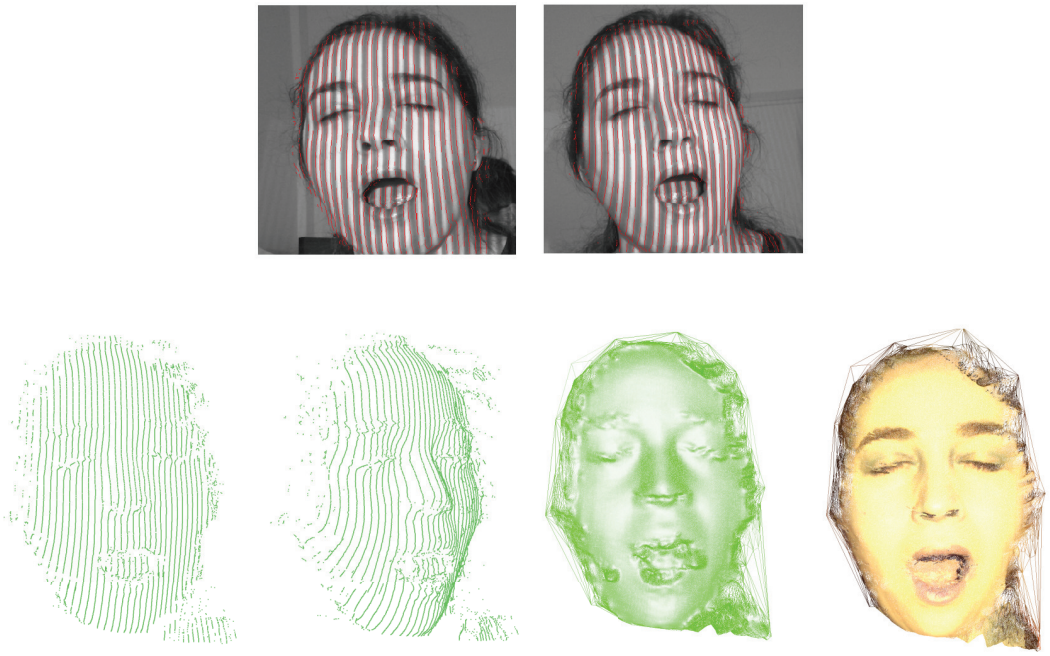


FIGURE 4.28 – Erreur de numérisation dans le cas d'occultation.

CHAPITRE 5

Super-résolution 3D Spatio-temporelle

5.1 Introduction

Toute technique de mesure 3D se caractérise par des erreurs de numérisation qui lui sont intrinsèques. On peut distinguer deux grandes familles : Les erreurs dites systématiques et les erreurs dites aléatoires. Les erreurs systématiques sont généralement celles qui posent le moins de problèmes à cause de leur caractère déterministe. Une opération d'étalonnage permet de les corriger. Les erreurs systématiques générées par notre technique sont d'une part la distorsion radiale et tangentielle propres aux caméras utilisées. D'autre part, la distorsion gamma engendrée par le vidéoprojecteur et les caméras. L'étalonnage stéréo et celui de la distorsion gamma nous permettent de les corriger.

Le caractère imprévisible des erreurs aléatoires se traduit par un écart entre le signal effectivement observé et le signal réel. Cet écart varie aléatoirement d'une numérisation à l'autre. Ainsi, le bruit aléatoire crée par notre technique de numérisation varie selon le degré d'occultation du visage sur les différentes vues capturées par les caméras, par exemple. En effet, ceci engendre de faux appariements stéréo et une estimation erronée du modèle non-dense de visage. Une erreur qui impacte la qualité du modèle dense final puisque l'estimation des paramètres du vidéoprojecteur ainsi que la densification se basent sur le modèle non-dense préalablement calculé. D'ailleurs, la projection successive des patrons de lumière sur un visage animé mène à une localisation imprécise des points d'intersection de franges. Nous obtenons ainsi une estimation biaisée de la disparité ce qui crée des ondulations verticales sur le visage numérisé. Enfin, une faible illumination de la scène diminue le contraste des franges sinusoïdales surtout sur les sourcils. Les points d'intersection de

franges situés sur les sourcils ne sont pas tous localisés ce qui cause de faux appariements et des points aberrants sur le visage 3D final.

Nous proposons de corriger ces erreurs aléatoires par une technique de super-résolution spatio-temporelle. La numérisation d'un visage en mouvement présente des déformations locales non rigides de la surface faciale ce qui rend l'application de la super-résolution temporelle classique insuffisante. Dans ce chapitre, nous suggérons de traiter l'aspect non-rigide de la déformation de la surface faciale.

5.2 Principe

La technique utilise N caméras étalonnées et un vidéoprojecteur non-étalonné et propose de traiter les caméras par couple. La figure 5.1 décrit les différentes étapes nécessaires pour la génération de séquences texturées d'un visage. Les étapes d'étalonnage du banc multi-caméra ainsi que l'étalonnage de la distorsion gamma sont effectués hors-ligne. La lumière structurée, que nous utilisons, comporte deux patrons sinusoïdaux et un dernier patron uniformément blanc. Leur projection successive permet de numériser un visage fixe. Ainsi, pour capturer un visage en mouvement, nous projetons cette lumière structurée en continu sur le visage et nous capturons au fur et à mesure les trames 2D correspondantes en utilisant le banc multi-caméra.

A un instant t, une trame 3D F_t du visage est calculée en utilisant les 3 trames 2D capturées aux instants $t-2$, $t-1$, et t, par chacune des N caméras. L'ensemble des 3xN trames correspondent à la distorsion des trois patrons sur le visage capturé. Pour obtenir F_t, nous estimons un nuage de points 3D dense du visage à partir de chaque couple de caméras par la technique de numérisation hybride proposée dans le chapitre précédent. Ainsi, nous calculons en premier lieu un nuage de points 3D non-dense du visage par stéréovision. Ensuite, une paramétrisation du vidéoprojecteur enligne permet une densification intra-frange en utilisant le modèle non-dense déjà calculé. Ceci permet d'estimer la profondeur de chaque pixel intra-frange de la caméra gauche et de celle droite séparément.

Chaque couple de caméras fournit une numérisation dense d'une vue 3D partielle du visage. La super-résolution spatiale consiste à fusionner d'abord les $N-1$ vues acquises à l'instant t pour former F_t. Ensuite, il s'agit d'utiliser la trame précédente F_{t-1} pour une

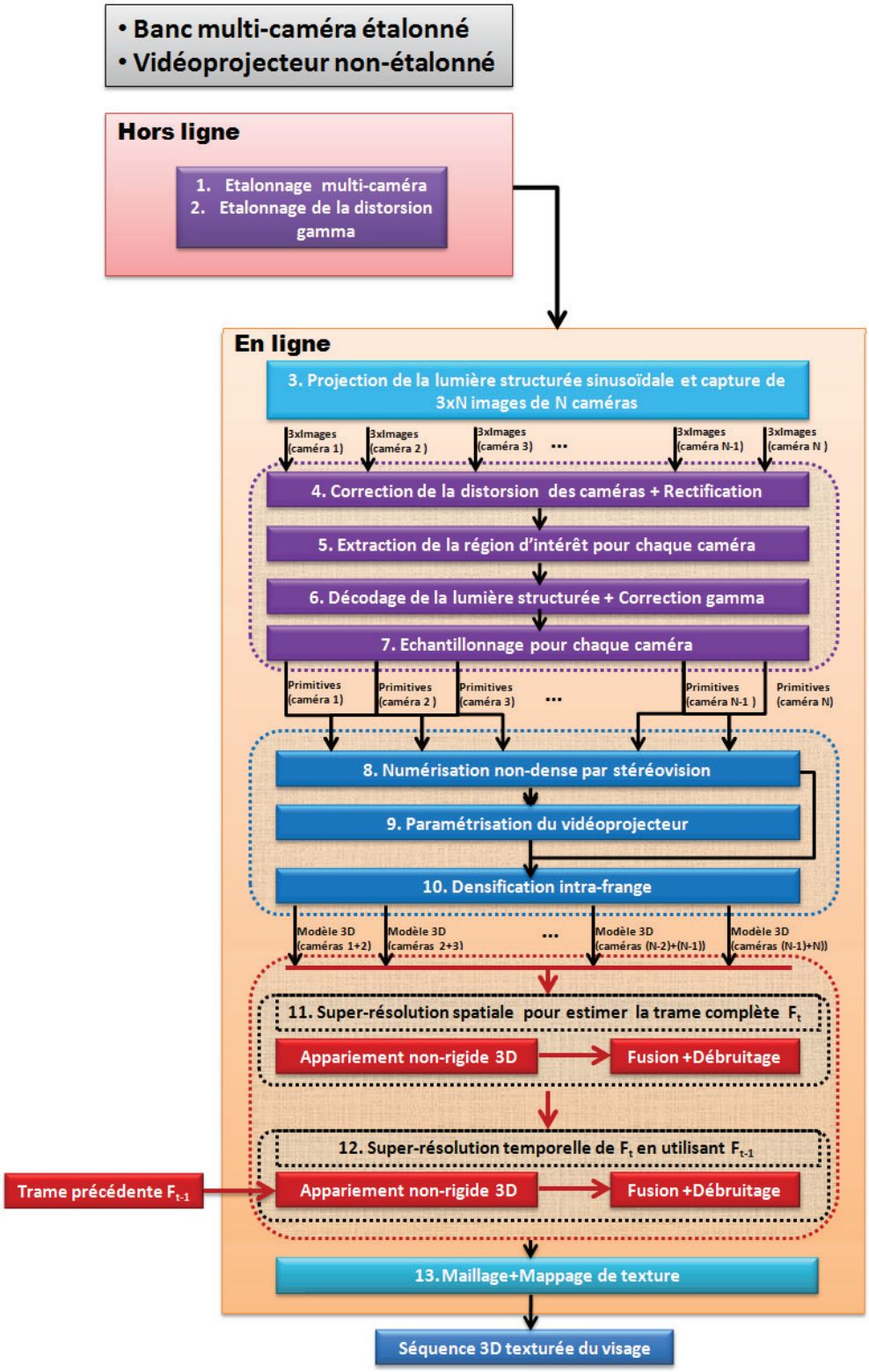


FIGURE 5.1 – Architecture détaillée du système multi-caméras proposé.

super-résolution/correction temporelle de F_t. La super-résolution spatio-temporelle nécessite une étape d'appariement suivie par une étape de fusion et de débruitage. Pour traiter l'aspect déformable de la surface faciale, nous suggérons une approche d'appariement 3D non-rigide. Le maillage et le plaquage de la texture permettent de rendre la trame 3D texturée finale du visage.

5.3 Super-résolution spatio-temporelle

La super-résolution spatiale permet la construction de la trame F_t par une fusion des $N-1$ vues 3D acquises à l'instant t. Ceci permet de compléter la vue 3D capturée, de corriger les occultations et de traiter les artéfacts générés par une réflection de la lumière sur la surface faciale, par exemple. Pour estimer F_t, nous utilisons une approche non-rigide d'appariement 3D pour recaler les $N-1$ vues 3D. Une étape de fusion des vues 3D appariées suivie d'une étape de débruitage permettent d'obtenir la trame complète F_t.

La super-résolution temporelle permet de corriger les modèles 3D de visages puisque la numérisation 3D peut engendrer des distorsions et des artéfacts causés essentiellement par des occultations, par une variation de pose, ou une expression faciale. Cette étape est appliquée pour chaque trame F_t à chaque instant t en utilisant sa trame précédente F_{t-1}. D'abord, nous assurons un appariement 3D non-rigide pour aligner le plus possible les deux trames 3D F_{t-1} et F_t en étant robuste à une éventuelle déformation non-rigide comme par exemple une variation de l'expression faciale. Une deuxième étape consiste à fusionner les deux visages appariés et à débruiter le résultat final.

5.4 Appariement 3D non-rigide

Le problème de l'appariement est formulé comme un problème probabiliste d'estimation du maximum de vraisemblance. La forme et la texture sont utilisées jointement pour caractériser les deux nuages de points 3D à apparier. Nous proposons un appariement non-rigide forme+texture basé sur l'algorithme CPD (Coherent Point Drift) développé par [Myronenko *et al.* 2007] pour aligner les deux nuages de points 3D source et destination en déformant itérativement le nuage source.

L'avantage essentiel de l'approche CPD est qu'elle permet un appariement dense de deux nuages de points 3D directement à la différence des approches classiques d'appariement dense non-rigide. Nous citons par exemple l'approche géodésique [Bronstein *et al.* 2006] et l'approche conformale [Wang *et al.* 2008]. D'abord, ces deux approches exigent une étape de maillage 3D pour apparier deux nuages de points 3D. En plus, elles transforment les deux maillages vers deux nouvelles représentations surfaciques dites isométriques sur l'hypothèse de la non-élasticité des surfaces à apparier. La technique $TPS - RPM$ (Thin Plate Spline - Robust non-rigid Point Matcher) proposée par [Chui & Rangarajan 2000] permet aussi un appariement non-rigide de deux nuages de points en paramétrisant la transformation non-rigide par la technique TPS [Bookstein 1989]. Cependant, une étude comparative réalisée par [Myronenko *et al.* 2007, Myronenko & Song 2010], montre que l'algorithme CPD est plus robuste que l'algorithme $TPS - RPM$ en présence de bruit et de points aberrants.

5.4.1 Modélisation du problème

Soient F_{src} et F_{dst} deux nuages de points texturés source et destination, avec $F_{src} = \{s_n | n = 1, ..., N_{src}\}$ et $F_{dst} = \{d_n | n = 1, ..., N_{dst}\}$. N_{src} constitue le nombre de points de F_{src} et N_{dst} est le nombre de points de F_{dst}. Chaque point texturé $M \in F_{src} \cup F_{dst}$ est un vecteur 1x6 qui concatène l'information de forme et de texture $M(XYZRGB)$. Cette approche suggère tout d'abord de représenter chaque point texturé s_n du nuage source F_{src} par un centroïde $Ctroid(s_n)$ d'un modèle de mélange gaussien GMM (Gaussian Mixture Model). Ce modèle statistique est défini par une densité mélange $p(x)$. Cette densité permet d'estimer la distribution de variables aléatoires en les modélisant comme une somme de plusieurs gaussiennes appelées centroïdes. Il s'agit alors de déterminer la variance, la moyenne et l'amplitude de chaque gaussienne. Ces paramètres sont optimisés selon un critère de maximum de vraisemblance pour approcher le plus possible la distribution recherchée.

L'appariement 3D entre les deux nuages, source et destination, se ramène ainsi à un alignement dense entre les centroïdes du modèle de mélange gaussien source et les points 3D du nuage destination F_{dst}. Pour créer le GMM pour F_{src}, une gaussienne multi-variée

est centrée en chaque point de F_{src}. Ainsi, tout le nuage de points F_{src} constitue un modèle de mélange gaussien qui se caractérise par une densité de probabilité mélange $p(x)$ comme le définit l'équation (5.1).

$$p(x) = \sum_{v=1}^{N_{src}+1} P(v)p(x|v), \quad p(x|v) = \frac{1}{(2\pi\sigma^2)^3} exp^{-\frac{\|x-s_v\|^2}{2\sigma^2}}. \tag{5.1}$$

Aussi, une distribution uniforme $p(x|N_{src}+1)$ est ajoutée au modèle de mélange gaussien pour tenir compte du bruit et des points aberrants, $p(x|N_{src}+1) = \frac{1}{N_{dst}}$. Nous utilisons des covariances isotropiques égales σ^2 et des probabilités d'adhésion égales $P(v) = \frac{1}{N_{src}}$ pour tous les centroïdes du GMM ($v = 1, ..., N_{src}$). En considérant le poids de la distribution uniforme w, $0 \leq w \leq 1$, le modèle de mélange gaussien prend la forme suivante :

$$p(x) = w\frac{1}{N_{dst}} + (1-w)\sum_{v=1}^{N_{src}} \frac{1}{N_{src}} p(x|v). \tag{5.2}$$

5.4.2 Principe de la déformation

Cette méthode consiste à rapprocher F_{src} et F_{dst} en déformant F_{src} itérativement et forcer les centroïdes du GMM à bouger en cohérence comme un groupe pour préserver la structure topologique du nuage de points F_{src} [Myronenko & Song 2010]. Les positions des centroïdes du GMM sont reparamétrisées par un ensemble de paramètres non-rigides θ. Nous les estimons par une maximisation de la vraisemblance ou, en équivalence, par une minimisation de la fonction négative $E(\theta, \sigma^2)$ du logarithme de la vraisemblance définie par l'équation (5.3).

$$E(\theta, \sigma^2) = -\sum_{u=1}^{N_{dst}} log \sum_{v=1}^{N_{src}+1} P(v)p(d_u|v). \tag{5.3}$$

La probabilité de la correspondance entre deux points s_v et d_u est définie par une probabilité à postériori du centroïde $Ctroid(s_v)$ du GMM étant donné le point : $P(v|d_u) = P(v)p(d_u|v)/p(d_u)$.

5.4.3 Estimation des probabilités à postériori

Nous utilisons l'algorithme d'espérance-maximisation EM(Expectation-Maximisation) [Dempster *et al.* 1977, Bishop 1995] pour estimer les probabilités à postériori. L'idée de l'algorithme EM est d'estimer les valeurs des paramètres θ^{old} et σ^{old} et d'utiliser ensuite le théorème de Bayes pour calculer les distributions des probabilités à postériori $P^{old}(v|d_u)$ des centroïdes du mélange. Ceci constitue l'étape d'espérance $E-step$ de l'algorithme. Les nouvelles valeurs des paramètres θ^{new} et σ^{new} sont ensuite calculées par une minimisation de l'espérance de la fonction négative du logarithme de la vraisemblance [Bishop 1995]. Cette étape est appelée maximisation ou $M-step$ de l'algorithme. La fonction Q , appelée la fonction objective, n'est autre que la dividende de la fonction négative du logarithme de vraisemblance((5.4)).

$$Q = -\sum_{u=1}^{N_{dst}} \sum_{v=1}^{N_{src}+1} P^{old}(v|d_u) log(P^{new}(v) P^{new}(d_u|v)). \tag{5.4}$$

L'algorithme EM alterne les étapes d'espérance $E-$ et de maximisation $M-steps$ jusqu'à la convergence. En ignorant les constantes qui sont indépendantes de θ et σ^2, nous réécrivons l'équation (5.4) comme l'équation (5.5) où $N_{dst,p} = \sum_{u=1}^{N_{dst}} \sum_{v=1}^{N_{src}} P^{old}(v|d_u) \leq N_{src}$ (avec $N_{dst} = N_{dst,p}$ si seulement $w = 0$). $\tau(s_v, \theta)$ définit la transformation τ appliquée à s_v en considérant l'ensemble θ des paramètres de la transformation non-rigide.

$$\begin{aligned} Q(\theta, \sigma^2) = & -\tfrac{1}{\sigma^2} \textstyle\sum_{u=1}^{N_{dst}} \sum_{v=1}^{N_{src}} P^{old}(v|d_u) \|d_u - \tau(s_v, \theta)\|^2 \\ & + \tfrac{3N_{dst,p}}{2} log(\sigma^2). \end{aligned} \tag{5.5}$$

P^{old} définit les probabilités à postériori des composants du GMM calculées en utilisant les anciennes valeurs des paramètres comme le décrit l'équation (5.6).

$$P^{old}(v|d_u) = \frac{exp^{-\frac{1}{2}\| \frac{d_u - \tau(s_v, \theta^{old})}{\sigma^{old}} \|^2}}{\sum_{k=1}^{N_{src}} exp^{-\frac{1}{2}\| \frac{d_u - \tau(s_k, \theta^{old})}{\sigma^{old}} \|^2} + C}. \tag{5.6}$$

avec $C = (2\pi\sigma^2)^3 \frac{w}{1-w} \frac{N_{src}}{N_{dst}}$. En minimisant la fonction Q, nous réduisons nécessaire-

ment la fonction E du logarithme de vraisemblance sauf si elle est déjà dans un minimum local. Pour résoudre l'aspect non-rigide des déformations, une régularisation de Tikhonov est utilisée [Tikhonov & Arsenin 1977, Myronenko & Song 2010]. La transformation τ est définie comme étant la position initiale ajoutée à une fonction de déplacement V, $\tau(F_{src}, V) = F_{src} + V$. La fonction de déplacement V est estimée en utilisant un calcul variationel et la norme de V est régularisée pour renforcer la continuité de la déformation [Myronenko & Song 2010].

5.5 Fusion et débruitage

Nous assurons la fusion des deux modèles appariés F_{src} et F_{dst} et l'étape de débruitage par la résolution d'un problème d'optimisation. La représentation d'un modèle 3D de visage par une image 2D de profondeur cause une perte de précision puisque les points 3D que nous obtenons sont d'une précision sous-pixélique. Aussi, les pixels de chaque caméra participent séparément dans le modèle 3D puisque leur numérisation utilise seulement leurs phases. Ainsi, nous représentons chaque modèle 3D par 4 cartes 2D définies par les coordonnées X, Y, Z et la texture de ses points. Pour chaque type de carte 2D, nous définissons une fonction d'optimisation qui utilise F_{src} et F_{dst} comme des données de faible résolution pour estimer une carte 2D de haute résolution de F_{dst}.

5.5.1 Modélisation du problème

Nous proposons une approche d'optimisation basée sur la technique de super-résolution 2D proposée dans [Schuon *et al.* 2009, Farsiu *et al.* 2004]. La fonction d'énergie à minimiser emploie conjointement deux termes $E_{data}(H)$ et $E_{regular}(H)$ comme le définit l'équation (5.7).

$$minimize \qquad E_{data}(H) \quad + \quad \alpha \quad E_{regular}(H). \tag{5.7}$$

Le premier terme $E_{data}(H)$ mesure l'accord de la reconstruction H avec les nuages de points 3D de faible résolution alignés. $E_{regular}(H)$ est un terme d'énergie de régularisation qui guide l'algorithme d'optimisation vers une reconstruction plausible H. Le terme α

$(0 \leq \alpha \leq 1)$ caractérise le taux de la régularisation dans la fonction d'énergie à minimiser.

5.5.2 Le terme de données

Le premier terme $E_{data}(H)$ est défini par l'équation (5.8). Il mesure la fidélité et la cohérence de la reconstruction H de haute résolution à estimer avec les données de faible résolution de départ. N constitue le nombre de cartes 2D de faible résolution. Les cartes 2D I_k représentent X, Y, Z ou texture des modèles 3D de faible résolution déjà recalés par le CPD.

$$E_{data}(H) = \sum_{k=1}^{N} \|W_k . * G_k . * (I_k - H)\|_2. \tag{5.8}$$

L'opérateur $.*$ constitue un opérateur de multiplication élément par élément. W_k est une matrice bande qui encode les positions de I_k affectées lors du rééchantillonnage sur la grille cible H à haute résolution. G_k est une matrice diagonale contenant des entrées nulles pour tous les échantillons de I_k qui ont été estimés non fiables selon le résultat de l'algorithme CPD d'alignement non-rigide. En effet, l'appariement non-rigide fournit une liste de correspondance dense entre les deux modèles appariés qui caractérise la déviation spatiale obtenue pour chaque couple de points 3D correspondants. Nous utilisons un seuil et nous affectons la valeur 0 dans la matrice G_k pour chaque couple de points 3D correspondants ayant une déviation spatiale supérieure à la valeur du seuil.

5.5.3 Le terme de régularisation

Le terme $E_{regular}(H)$ définit une énergie de régularisation nécessaire dans le processus d'optimisation pour converger vers une solution admissible H. Il est défini comme une nouvelle somme de normes adaptée pour reconstruire une solution de haute résolution crédible comme le décrit l'équation (5.9).

$$E_{regular}(H) = \sum_{u,v} \|\nabla H_{u,v}\|_2. \tag{5.9}$$

La quantité $\nabla H_{u,v}$ constitue un vecteur rassemblant des approximations du gradient spatial de différences finies sur différentes échelles (l, m) pour une position pixélique don-

née (u, v) comme le décrit l'équation (5.10).

$$\nabla H_{u,v} = \begin{pmatrix} Q_{u,v}(0,1) \\ Q_{u,v}(1,0) \\ . \\ . \\ . \\ Q_{u,v}(l,m) \end{pmatrix}. \tag{5.10}$$

$Q_{u,v}(l, m)$ constitue une différence finie définie par l'équation (5.11).

$$Q_{u,v}(l,m) = \frac{H_{u,v} - H_{u+l,v+m}}{\sqrt{l^2 + m^2}}. \tag{5.11}$$

5.6 Etude expérimentale

Nous commençons par une évaluation de l'appariement non-rigide de deux nuages de points 3D F_{src} et F_{dst}. Nous étudions ensuite la consistance physique de la déformation assurée par l'appariement. Puis, nous présentons les résultats de la numérisation d'un visage 3D en mouvement. En un premier temps, nous utilisons un système bi-caméra. Dans ce cas, nous appliquons une super-résolution temporelle uniquement. Elle s'applique entre les deux modèles 3D F_{t-1} et F_t calculés respectivement à l'instant $t-1$ et t. Dans un deuxième temps, un système tri-caméra est utilisé. Dans ce cas, deux vues 3D sont générées à un instant t. La première vue est calculée à partir des deux caméras gauche et centrale, l'autre est générée par les deux caméras centrale et droite. D'abord, les deux vues sont fusionnées par une super-résolution spatiale pour trouver F_t. Une super-résolution/correction temporelle de F_t est ensuite assurée en utilisant F_{t-1}. Finalement, une évaluation de la super-résolution/correction assurée est décrite.

Notre approche d'appariement non-rigide emploie un paramètre w,($0 \leq w \leq 1$) qui reflète le taux de bruit qui caractérise les nuages de points à apparier. La fonction de déplacement V définie dans la section 5.4.3 implique deux paramètres λ et β définis dans [Myronenko & Song 2010] pour renforcer la continuité de la déformation. Dans nos expérimentations, nous utilisons $w = 0.3$, $\lambda = 2$ et $\beta = 3$. Pour la convergence de l'algorithme

d'appariement, 30 itérations sont utilisées. Pour la fusion et le débruitage, le poids de la régularisation α défini dans l'équation (5.7) est $\alpha = 0.25$.

5.6.1 Appariement non-rigide 3D

L'appariement de deux nuages de points 3D F_{src} et F_{dst} fournit une liste de correspondance qui réunit chaque point 3D de F_{dst} avec le point 3D le plus proche de F_{src} recalé. Les deux points sont dits homologues. Nous proposons d'utiliser, comme critère d'évaluation, la déviation spatiale qui sépare F_{src} et F_{dst} suite à leur appariement. La déviation spatiale entre les deux nuages est représentée par une carte de couleur que nous mappons sur F_{dst}. A chaque point 3D de F_{dst}, la carte de couleur associe une couleur qui décrypte sa déviation spatiale avec son homologue de F_{src}. Les figures 5.2 et 5.3 montrent respectivement le résultat de l'appariement rigide et celui de l'appariement non-rigide entre F_{src} et F_{dst}. Les figures 5.4 et 5.5 illustrent les deux cartes de couleur de la déviation spatiale entre F_{src} et F_{dst} suite à leur appariement rigide et non-rigide. Les distributions de la déviation spatiale rigide et non-rigide sont illustrées par les figures 5.6 et 5.7.

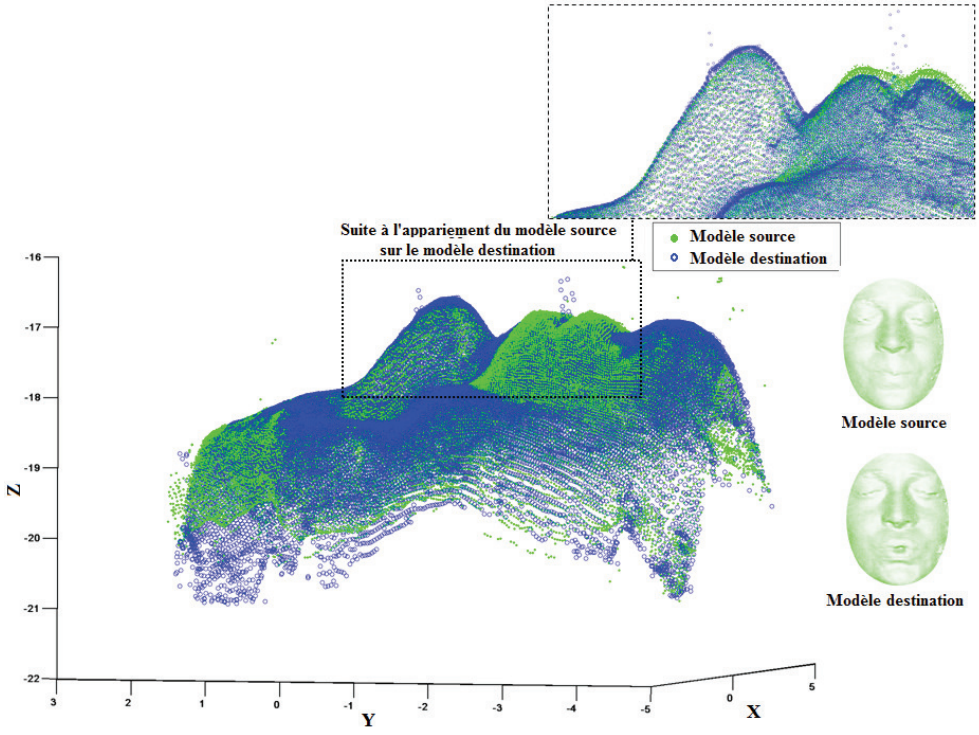


FIGURE 5.2 – Le résultat de l'appariement 3D rigide entre F_{src} et F_{dst}.

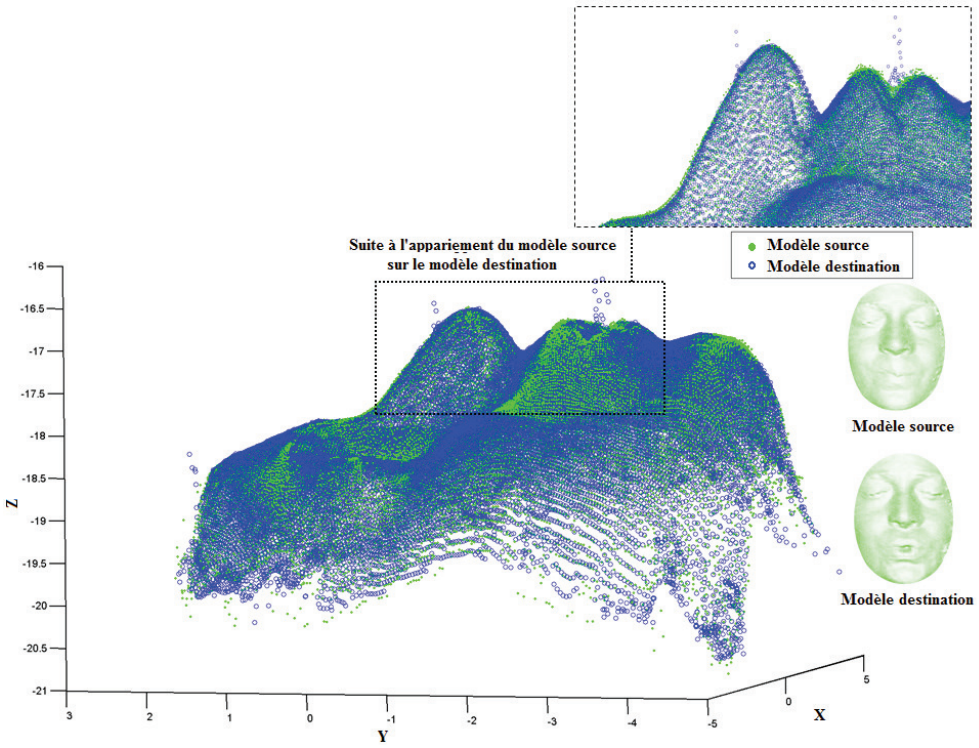


FIGURE 5.3 – Le résultat de l'appariement 3D non-rigide entre F_{src} et F_{dst}.

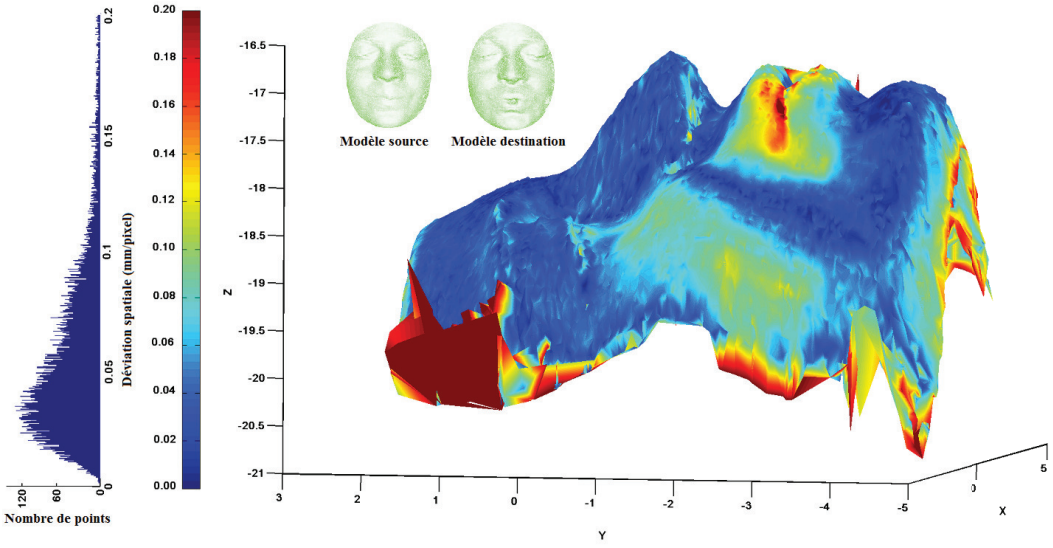


FIGURE 5.4 – Carte de la déviation spatiale entre F_{src} et F_{dst} suite à leur appariement rigide.

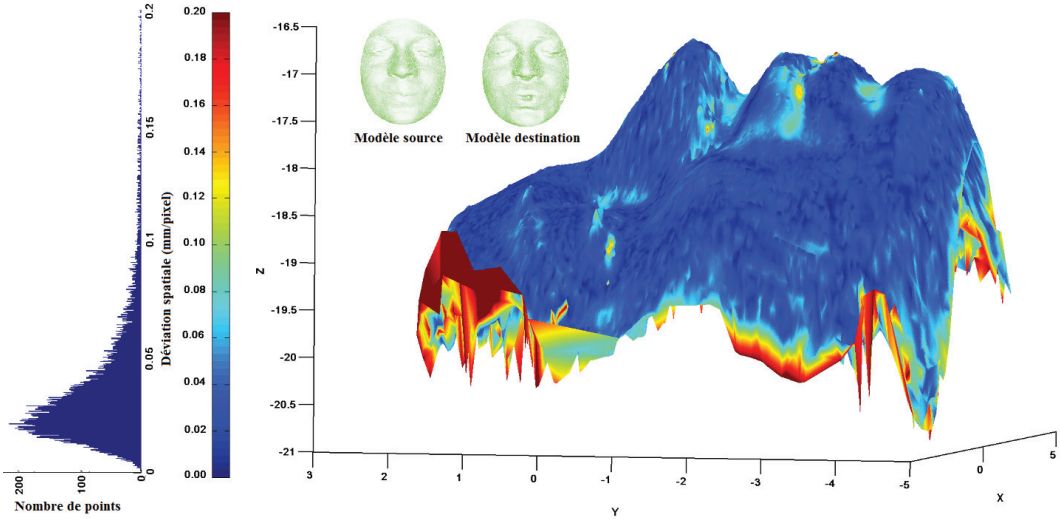


FIGURE 5.5 – Carte de la déviation spatiale entre F_{src} et F_{dst} suite à leur appariement non-rigide.

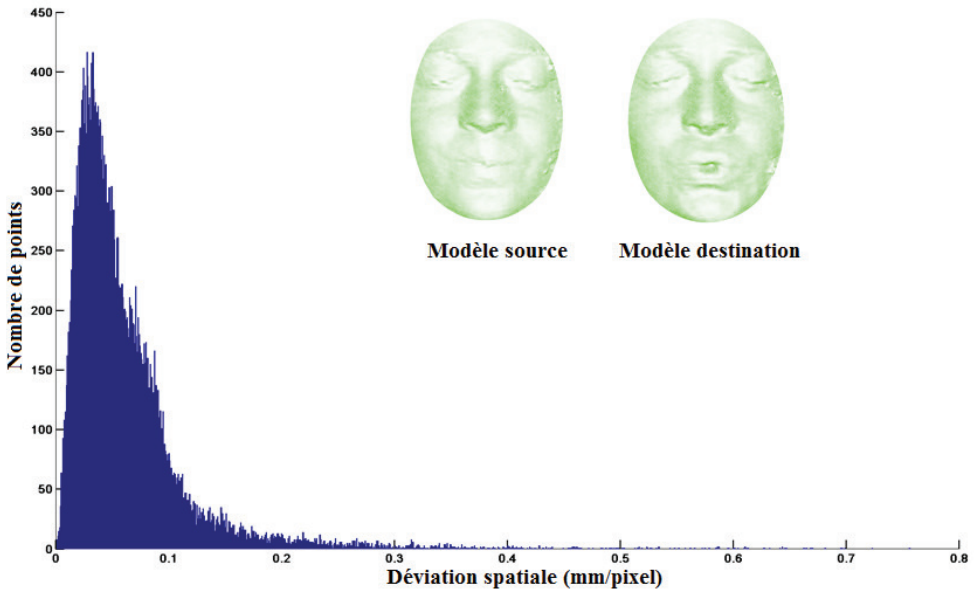


FIGURE 5.6 – Distribution de la déviation spatiale entre F_{src} et F_{dst} suite à leur appariement rigide.

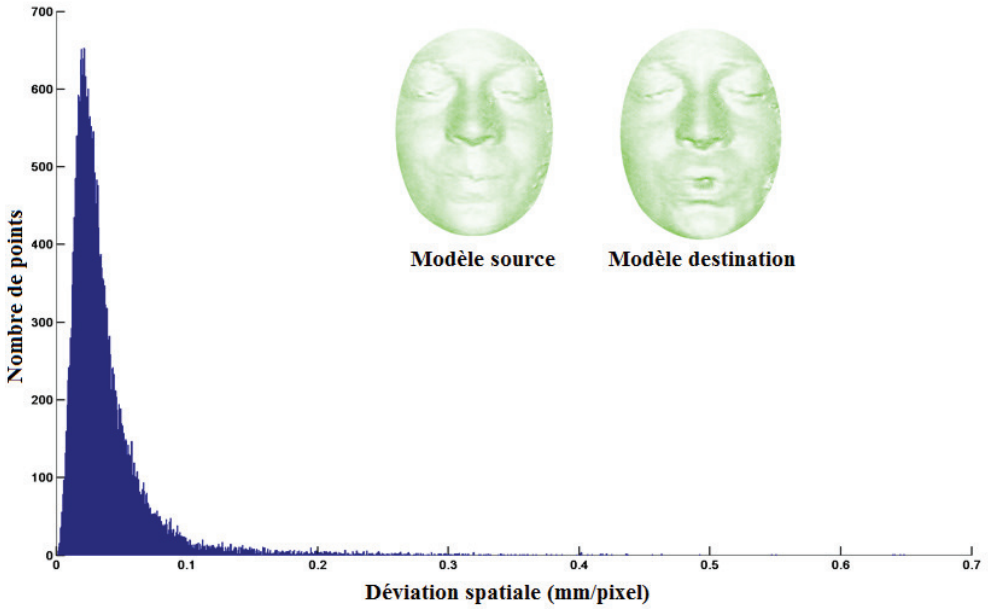


FIGURE 5.7 – Distribution de la déviation spatiale entre F_{t-1} et F_t suite à leur appariement non-rigide.

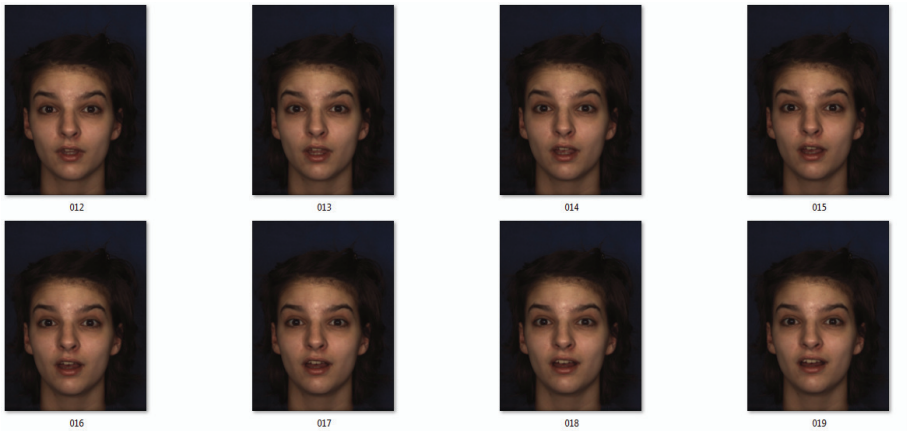
La variante non-rigide de l'algorithme CPD fournit une déviation spatiale moyenne de $0.0387mm/pixel$ et un écart-type de $0.0371mm/pixel$ et la variante rigide fournit une déviation spatiale moyenne de $0.0616mm/pixel$ avec un écart-type $0.0566mm/pixel$. Ainsi, l'appariement non-rigide permet de rapprocher F_{src} et F_{dst} même en présence d'une déformation non-rigide comme l'illustre la carte de couleur correspondante sur la figure 5.5 à la différence de l'appariement rigide.

5.6.2 Consistance de la déformation

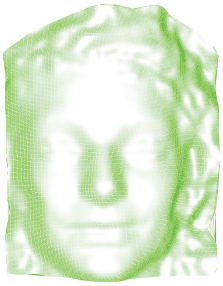
Nous suggérons de valider la consistance physique de l'appariement non-rigide dans le sens que chaque point de F_{dst} définit-il ou non le même point physique que son homologue sur F_{src} apparié. Pour ce faire, nous proposons d'utiliser une base standard de séquences 3D texturées de visages surtout parce que notre système actuel de numérisation n'a pas été conçu pour l'élaboration d'une base de vidéos 3D de visages. Nous utilisons la base BU de vidéos 3D qui comporte 101 sujets avec 58 femmes et 43 hommes [Yin *et al.* 2008, Yin *et al.* 2006].

La base se caractérise par une variété ethnique de 28 asiatiques, 8 noirs, 3 d'origine

latine et 62 blancs avec des âges entre 18 et 45 ans. Elle dispose de six types d'expressions faciales pour chaque sujet et d'une vidéo 3D texturée pour chaque expression faciale et chaque sujet. Les expressions sont classées sous six catégories : la joie, la tristesse, la peur, le dégoût, la colère et la surprise. Ainsi, la base contient 606 séquences vidéo texturées de format AVI. Une séquence vidéo 3D est d'une résolution approximative de 35000 points et la séquence de texture a une résolution de 1040x1329 pixels par trame. La figure 5.8 présente quelques trames d'un visage de la base BU.



(a) une séquence vidéo d'un visage ayant une expression de surprise



(b) Forme

(c) Texture

FIGURE 5.8 – La base BU de séquences 3D texturée de visages.

Pour valider la correspondance physique entre deux trames successives F_t et F_{t-1} après leur appariement non-rigide, nous localisons manuellement un ensemble de points ancres sur le visage 3D F_t. La trame F_{t-1} est considérée comme un modèle source et la trame F_t comme un modèle destination. Ainsi, le processus de l'appariement consiste à déformer F_{t-1} pour minimiser sa déviation spatiale de F_t. Nous suivons les positions physiques des

points ancres sur F_{t-1} avant et après sa déformation non-rigide.

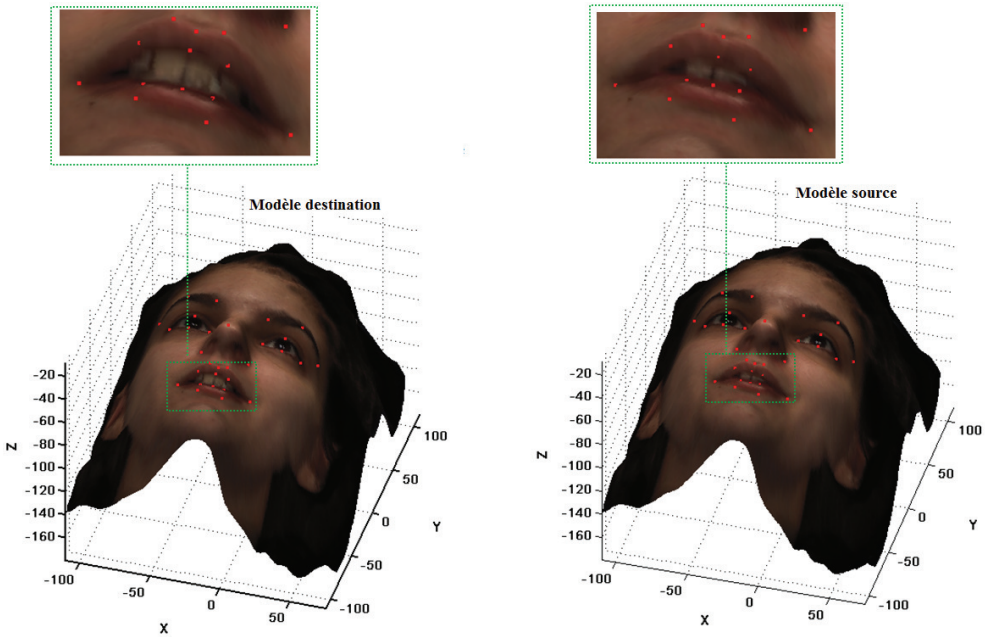


FIGURE 5.9 – Les localisations des points ancres pour les deux trames destination F_t et source F_{t-1} avant l'appariement non-rigide, F_t et F_{t-1} étant des données de la base BU de visages.

La figure 5.9 présente les localisations des points ancres sur les deux trames F_t et F_{t-1} avant sa déformation. La figure 5.10 illustre les localisations des points ancres sur la trame F_{t-1} après sa déformation, la carte couleur de la déviation spatiale et la distribution des points de F_{t-1} en fonction de leurs déviations spatiales avec leurs homologues de la trame F_t. Avant l'appariement non-rigide, les positions des points ancres sur F_{t-1} sont définies par leurs coordonnées sur F_t. Ainsi, les points ancres ne sont pas situés à leurs localisations physiques légitimes sur F_{t-1}.

Après la déformation non-rigide, les positions des points ancres sur F_{t-1} sont définies par la liste de correspondance entre la trame F_t et la trame F_{t-1} déformée. La figure 5.10 montre que les localisations des points ancres sur la trame déformée F_{t-1} correspondent à leurs positions physiques légitimes.

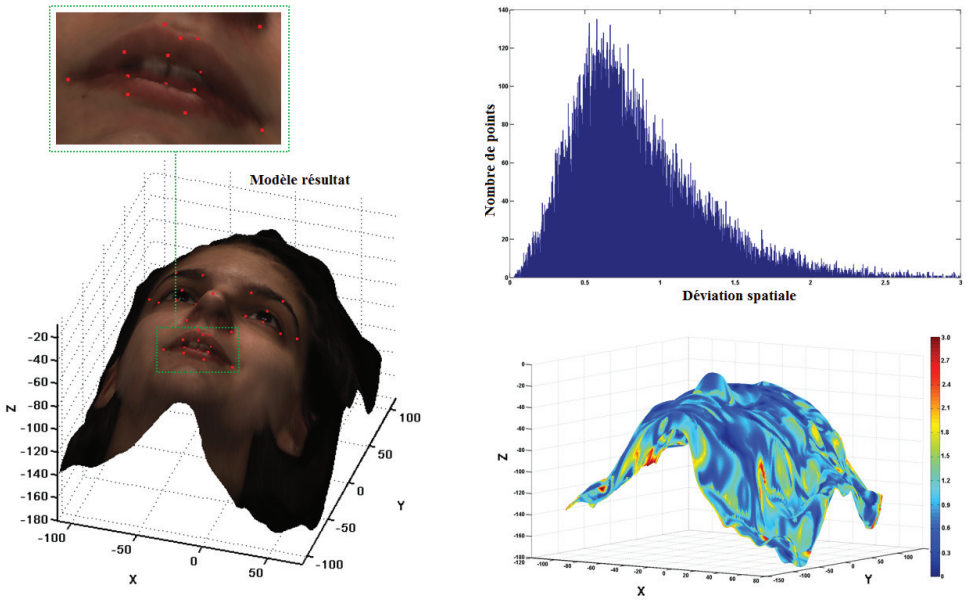


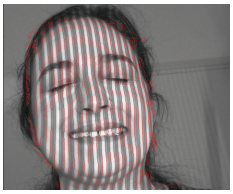
FIGURE 5.10 – Les localisations des points ancres pour les deux trames destination F_t et source F_{t-1} appariées, la carte couleur de leur déviation spatiale ainsi que la distribution des points de la trame F_t suivant leurs déviations spatiales avec leurs homologues sur la trame F_{t-1} appariée.

5.6.3 Super-résolution temporelle

La figure 5.11 présente un modèle 3D reconstruit par une super-résolution temporelle à partir de deux trames successives F_{t-1} et F_t d'un visage en mouvement avec une variation d'expression.

A un instant t, l'échantillonnage des primitives gauches et droites du visage n'a pas pu localiser tous les points d'intersection de franges ce qui génère quelques régions occultées et des artéfacts comme le montre les figures 5.11.c et 5.11.d. Les régions occultées se situent au niveau du nez et sur les bords du visage. Les deux figures 5.12 et 5.13 présentent les résultats de la numérisation dense de F_{t-1} et F_t par notre approche hybride de stéréovision et de codification sinusoïdale. La super-résolution temporelle non-rigide permet de corriger les artéfacts au niveau du nez et de l'œil droite comme le montre la figure 5.14.

(a) Image gauche de F_{t-1}



(b) Image droite de F_{t-1}

(c) Image gauche de F_t

(d) Image droite de F_t

FIGURE 5.11 – Les primitives droites et gauches de F_{t-1} et F_t.

(a) Maillage F_{t-1}

(b) Maillage F_t

FIGURE 5.12 – Les maillages de F_{t-1} et F_t calculés par notre approche hybride de stéréo-vision et de codification sinusoïdale.

(c) Texture F_{t-1}

(d) Texture F_t

FIGURE 5.13 – Les deux surfaces texturées de F_{t-1} et F_t.

(a) Maillage corrigé F_t

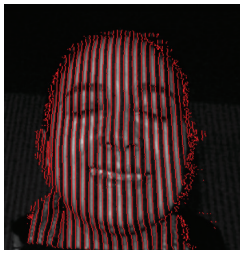
(b) Texture corrigée F_t

FIGURE 5.14 – Les résultats de la super-résolution temporelle de F_t en utilisant F_{t-1}.

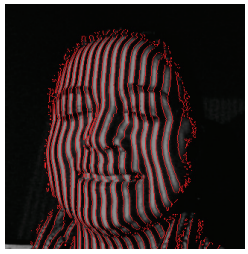
5.6.4 Super-résolution spatio-temporelle

La figure 5.15 présente deux ensembles de trois images 2D capturées par les trois caméras gauche, centrale, et droite. Le premier ensemble est capturé à l'instant t et le deuxième à l'instant $t-1$. Chaque vue est représentée par un ensemble de trois images contenant respectivement le premier patron sinusoïdal, le deuxième patron déphasé de π et le dernier patron blanc.

(a) Vue gauche de F_{t-1}



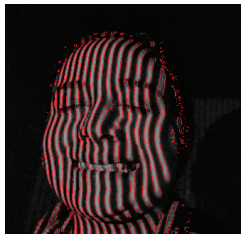
(b) Vue centrale de F_{t-1}



(c) Vue droite de F_{t-1}

(d) Vue gauche de F_t

(e) Vue centrale de F_t

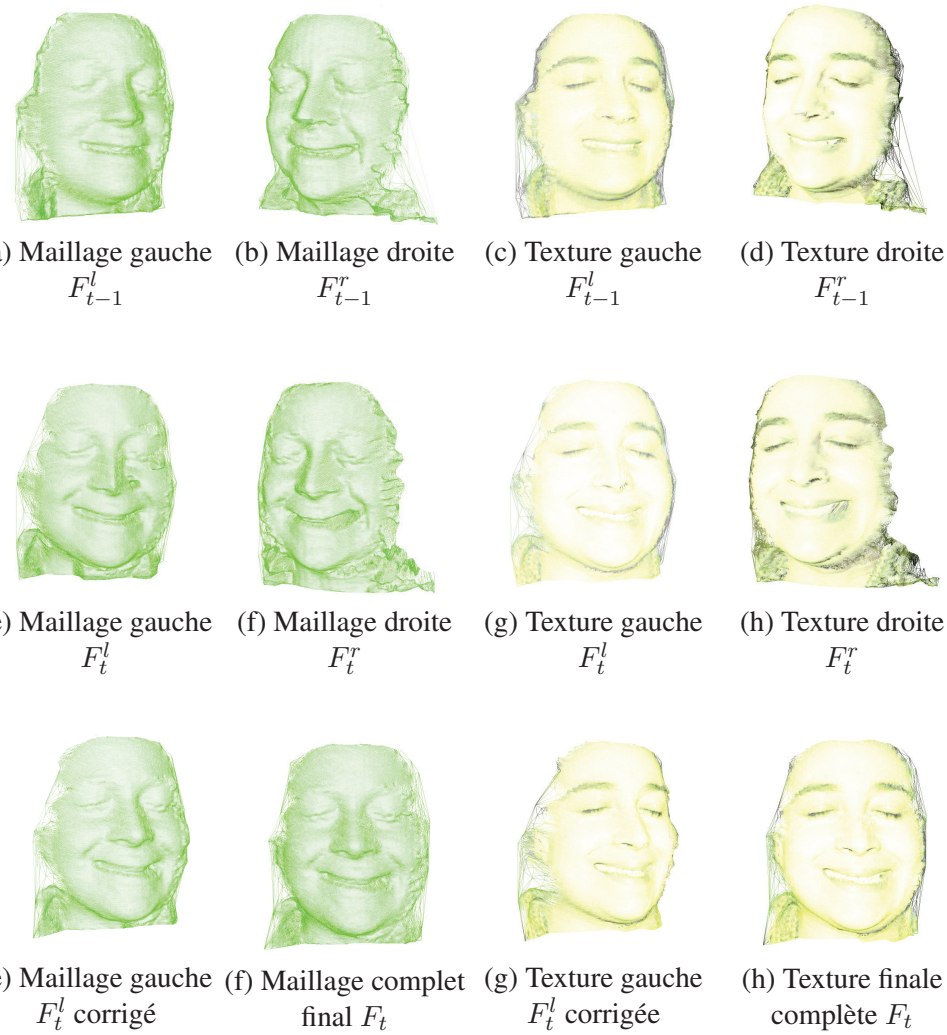


(f) Vue droite de F_t

FIGURE 5.15 – Primitives échantillonnées sur les vues gauche, centrale et droite pour deux trames successives de F_{t-1} et F_{t-1} en présence d'une variation d'expression faciale.

A l'instant $t-1$, les trois images $2D$ fournissent deux vues faciales $3D$ gauche F^l_{t-1}

et droite F^r_{t-1}. Aussi, à l'instant t, les trois images $2D$ fournissent deux vues faciales $3D$ gauche F^l_t et droite F^r_t. Quelques artéfacts peuvent être générés comme le montre la figure 5.16 particulièrement dans la vue $3D$ gauche F^l_t qui s'affiche dans la figure 5.16.e. Ici, ces erreurs de reconstruction 3D sont causés par une occultation.



(a) Maillage gauche F^l_{t-1} (b) Maillage droite F^r_{t-1} (c) Texture gauche F^l_{t-1} (d) Texture droite F^r_{t-1}

(e) Maillage gauche F^l_t (f) Maillage droite F^r_t (g) Texture gauche F^l_t (h) Texture droite F^r_t

(e) Maillage gauche F^l_t corrigé (f) Maillage complet final F_t (g) Texture gauche F^l_t corrigée (h) Texture finale complète F_t

FIGURE 5.16 – Les résultats de la super-résolution spatio-temporelle de F_t en utilisant F_{t-1}.

Pour corriger les données 3D obtenues, la première trame 3D complète F_{t-1} est utilisée pour corriger les deux vues 3D gauche et droite F^l_t et F^r_t et leur variation d'expression est considérée grâce à l'appariement non-rigide proposé dans la section 4. Ensuite, F^l_t et F^r_t sont fusionnées pour construire F_t. La figure 5.17 montre les localisations des points

ancres sur les deux trames destination F_t^l et source F_{t-1} avant sa déformation. La figure 5.18 présente les positions des points ancres sur la trame source F_{t-1} après sa déformation et la carte couleur de la déviation spatiale entre les deux trames recalées. Elle affiche aussi la distribution des points de F_t^l suivant leurs déviations spatiales avec leurs homologues sur F_{t-1}.

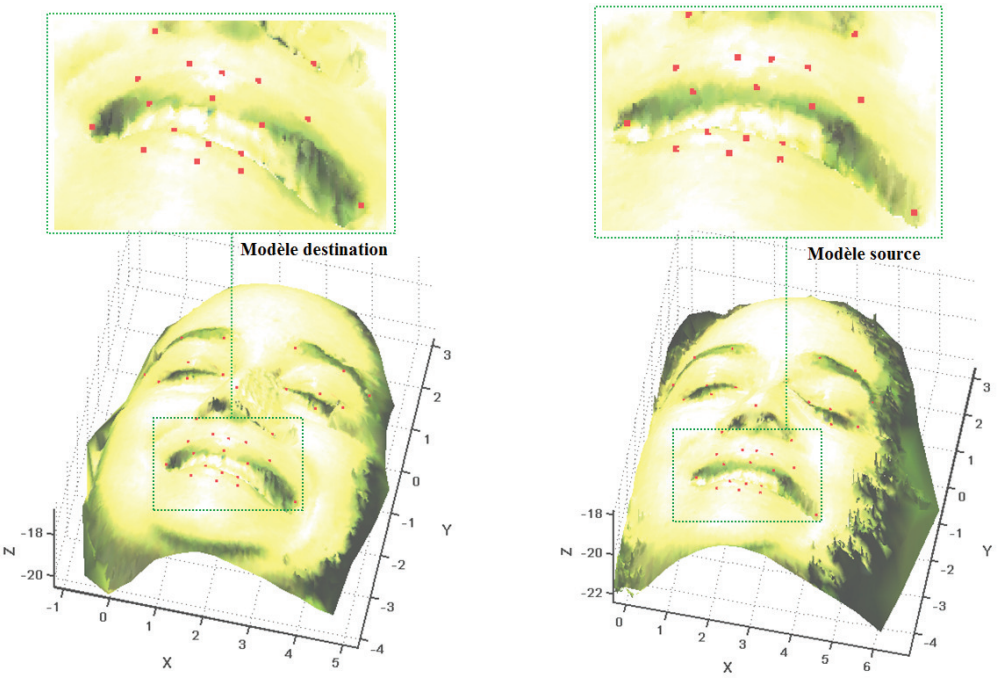


FIGURE 5.17 – La localisation des points ancres sur les deux trames destination F_t^l et source F_{t-1} avant sa déformation, les deux trames étant acquises par notre système tri-caméra.

Comme le montre la figure 5.19, l'appariement non-rigide aligne F_{t-1} avec F_t^l avec une déviation spatiale moyenne de $0.1222mm/pixel$. Au cours du processus itératif de l'algorithme CPD, une liste de correspondance dense entre les deux nuages de points est créée et mise à jour au fur et à mesure. Nous proposons un débruitage itératif basé sur les distances séparant les couples de points homologues définis par la liste de correspondance. Ceci permet de fournir un appariement plus efficace avec une nouvelle déviation spatiale moyenne de $0.0437mm/pixel$.

De plus, l'étape de débruitage itératif permet de mieux localiser les artéfacts et donc de les enlever efficacement puisque les points qui les forment présentent une haute déviation

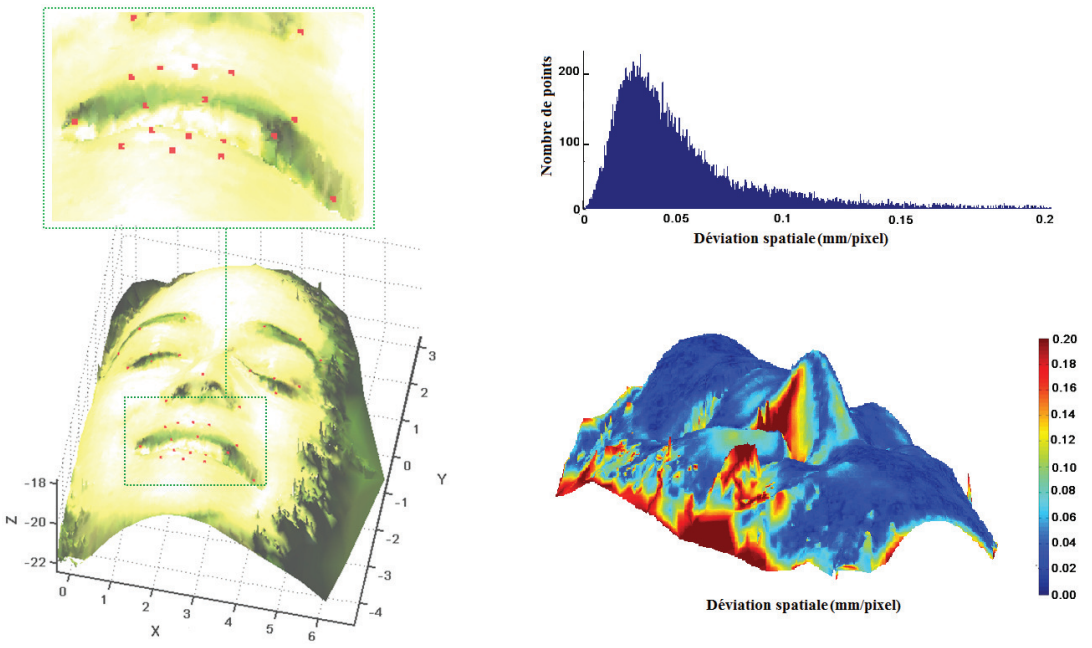


FIGURE 5.18 – La localisation des points ancres sur la trame source F_{t-1} après sa déformation, la carte couleur de sa déviation spatiale avec F_t^l, et la distribution des points de F_t^l en fonction de leurs déviations spatiales avec leurs homologues sur F_{t-1}.

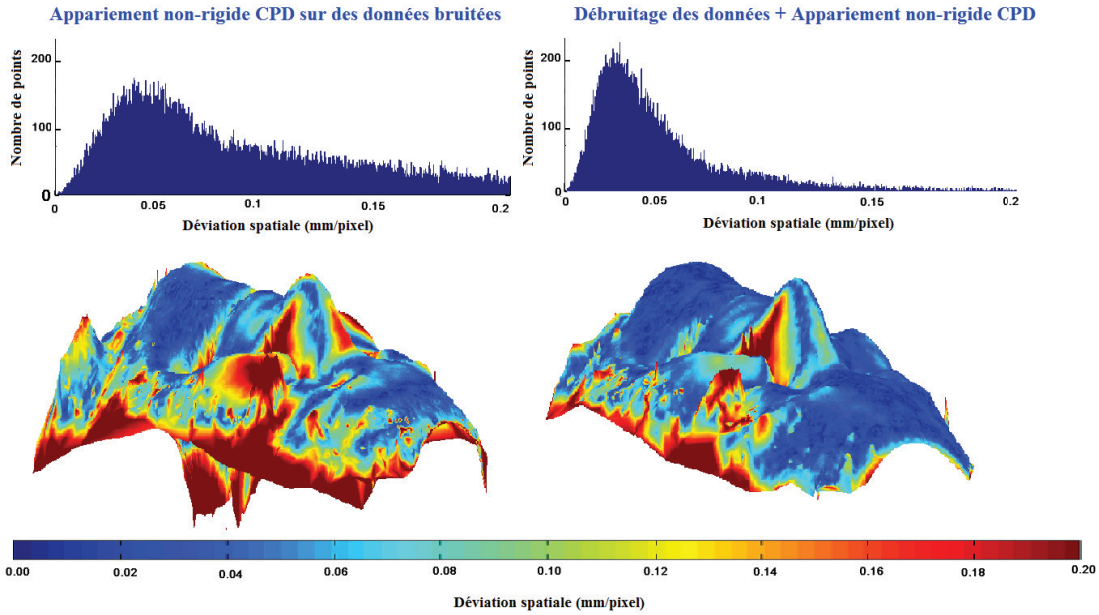


FIGURE 5.19 – Les résultats de l'appariement CPD non-rigide.

spatiale avec leurs points homologues de la trame 3D précédente. La figure 5.20 présente quelques trames 3D calculées par l'approche proposée.

FIGURE 5.20 – Quelques trames 3D reconstruites par la technique proposée.

5.6.5 Evaluation de la super-résolution/correction

Pour évaluer quantitativement le taux de l'erreur corrigée après la super-résolution temporelle, nous introduisons quelques erreurs sur un modèle 3D sans défauts préalables, nous nous proposons de le corriger comme le montre la figure 5.21 qui affiche les trois modèles origine, erroné et corrigé. Ensuite, nous procédons par un appariement rigide entre le modèle erronné et le modèle origine d'une part et un appariement rigide entre le modèle corrigé et le modèle origine d'autre part comme l'illustrent les figures 5.22 et 5.23.

(a)Modèle origine (b)Modèle erroné (c)Modèle corrigé

FIGURE 5.21 – Modèle origine, modèle erroné et modèle corrigé.

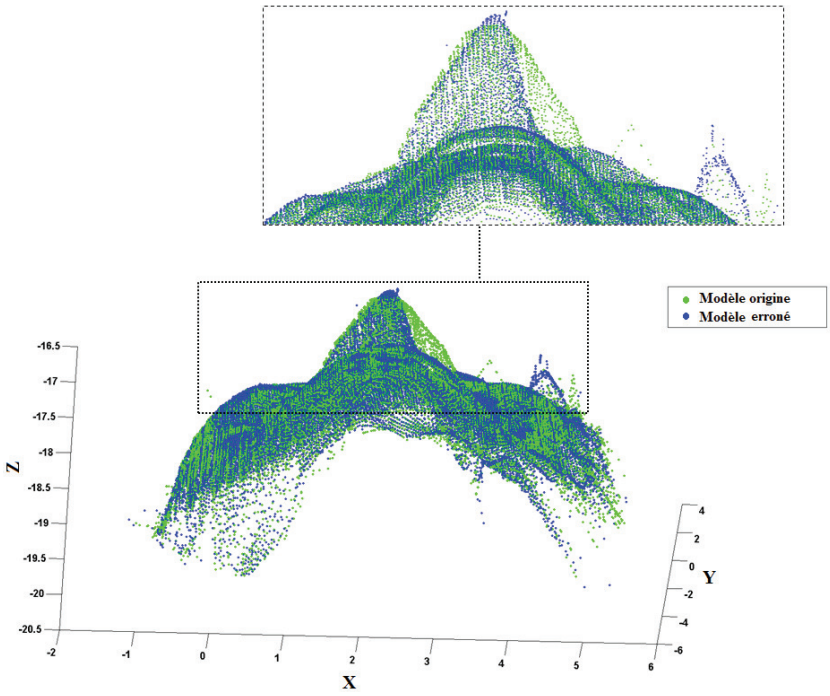


FIGURE 5.22 – Appariement rigide entre le modèle origine et le modèle erroné.

La déviation spatiale entre le modèle origine et le modèle erroné s'affiche sur la figure 5.24. La distribution se caractérise par une moyenne de $0.0369mm$, un écart-type de $0.0294mm$, une valeur minimale de $8.4894e-004mm$ et une valeur maximale de $0.1997mm$. La déviation moyenne relative est estimée à $18,129\%$. La figure 5.25 présente la déviation spatiale entre le modèle origine et le modèle corrigé avec une moyenne de $0.0311mm$, un écart-type de $0.0290mm$, une valeur minimale de $5.3265e-004mm$ et une valeur maximale de 0.1999. La déviation moyenne relative est évaluée à $15,332\%$.

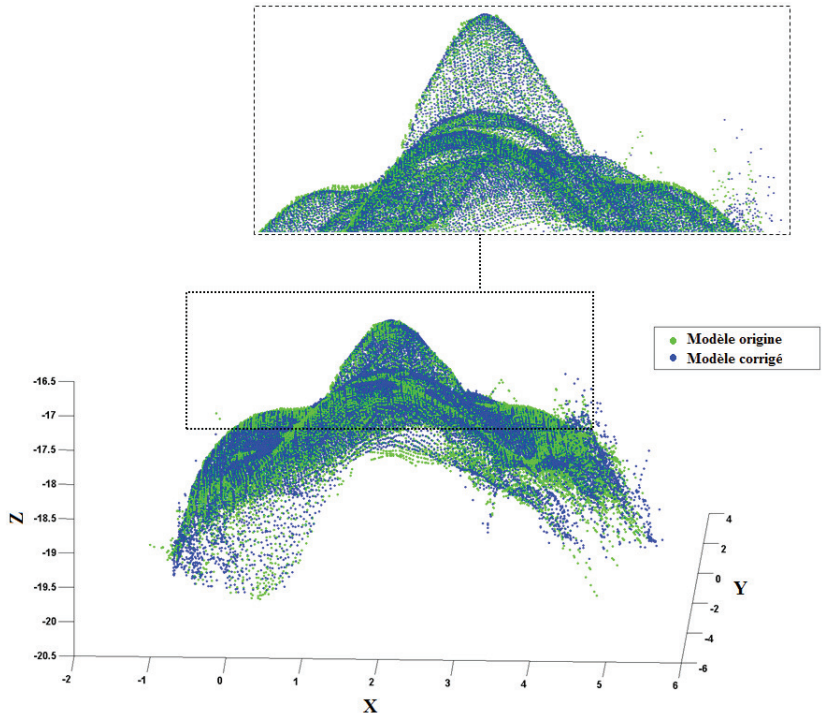


FIGURE 5.23 – Appariement rigide entre le modèle origine et le modèle corrigé.

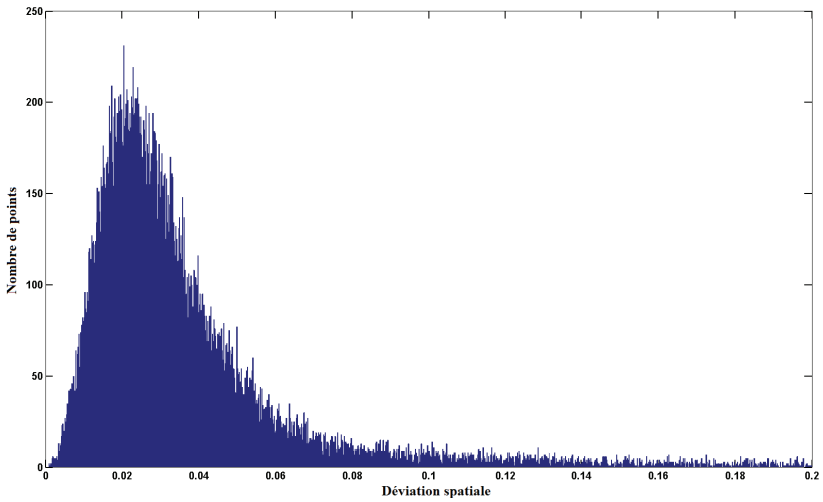


FIGURE 5.24 – Déviation spatiale entre le modèle origine et le modèle erroné.

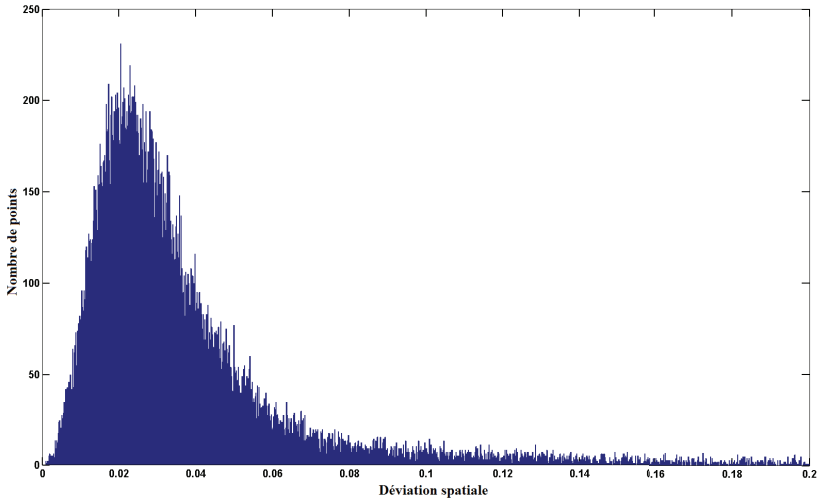


FIGURE 5.25 – Déviation spatiale entre le modèle origine et le modèle corrigé.

5.7 Conclusion

Dans ce chapitre, nous avons suggéré de corriger les erreurs aléatoires générées au cours de la numérisation d'un visage animé par une déformation rigide ou non-rigide. Ainsi, un système multi-caméras a été mis en place et une approche de super-résolution/correction spatio-temporelle a été proposée. Nous avons introduit une méthode d'appariement 3D non-rigide, qui utilise conjointement la forme et la texture du visage. La solution proposée est mobile, rapide et peu-coûteuse puisqu'elle utilise des caméras grand public.

Néanmoins, la super-résolution temporelle ne permet pas de corriger une trame 3D si elle présente de sévères artéfacts. Alors, ces derniers impactent les trames 3D suivantes. Nous projetons, dans nos travaux futurs, de considérer plus de caméras et plus de trames 3D dans la super-résolution temporelle pour renforcer la précision de la numérisation. Une synchronisation entre les caméras et le vidéoprojecteur ainsi qu'une implémentation GPU sont aussi envisagées pour assurer une numérisation en temps-réel et une super-résolution/correction spatio-temporelle multi-caméras plus rapide. Finalement, une source de projection infrarouge doit remplacer le vidéoprojecteur actuel puisque la projection des franges visibles est intrusive et dérangeante. En plus, cette projection cause une dégradation de la texture faciale.

CHAPITRE 6

Conclusion et Travaux Futurs

Ce travail de recherche a été réalisé dans le cadre du projet FAR 3D ANR-07-SESU-004 (Face Analysis and Recognition using 3D). Il adresse le problème de la numérisation 3D de surfaces mal-texturées. La solution proposée est à faible coût puisqu'elle utilise un hardware grand public formé d'un banc stéréo ou multi-caméras étalonné assisté par un vidéoprojecteur non-étalonné. Elle est particulièrement adaptée à la numérisation 3D de visages en mouvement, c.-à-d. rapide, mobile et robuste à la variation de la pose et de l'expression faciale. Cette solution trouve son application essentiellement en vidéosurveillance pour une identification faciale ou encore pour le suivi postopératoire d'une chirurgie faciale.

6.1 Contributions

Dans ce manuscrit de thèse, nous avons introduit deux contributions majeures. D'abord, nous avons conçu une approche de numérisation 3D dense d'un visage statique par une projection d'une lumière structurée codée par un multiplexage temporel. Ensuite, nous avons proposé une approche de super-résolution spatio-temporelle qui renforce la qualité du rendu facial numérisé par une correction des aberrations/artéfacts crées au moment de la numérisation. Notre solution de correction/super-résolution permet une numérisation plus précise en présence de déformations non-rigides de la surface faciale et de distorsions systématiques.

6.1.1 Technique hybride de numérisation

Notre approche de numérisation hybride utilise la stéréovision active et la codification sinusoïdale par décalage de phase pour reconstruire des surfaces mal-texturées comme le

visage. Dans un premier temps, la stéréovision active permet une première estimation de la surface faciale par un échantillonnage sous-pixélique des points d'intersection de franges. Cette première estimation est dense sur la direction verticale parallèle aux franges sinusoïdales projetées et non-dense dans sa direction orthogonale.

Dans un deuxième temps, une estimation en ligne des paramètres du vidéoprojecteur nous accorde plus de flexibilité permettant de déplacer le vidéoprojecteur en cours de la numérisation pour couvrir plus de surface. En effet, à la différence des approches classiques de codification sinusoïdale, un étalonnage préalable du projecteur doit se faire hors ligne. Ainsi, le système projecteur-caméra reste figé au cours de la numérisation. La paramétrisation enligne est assurée en utilisant le nuage de points 3D non-dense du visage. La stéréovision active permet une résolution 3D spatiale qui dépend du nombre de franges sinusoïdales projetées sur la surface faciale.

La codification sinusoïdale par décalage de phase permet de densifier le nuage de points non-dense et de renforcer sa précision. Ainsi, nous calculons les coordonnées 3D de chaque pixel intrafrange par une reconstruction géométrique, pour chaque caméra séparément. La complexité de l'appariement stéréo dépend uniquement du nombre de franges projetées sur la surface faciale et ne dépend pas de la résolution pixélique des caméras. Ceci permet de numériser de la vidéo 3D de visages en temps réel même avec des caméras de très haute résolution.

6.1.2 Super-résolution spatio-temporelle non-rigide

Les erreurs/aberrations, qui caractérisent notre technique de numérisation hybride, peuvent être causés par une occultation du visage sur les différentes vues capturées par les caméras, par exemple. Les occultations causent de faux appariements gauche/droite et une estimation erronée du modèle non-dense de visage. Ceci fait apparaitre des aberrations/artéfacts sur le visage 3D dense puisque l'estimation des paramètres du vidéoprojecteur ainsi que la densification intrafrange se basent sur le modèle non-dense préalablement calculé. En plus, le multiplexage temporel que nous utilisons pour projeter la lumière structurée ne s'adapte pas à la numérisation de visages en mouvement. En effet, une légère variation de la pose, ou de l'expression du visage mène à une localisation imprécise

des points d'intersection de franges. Ceci engendre des ondulations verticales sur le rendu 3D justifiées par une estimation biaisée de la disparité. Dans ce manuscrit de thèse, nous avons introduit une correction/super-résolution spatio-temporelle pour traiter les erreurs aléatoires. Nous proposons de traiter les déformations non-rigides d'un visage animé pour permettre une numérisation plus précise.

Nous avons proposé dans ce travail de thèse, un schéma multi-caméra formé de N caméras pour une numérisation 3D dense d'un visage en mouvement. Dans un premier temps, et à un instant t, chaque couple de caméras voisines permet de calculer un nuage de points 3D dense du visage en utilisant notre technique hybride de numérisation. Dans un deuxième temps, la super-résolution spatiale permet la construction de la trame courante par une fusion des $N-1$ vues 3D acquises à l'instant t. Enfin, la super-résolution temporelle permet de corriger la trame courante en utilisant sa trame précédente, en étant robuste à une éventuelle déformation non-rigide comme par exemple une variation de l'expression faciale.

La super-résolution spatio-temporelle s'effectue en une étape d'appariement suivie par une fusion et un débruitage. D'abord, une approche non-rigide d'appariement 3D basée sur l'algorithme CPD, est assurée en utilisant jointement la forme et la texture pour recaler les nuages deux nuages de points 3D. L'algorithme CPD est sensible à la présence de bruit et ne réussit pas à annuler efficacement la déformation faciale. Au cours du processus itératif de l'algorithme, une liste de correspondance dense entre les deux nuages de points est créée et mise à jour au fur et à mesure. Un débruitage itératif est assuré en utilisant les distances séparant les couples de points homologues définis par la liste de correspondance. Ceci facilite la convergence de l'appariement non-rigide vers le minimum global. Une étape de fusion des vues 3D appariées suivie d'une étape de débruitage permettent d'obtenir la trame complète courante. Les étapes de fusion et de débruitage sont assurées sur des cartes séparées contenant les informations X, Y, Z et la texture séparément puisque les points 3D du modèle numérisé ont une précision sous-pixélique et ne peuvent pas être représentés par une image de profondeur classique.

6.2 Perspectives

D'abord, nous envisageons une étude plus avancée de la performance de notre technique de numérisation. Ensuite, nos perspectives ciblent d'une part une numérisation plus efficace d'un visage statique et d'autre part une numérisation d'un visage en mouvement en temps réel et sans artéfacts.

6.2.1 Etude de la performance de la numérisation

L'utilisation d'un plan pour l'estimation de la qualité de notre reconstruction constitue une première étape pour évaluer notre système. Ainsi, nous envisageons d'utiliser un objet rigide de géométrie connue, telle qu'une sphère dont on connaît précisément le diamètre ou une cale étalon. Un tel objet constitue une vérité terrain permettant ainsi d'évaluer la précision exacte d'un système de numérisation. Il est souvent appelé objet fantôme. L'utilisation d'une cale étalon permet de mesurer la précision du système selon les 3 axes x, y et z séparément contrairement à l'utilisation de la sphère étalon.

Nous envisageons aussi d'analyser l'influence de l'amélioration de la résolution de la caméra sur la résolution du système de numérisation et plus particulièrement l'étude du rapport Résolution-caméra/Résolution-système et le rapport Résolution-caméra/Précision-système. En effet, plus la résolution de la caméra est bonne, plus le nombre de points obtenus par échantillonnage augmente. Ainsi, le nuage de points calculés par triangulation optique est plus dense. En d'autres termes, la résolution du système augmente. Ainsi, les erreurs de numérisation diminuent et la précision du modèle calculé augmente. Le modèle 3D numérisé ressemble donc de plus en plus au visage réel. Nous proposons aussi d'étudier l'influence de la variation de la largeur des franges sur la résolution et la précision du système.

6.2.2 Numérisation statique

Notre approche de numérisation hybride permet de produire un nuage 3D dense de points en utilisant un couple de caméras gauche et droite. Les points qui forment le modèle non-dense proviennent des primitives gauches et droites homologues. Néanmoins, les primitives échantillonnés dans l'image gauche et dans l'image droite ne participent pas

tous dans ce modèle non-dense. En fait, certaines primitives gauches sont occultées dans la caméra droite.

Nous envisageons de reconstruire l'information 3D pour toutes les primitives sans homologues ainsi que les pixels intrafranges de la région occultée. Il suffit de considérer la primitive la plus proche appartenant au modèle non-dense comme point de référence à phase nulle. Ainsi, pour chaque pixel de la région occultée, sa valeur de phase absolue par rapport à ce point de référence est d'abord calculée. Ensuite, les coordonnées 3D de ce pixel peuvent être estimées par notre approche de reconstruction géométrique décrite dans la section 4.5.3. Ceci permet de compléter la vue 3D en faisant participer les deux régions occultées gauche et droite. Ceci constitue un avantage majeur par apport à une approche de stéréovision classique qui numérise seulement la région commune des vues gauche et droite.

Pour la segmentation du visage, nous utilisons actuellement une fenêtre glissante fixe de 64 pixels pour les caméras de résolution 640x480 et de 128 pixels pour les caméras de résolution 1600x1200 et un seuil empirique égal à 0.6 pour la localisation de la région d'intérêt. Aussi, nous utilisons un nombre de 300 itérations pour assurer la convergence de l'algorithme RANSAC. Nous projetons d'une part d'automatiser l'estimation de ces paramètres et d'autre part de les optimiser. Nous comptons aussi évaluer d'autres patrons de lumière structurée sur le visage et étudier l'influence du choix du patron sur le taux de l'erreur gamma, sur l'efficacité de l'estimation de la profondeur, et aussi sur la complexité du calcul. Par exemple, nous envisageons de projeter des patrons triangulaires pour éviter l'utilisation de la fonction $arcsin$ et rendre la variation de la phase linéaire et non sinusoïdale. Aussi, une utilisation d'un codage spatial de la lumière structurée remplacera le multiplexage temporel pour capturer le mouvement 3D de visage, d'une manière plus fluide.

6.2.3 Numérisation de mouvement

Le multiplexage temporel que nous utilisons pour projeter la lumière structurée sur le visage exige une synchronisation à bas niveau entre les caméras et le vidéoprojecteur pour réussir une capture synchrone des différents patrons sur le visage. Aussi, la fréquence des

caméras doit être de l'ordre de 90 fps (trames/secondes) pour pouvoir capturer le mouvement du visage avec une fréquence de 30fps (trame3D/secondes). En plus, une implémentation GPU s'avère nécessaire pour accélérer les calculs et permettre une numérisation 3D en temps-réel et une super-résolution/correction spatio-temporelle multi-caméras rapide.

Finalement, la projection de patrons visibles sur le visage est intrusive et dérangeante. Cette projection cause même une dégradation de la texture faciale. Nous comptons remplacer le vidéoprojecteur actuel par une source de projection infrarouge. Ainsi, notre système de numérisation peut être utilisé pour un contrôle d'accès par identification faciale, par exemple.

Liste des tableaux

Table des figures

CHAPITRE 7

Publications

Mes travaux de recherche ont fait l'objet d'un papier revue, un dépôt logiciel, huit publications dans des conférences internationales, six publications dans des conférences francophones et nationales et d'un chapitre de livre.

Papiers Revues

1. **K.Ouji**, M.Ardabilian, L.Chen and F.Ghorbel : 3D Deformable Super-Resolution For Multi-camera 3D Face Scanning. Int. Journal of Mathematical Imaging and Vision, Soumis le 2 décembre 2011.

Dépôt logiciel

1. M. Ardabilian, L. Chen, **K.Ouji**, B Ben Amor. M.Ardabilian et L.Chen : Procédé d'acquisition et de modélisation 3D sans contact et avec ajout de texture, 2010.

Conférences Internationales

1. **K.Ouji**, M.Ardabilian, L.Chen et F.Ghorbel : Pattern-based Face Localization and Online Projector Parameterization for Multi-Camera 3D Scanning. International Conference on 3D Body Scanning Technologies, Lugano, Switzerland, 2011 ;
2. **K.Ouji**, M.Ardabilian, L.Chen et F.Ghorbel : Multi-Camera 3D Scanning with a Non-rigid and Space-Time Depth Super-Resolution capability, IAPR International

Conference on Computer Analysis of Images and Patterns (CAIP), Seville, Spain, 2011 ;

3. **K.Ouji**, M.Ardabilian, L.Chen et F.Ghorbel : A Space-Time Depth Super-Resolution Scheme For 3D Face Scanning, IEEE Advanced Concepts for Intelligent Vision Systems Conference (ACIVS), Het Pand, Ghent, Belgium, 2011 ;
4. D. Huang, **K.Ouji**, M. Ardabilian, Y. Wang et L. Chen : 3D Face Recognition based on Local Shape Patterns and Sparse Representation Classifier, IEEE MultiMedia Modeling Conference (MMM), Taipei, Taiwan, 2011 ;
5. **K.Ouji**, M.Ardabilian, L.Chen et F.Ghorbel : Pattern Analysis for an automatic and Low-Cost 3D Face acquisition Technique, IEEE Advanced Concepts for Intelligent Vision Systems Conference (ACIVS), Bordeaux, 2009 ;
6. **K.Ouji**, B.Ben Amor, M.Ardabilian, F.Ghorbel et L.Chen : 3D Face Recognition using R-ICP and Geodesic Computation Coupled Approach, IEEE MultiMedia Modeling Conference (MMM), Sophia Antipolis, 2009 ;
7. **K.Ouji**, B.Ben Amor, M.Ardabilian, F.Ghorbel et L.Chen : 3D Face Recognition using ICP and Geodesic Computation Coupled Approach, IEEE/ACM (SITIS), Hammamet, Tunisie, 2006 ;
8. B.Ben Amor, **K.Ouji**, M.Ardabilian and L.Chen : Face recognition by ICP-based shape matching, (ACIDCA-ICMI), Tozeur, Tunisie, 2005 ;

Conférences Francophones et Nationales

1. **K.Ouji**, M. Ardabilian, L. Chen et Faouzi Ghorbel : Une approche de super-résolution spatio-temporelle pour l'acquisition 3D de visages, Traitement et Analyse de l'Information Méthodes et Applications (TAIMA), Hammamet, 2011 ;
2. P. Lemaire, W. Ben Soltana, D. Huang, **K.Ouji**, M. Ardabilian, L.Chen : Reconnaissance rapide, robuste et résistante aux leurres de visages en 3D, Workshop Interdisciplinaire sur la Sécurité Globale, Troyes, 2011.
3. M. Ardabilian, **K.Ouji**, D. Huang, P. Szeptycki, P. Lemaire, W. Ben Soltana, L.Chen : Biométrie faciale 3D - Acquisition résistante aux leurres et reconnaissance,

Workshop Interdisciplinaire sur la Sécurité Globale, Troyes, 2010.

4. M. Ardabilian, **K.Ouji**, D. Huang, P. Szeptycki, P. Lemaire, W. Ben Soltana, L.Chen : Biométrie faciale 3D - Acquisition, prétraitement et reconnaissance, Workshop Interdisciplinaire sur la Sécurité Globale, Troyes, 2009.
5. Y. Desbois, **K.Ouji**, M. Ardabilian, R. Perrot : Contribution de la biométrique de similarité à l'identification des auteurs de vols à main armée : le projet IDASOR, Workshop Interdisciplinaire sur la Sécurité Globale, Troyes, 2009.
6. B.Ben Amor, **K.Ouji**, M.Ardabilian and L.Chen : Une nouvelle approche d'appariement 3D orienté régions pour la reconnaissance faciale, Workshop Interdisciplinaire sur la Sécurité Globale, (TAIMA), Hammamet, Tunisie, 2007.

Chapitres de livre

1. B.Ben Amor, **K.Ouji**, M.Ardabilian, F.Ghorbel and L.Chen : 3D Face Recognition using ICP and Geodesic Computation Coupled Approach, Book Chapter on Signal Processing for image enhancement and multimedia processing, Springer-Verlag Edition, 2007.

Bibliographie

[Abdul-Rahman *et al.* 2008] H.S. Abdul-Rahman, M.A. Gdeisata, D.R. Burtona, M.J. Lalora, F. Lilleya and A. Abid. *Three-dimensional Fourier Fringe Analysis*. J.Optics and Lasers in Engineering, vol. 46, pages 446–455, 2008. 22

[Aliomonos & Swain 1987] J. Aliomonos and M. Swain. *Shape from texture*. IJCAl, pages 926–931, 1987. 16

[Anke *et al.* 2008] B. Anke, H. Olaf, R. Volker and Y. Ulas. *A Benchmark dataset for performance evaluation of shape-from-X algorithms*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2008. 16, 26

[Arold *et al.* 2009] O. Arold, Z. Yang, S. Ettl and G. Häusler. *A new registration method to robustly align a series of sparse 3D data*, 2009. Deutschen Gesellschaft für angewandte Optik. 24

[Arold *et al.* 2010] O. Arold, Z. Yang, S. Ettl and G. Häusler. *How precise is 'Flying Triangulation' ?*, 2010. Deutschen Gesellschaft für angewandte Optik. 24

[Baker & Kanade 2002] S. Baker and T. Kanade. *Limits on super-resolution and how to break them*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pages 1167–1183, 2002. 31

[B.Horn & Brooks 1989] B.Horn and M.J. Brooks. *Shape from Shading*, 1989. MIT Press. 17

[Bishop 1995] C.M. Bishop. *Neural networks for pattern recognition*. Oxford Univ, Press, 1995. 119

[Blais 2004] F. Blais. *Review of 20 years of range sensor development*. J. Electronic Imaging, vol. 13, pages 231–240, 2004. 26

[Blake & Zisserman 1987] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987. 13

[Blanz & Vetter 1999] V. Blanz and T. Vetter. *A morphable model for the synthesis of 3d-faces*, 1999. SIGGRAPH. 36

[Blanz & Vetter 2003] V. Blanz and T. Vetter. *Face recognition based on fitting a 3d morphable model*. PAMI, vol. 25, pages 1063–1074, 2003. 28

[Bookstein 1989] F.L. Bookstein. *Principal warps : Thin-plate splines and the decomposition of deformations*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, pages 567–585, 1989. 117

[Borga & Knutsson 1998] M. Borga and H. Knutsson. *An adaptive stereo algorithm based on canonical correlation analysis*, 1998. IEEE International Conference on Intelligent Processing Systems, Gold Coast, Austalia. 13

[Bronstein *et al.* 2005] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Threedimensional face recognition*. International Journal of Computer Vision, vol. 64, pages 5–30, 2005. 37

[Bronstein *et al.* 2006] A. M. Bronstein, M. M. Bronstein and R. Kimmel. *Efficient computation of isometry-invariant distances between surfaces*. SIAM Journal of Scientific Computing, vol. 28, pages 1812–1836, 2006. 37, 117

[Chen *et al.* 2000] F. Chen, G. M. Brown and M. Song. *Overview of Three-Dimensional Shape Measurement Using Optical Method*. Optical Engineering, vol. 39, pages 10–22, 2000. 25

[Chuang *et al.* 2002] E.S. Chuang, H. Deshpande and C. Bregler. *Facial Expression Space Learning*, 2002. In IEEE Pacific conference on computer graphics and applications. 36

[Chui & Rangarajan 2000] H. Chui and A. Rangarajan. *A new algorithm for non-rigid point matching*. IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pages 44–51, 2000. 117

[Cootes *et al.* 2001] T.F. Cootes, G.J. Edwards and C.J. Taylor. *Active Appearance Models*. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pages 681–685, 2001. 35, 36

[C.Tomasi & R.Manduchi 1998] C.Tomasi and R.Manduchi. *Stereo matching as a nearest-neighbor problem*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, pages 330–340, 1998. 13

[Cui *et al.* 2010] Y. Cui, S. Schuon, D. Chan, S. Thrun and C. Theobalt. *3D Shape Scanning with a Time-of-Flight Camera*, 2010. 3DPVT Conference. 19, 33, 147

[Curless 2000] B. Curless. *Overview of active vision techniques*, 2000. SIGGRAPH Course on 3D Photography. 9

[D'apuzzo 2006] N. D'apuzzo. *State of the art of the methods for static 3D scanning of partial or full human body*, 2006. International Conference of 3D Modelling, Paris, France. 39

[Davis *et al.* 2003] J. Davis, R. Ramamoothi and S. Rusinkiewicz. *Spacetime Stereo : A Unifying Framework for Depth from Triangulation*, 2003. IEEE Computer Vision and Pattern Recognition. 31

[de Berg *et al.* 2000] M. de Berg, M. van Kreveld, M. Overmars and O. Schwarzkopf. *Computational Geometry : Algorithms and Applications, Second Edition*, 2000. Springer. 68

[de Berg *et al.* 2008] M. de Berg, M. van Kreveld, M. Overmars and O. Schwarzkopf. *Computational Geometry : Algorithms and Applications, Third Edition*, 2008. Springer. 68

[Delaunay 1934] B. Delaunay. *Sur la sphère vide*. Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk, vol. 7, pages 793–800, 1934. 68

[Dempster *et al.* 1977] A. Dempster, N. Laird and D. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. J. Royal Statistical Soc. Series B, vol. 39, pages 1–38, 1977. 119

[D.N.Bhat & S.K.Nayar 1998] D.N.Bhat and S.K.Nayar. *Ordinal measures for visual correspondence*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, pages 415–423, 1998. 13, 18

[D.Scharstein & R.Szeliski 1998] D.Scharstein and R.Szeliski. *Stereo matching with nonlinear diffusion*. Int'l Joint Conf. Artificial Intelligence, vol. 28, pages 155–174, 1998. 13

[Edwards *et al.* 1998] G.J. Edwards, T.F. Cootes and C.J. Taylor. *Face recognition using active appearance models*. In IEEE European Conference on Computer Vision, vol. 2, pages 581–695, 1998. 36

[Ettl *et al.* 2009] S. Ettl, O. Arold, P. Vogt, O. Hybl, Z. Yang, W. Xie and G. Häusler. *'Flying Triangulation' - a new optical 3D sensor enabling the acquisition of surfaces by freehand motion*, 2009. Deutschen Gesellschaft für angewandte Optik. 23, 24, 147

[Farsiu *et al.* 2004] S. Farsiu, D. Robinson, M. Elad and P. Milanfar. *Fast and robust multiframe super-resolution*, 2004. IEEE Trans. Image Processing. 32, 120

[Favaro & Soatto 2002] P. Favaro and S. Soatto. *Learning shape from defocus*, 2002. European Conference on Computer Vision. 17

[Fidaleo & Medioni 2007] D. Fidaleo and G. Medioni. *Model-Assisted 3D Face Reconstruction from Video*, 2007. IEEE International Workshop on Analysis and Modeling of Faces and Gestures. 28, 34

[Fischler & Bolles 1981] Martin A. Fischler and Robert C. Bolles. *Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography*. Comm. Of the ACM, vol. 24, pages 381–395, 1981. 93

[Fredricksen 1970] H. Fredricksen. *The lexicographically least debruijn cycle*. Journal of Combinatorial Theory, vol. 9, pages 509–510, 1970. 21

[Fua 2000] P. Fua. *Regularized bundle adjustment to model heads from image sequences without calibration data*. IJCV, vol. 38, pages 153–171, 2000. 15, 28

[Geiger *et al.* 1992] D. Geiger, B. Ladendorf and A. L. Yuille. *Occlusions and Binocular Stereo*. Second European Conference on Computer Vision Science, pages 425–433, 1992. 13

[Goulette 1999] F. Goulette. *Modélisation 3D automatique : outils de géométrie différenctielle*. ISBN 2–911762–18–5, 1999. 12

[Gu & Yau 2003] X. Gu and S. Yau. *Surface classification using conformal structures*. IEEE international conference on computer vision, pages 701–708, 2003. 36

[Guenter *et al.* 1998] B. Guenter, C. Grimm and D. Wood. *Making Faces*. SIGGRAPH, 1998. 35

[Han *et al.* 2009] X. Han, P. Huang, Z. Deng and L. Xu. *Combined stereovision and phase shifting method : use of a color visibility-modulated fringe pattern*. Proceedings of SPIE, vol. 7432, 2009. 30

[Hernàndez *et al.* 2008] C. Hernàndez, G. Vogiatzis and R. Cipolla. *Shadows in three-source photometric stereo*, 2008. ECCV. 18

[Hertzmann & Seitz 2003] A. Hertzmann and S.M. Seitz. *Shape and materials by example : A photometric stereo approach*. IEEE Conference on Computer Vision and Pattern Recognition, Madison, vol. 1, pages 533–540, 2003. 26

[Hirschmüller & Scharstein 2009] H. Hirschmüller and D. Scharstein. *Evaluation of Stereo Matching Costs on Images with Radiometric Differences*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pages 1582–1599, 2009. 12, 13, 17

[Huang & Zhang 2006] P.S. Huang and Song Zhang. *Fast Three-Step Phase-Shifting Algorithm*. J. Applied Optics, vol. 45, pages 5086–5091, 2006. 22

[Huang *et al.* 2003] P. S. Huang, C. P. Zhang and F. P. Chiang. *High speed 3-d shape measurement based on digital fringe projection*. Journal of Optical Engineering, vol. 42, pages 163–168, 2003. 22, 23, 147

[Huntley & Saldner 1997] J.M. Huntley and H.O. Saldner. *Shape measurement by temporal phase unwrapping : comparision of unwrapping algorithms*. J. Measurement Science and Technology, vol. 8, pages 986–992, 1997. 23

[Ilic & Fua 2006] S. Ilic and P. Fua. *Implicit Meshes for Surface Reconstruction*. PAMI, 2006. 28

[Jones 1997] G. A. Jones. *Constraint, Optimization, and Hierarchy : Reviewing Stereoscopic Correspondence of Complex Features*. International Journal of Computer Vision, vol. 65, pages 57–58, 1997. 13

[Kang & Byun 2008] B. Kang and H. Byun. *Multi-resolution 3D Morphable Models and Its Matching Method*, 2008. International Conference on Pattern Recognition. 28

[Kang *et al.* 2002] H. Kang, T.F. Cootes and C.J. Taylor. *Face Expression Detection and Synthesis using Statistical Models of Appearance*, 2002. In Measuring Behavior, Amsterdam, The Netherlands. 36

[Kil *et al.* 2006] Y. Kil, Y. Mederos and N. Amenta. *Laser scanner super-resolution*, 2006. Eurographics Symposium on Point-Based Graphics. 33

[Klaudiny *et al.* 2010] M. Klaudiny, A. Hilton and J. Edge. *High-detail 3D capture of facial performance*, 2010. 3DPVT Conference. 28, 29, 44, 147

[Kolmogorov *et al.* 2003] V. Kolmogorov, R. Zabih and S. J. Gortler. *Generalized Multi-Camera Scene Reconstruction Using Graph Cuts*. Computer Vision and Pattern Recognition, vol. 2683, pages 501–516, 2003. 13

[Lee *et al.* 2003] J. Lee, B. Moghaddam, H. Pfister and R. Machiraju. *Silhouette-based 3D face shape recovery*, 2003. Proceedings of Graphics Interface, Halifax, Nova Scotia, Canada. 14, 147

[Marr & Poggio 1976] D. Marr and T. Poggio. *A cooperative stereo algorithm*. J. Science, vol. 194, 1976. 13

[Myronenko & Song 2010] A. Myronenko and X. Song. *Point set registration : Coherent point drift*. IEEE Trans. PAMI, vol. 32, pages 2262–2275, 2010. 117, 118, 120, 122

[Myronenko *et al.* 2007] A. Myronenko, X. Song and M. A. Carreira-Perpinan. *Non-rigid point set registration : Coherent Point Drift*, 2007. NIPS Conference. 116, 117

[Nayar *et al.* 1996] S.K. Nayar, M. Watanabe and M. Noguchi. *Real-time focus range sensor*. PAMI, vol. 12, pages 1186–1119, 1996. 16, 26

[Niem & Wingbermuhle 1997] W. Niem and J. Wingbermuhle. *Automatic reconstruction of 3d objects using a mobile monoscopic camera*, 1997. Conf. on Recent Advances in 3D Imaging and Modeling. 15

[Nitzan 1988] D. Nitzan. *Three dimensional vision structure for robot applications*. IEEE Trans. PAMI, 1988. 16

[O.Faugeras 1993] O.Faugeras. *Three-Dimensional Computer Vision*, 1993. MIT Press. 12

[Ohta & Kanade 1985] Y. Ohta and T. Kanade. *Stereo intra- and interscanline search using dynamic programming*, 1985. IEEE Trans. PAMI. 13, 62

[O'Neill 2001] B. O'Neill. *Elementary differential geometry*. New York, Academic Press, 2001. 36

[Pan *et al.* 2006] G. Pan, S. Han, Z. Wu and Y. Wang. *Super-Resolution of 3D Face*, 2006. ECCV. 28

[Park *et al.* 2003] S. Park, M. Park and M. Kang. *Super-resolution image reconstruction : a technical overview*. IEEE Signal Processing Magazine, vol. 20, pages 21–36, 2003. 32

[Pollard *et al.* 1985] S.B. Pollard, J.E.W.Mayhew and J.P.Frisby. *A stereo correspondence algorithm using a disparity gradient constraint*. J. Perception, vol. 14, pages 449–470, 1985. 13

[Potmesil 1987] M. Potmesil. *Generating octree models of 3d objects from their silhouettes in a sequence of images*, 1987. CVGIP. 15

[Rajagopalan *et al.* 2008] A. Rajagopalan, A. Bhavsar, F. Wallhoff and G. Rigoll. *Resolution Enhancement of PMD Range Maps*. Lecture Notes in Computer Science, vol. 5096, pages 304–313, 2008. 34

[Romdhani & Vetter 2003] S. Romdhani and T. Vetter. *Efficient, Robust and Accurate Fitting of a 3D Morphable Model*, 2003. ICCV. 36

[Rosenbush *et al.* 2007] G. Rosenbush, T. Hong and R. Eastman. *Super-resolution enhancement of flash LADAR range data*. Proceedings of SPIE, vol. 6736, pages 673614.1–673614.10, 2007. 32

[Sadeghi *et al.* 2008] H. Sadeghi, P. Moallem and S.A Monadjemi. *Feature Based Dense Stereo Matching using Dynamic Programming and Color*. International Journal of Computational Intelligence, vol. 4, pages 179–186, 2008. 13

[Salvi *et al.* 2004] J. Salvi, J. Pagès and J. Batlle. *Pattern codification strategies in structured light systems*. Pattern Recognition, vol. 37, pages 827–849, 2004. 20, 22

[Sansoni *et al.* 2009] G. Sansoni, M. Trebeschi and F. Docchio. *State-of-The-Art and Applications of 3D Imaging Sensors in Industry, Cultural Heritage, Medicine, and Criminal Investigation*. J. Sensors, vol. 9, pages 568–601, 2009. 16, 19

[S.Birchfield & C.Tomasi 1999] S.Birchfield and C.Tomasi. *Multiway cut for stereo and motion with slanted surfaces*. Int'l Joint Conf. Artificial Intelligence, vol. 1, pages 489–495, 1999. 13, 16

[Scharstein & Szeliski 2001] D. Scharstein and R. Szeliski. *A Taxonomy and Evaluation of Dense Two–Frame Stereo Correspondence Algorithms*. International Journal of Computer Vision, vol. 47, pages 7–42, 2001. 12, 13

[Schmitt & Yemez 1999] F. Schmitt and Y. Yemez. *3d color object reconstruction from 2d image sequences*, 1999. IEEE Interational Conference on Image Processing. 15

[Schoen & Yau 1997] R. Schoen and S. T. Yau. *Lectures on harmonic maps*. Cambridge : International Press, Harvard University, 1997. 36

[Schuon *et al.* 2008] S. Schuon, C. Theobalt, J. Davis and S. Thrun. *High-quality scanning using time-of-flight depth superresolution*, 2008. CVPR TOF Workshop. 32

[Schuon *et al.* 2009] S. Schuon, C. Theobalt, J. Davis and S. Thrun. *LidarBoost : Depth Superresolution for ToF 3D Shape Scanning*, 2009. CVPR Conference. 19, 28, 45, 120

[Seitz *et al.* 2006] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski. *A comparison and evaluation of multi-view stereo reconstruction algorithms*. IEEE Conference on Computer Vision and Pattern Recognition, New York, pages 519–526, 2006. 26

[Sharon & Mumford 2004] E. Sharon and D. Mumford. *2D shape analysis using conformal mapping*. In IEEE computer vision and pattern recognition, vol. 2, pages 350–357, 2004. 36

[Shechtman *et al.* 2002] E. Shechtman, Y. Caspi and M. Irani. *Increasing space-time resolution in video*, 2002. ECCV. 31

[Surrel 2003] Y. Surrel. *Images optiques : mesures 2D et 3D*. 2003. 25

[Surrel 2004] Y. Surrel. *Les techniques optiques demesure de champ : essai de classification*. Instrumentation Mesure Métrologie, vol. 4, pages 11–42, 2004. 25, 147

[S.Zhang & Huang 2006] S.Zhang and P.S. Huang. *High-resolution, real-time three-dimensional shape measurement*. J. Optical Engineering, vol. 45, no. 123601, 2006. 22

[Tikhonov & Arsenin 1977] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems*. V.H. Winston and Sons, Wachington, DC, 1977. 120

[Uffenkamp 1993] V. Uffenkamp. *State of the art of high precision industrial photogrammetry*. International Workshop on Accelerator Alignment, Annecy, France, 1993. 12

[Viola & Jones 2001] P. Viola and M. Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features*. International Conference of Pattern Recognition, CVPR, 2001. 72

[Wang *et al.* 2005] S. Wang, L. Zhang and D. Samaras. *Face Reconstruction Accross Different Poses and Arbitrary Illumination Conditions*, 2005. AVBPA. 28

[Wang *et al.* 2008] Y. Wang, M. Gupta, S. Zhang, S. Wang, X. Gu, D. Samaras and P. Huang. *High Resolution Tracking of Non-Rigid Motion of Densely Sampled 3D Data Using Harmonic Maps*. Int. Journal Computer Vision, vol. 76, pages 283–300, 2008. 36, 117

[Weise *et al.* 2007] T. Weise, B. Leibe and L. Van Gool. *Fast 3D scanning with automatic motion compensation*. In IEEE computer vision and pattern recognition, 2007. 30

[Williams 1990] L. Williams. *Performance-driven facial animation*. International Journal Computer Vision, vol. 24, pages 235–242, 1990. 35

[Willomitzer *et al.* 2010] F. Willomitzer, Z. Yang, O. Arold, S. Ettl and G. Häusler. *3D face scanning with 'Flying Triangulation'*, 2010. Deutschen Gesellschaft für angewandte Optik. 24

[Willomitzer *et al.* 2011] F. Willomitzer, Z. Yang, O. Arold, S. Ettl and G. Häusler. *Options and limitations of 'Flying Triangulation'*, 2011. Deutschen Gesellschaft für angewandte Optik. 24

[Wu *et al.* 2011] C. Wu, B. Wilburn, Y. Matsushita and C. Theobalt. *High-quality shape from multi-view stereo and shading under general illumination*, 2011. IEEE Computer Vision and Pattern Recognition. 29

[Xiong & Shafer 1993] Y. Xiong and S. Shafer. *Depth from focusing and defocusing*, 1993. IEEE Computer Vision and Pattern Recognition. 17

[Y. Ohta & Sakai 1981] K. Maenobu Y. Ohta and T. Sakai. *Obtaining surface orientation of from texels under perspective projection*. IJCAl, vol. 2, pages 746–751, 1981. 16

[Yin *et al.* 2006] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato. *A 3D facial expression database for facial behavior research*, 2006. IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA. 126

[Yin *et al.* 2008] L. Yin, X. Chenand Y. Sun, T. Worm and M. Reale. *A high-resolution 3d dynamic facial expression database*, 2008. IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands. 126

[Yoshiyasu & Yamazaki 2011] Y. Yoshiyasu and N. Yamazaki. *Topology-adaptive Multi-view Photometric Stereo*, 2011. IEEE Computer Vision and Pattern Recognition. 30

[Zhang & Yau 2007] S. Zhang and S. Yau. *Generic nonsinusoidal phase error correction for three-dimensional shape measurement using a digital video projector*. J. Applied Optics, vol. 46, pages 36–43, 2007. 23, 90

[Zhang & Yau 2008] S. Zhang and S. Yau. *Absolute phase-assisted three-dimensional data registration for a dual-camera structured light system*. J. Applied Optics, vol. 47, pages 3134–3142, 2008. 22, 26

[Zhang *et al.* 1999] R. Zhang, P. Tsai, J. Cryer and M. Shah. *Shape from shading : A survey*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, pages 690–706, 1999. 17

[Zhang *et al.* 2003] L. Zhang, B. Curless and S. M. Seitz. *Spacetime Stereo : Shape Recovery for Dynamic Scenes*, 2003. CVPR. 21, 30, 31, 147

[Zhang *et al.* 2004] L. Zhang, N. Snavely, B. Curless and S. M. Seitz. *Spacetime Faces : High-resolution capture for modeling and animation*. SIGGRAPH, 2004. 26

[Zhang 1999] Z. Zhang. *Flexible Camera Calibration by Viewing a Plane from Unknown Orientations*, 1999. ICCV Conference. 56, 84

[Zhang 2005] S. Zhang. *High-resolution, Real-time 3-D shape measurement*. Thesis dissertation, 2005. 25

[Zhang 2010] S. Zhang. *Recent progresses on real-time 3D shape measurement using digital fringe projection techniques*. J. Optics an Lasers in Engineering, vol. 48, pages 149–158, 2010. 20, 22, 88