

ÉCOLE DOCTORALE INFORMATIQUE ET INFORMATION POUR LA SOCIÉTÉ

THÈSE

pour obtenir le grade de
DOCTEUR EN
INFORMATIQUE DE L'ÉCOLE CENTRALE DE LYON



présentée et soutenue publiquement par

Boulbaba BEN AMOR

le 08 DÉCEMBRE 2006

**Contributions à la reconstruction, la reconnaissance
et l'authentification faciale 3D**

préparée au sein du laboratoire LIRIS

sous la direction de

Liming CHEN

Mohsen ARDABILIAN

COMPOSITION DU JURY

XXX	Président	(XXX)
XXX	Rapporteur	(XXX)
XXX	Rapporteur	(XXX)
XXX	Examineur	(XXX)
M. Liming CHEN	Directeur de thèse	(Professeur, École Centrale de Lyon)
M. Mohsen ARDABILIAN	Co-directeur de thèse	(Maître de conférences, École Centrale de Lyon)

Table des matières

Table des matières	iv
1 Introduction	1
1.1 Contexte et problématiques	1
1.2 Objectifs et contributions	3
1.3 Organisation de la thèse	4
2 État de l’art en modélisation et reconnaissance faciales 3D	7
2.1 Introduction	7
2.2 Technologies de numérisation tridimensionnelle	8
2.2.1 Techniques de reconstruction actives	10
2.2.2 Techniques de reconstruction passives	16
2.3 Identification/vérification de personnes par visages 3D	18
2.3.1 2D vs. 2D	19
2.3.1.1 Méthodes de sous-espace (<i>subspace methods</i>)	20
2.3.1.2 Méthodes basées sur des attribues géométriques (<i>geometric feature-based methods</i>)	21
2.3.2 2D vs. 2D via 3D	22
2.3.3 3D vs. 3D	23
2.3.3.1 Méthodes de sous espace (<i>subspace methods</i>)	23
2.3.3.2 Méthodes basées sur l’alignement de surfaces	23
2.3.3.3 Méthodes basée sur le calcul de courbure	24
2.3.3.4 Méthodes basées sur le calcul géodésique (<i>bending-invariant canonical form</i>)	25
2.3.4 Bi-modalité 2D+3D	26
2.4 Conclusion	27
3 Reconstruction faciale 3D par vision active et modélisation géométrique	29

3.1	Introduction	29
3.2	Vue d'ensemble de l'approche proposée	30
3.3	Calibration du capteur stéréo	33
3.3.1	Modèle mathématique de caméra	33
3.3.1.1	Transformation monde/camera	34
3.3.1.2	Projection 3D-2D	35
3.3.1.3	Prise en compte des distorsions	35
3.3.1.4	Transformation camera/image	36
3.3.1.5	Transformation globale	36
3.3.2	Au-delà d'une caméra	36
3.3.2.1	Calibrage stéréo	37
3.3.2.2	Relation droite/gauche	37
3.3.3	Méthode de calibration	37
3.3.3.1	Classification des méthodes de calibration	38
3.3.3.2	Méthode de calibration	39
3.4	Stratégie de mise en correspondance stéréo	40
3.4.1	Géométrie épipolaire et rectification géométrique	41
3.4.2	Lumière et détection sous-pixélique de frontières de franges	42
3.4.3	Appariement par programmation dynamique	44
3.4.3.1	Formulation mathématique du problème	46
3.4.3.2	Application à l'appariement stéréo	47
3.5	Modélisation partielle du visage	49
3.5.1	Triangulation optique	50
3.5.2	Interpolation via des modèles de <i>splines cubiques</i>	51
3.5.3	Génération de la forme 3D	53
3.6	Modélisation complète du visage	55
3.6.1	Recalage grossier des différentes vues	58
3.6.2	Recalage fin moyennant <i>ICP</i>	58
3.7	Résultats et procédure de validation	59
3.8	Conclusion	63
4	Reconnaissance/Authentification de personnes via leurs surfaces faciales	65
4.1	Introduction	65
4.2	Vue d'ensemble de l'approche proposée	67
4.3	Appariement de surfaces 3D via <i>ICP</i>	69
4.3.1	Présentation de l'algorithme <i>ICP</i>	70

4.3.2	Initialisation de l'algorithme <i>ICP</i>	78
4.3.2.1	Utilisation de points d'intérêts	79
4.3.2.2	Utilisation des barycentres	80
4.3.3	Notre variante de l'algorithme <i>ICP</i> : <i>R-ICP</i>	80
4.4	Segmentation du visage 3D en régions	82
4.4.1	Étude anatomique du visage	82
4.4.2	Influences des muscles et des régions sur la forme faciale 3D	86
4.4.3	Schéma de segmentation adapté	87
4.5	Calcul de carte géodésique en vue de la reconnaissance	90
4.5.1	Calcul Géodésique	90
4.5.1.1	Définition d'une géodésique	91
4.5.1.2	Théorie de Sethian	91
4.5.1.3	Algorithme de Cheminement Rapide	92
4.5.1.4	Cheminement Rapide sur une triangulation aiguë	93
4.5.1.5	Extension vers une triangulation quelconque	95
4.5.2	Application de l'algorithme du cheminement rapide dans le cas des surfaces faciales 3D	97
4.6	Conclusion	99
5	Expérimentations, évaluations et discussions	101
5.1	Introduction	101
5.2	Protocole d'expérimentation et d'évaluation	102
5.2.1	Nouvelle base de données <i>ECL-IV2</i>	102
5.2.2	Comparaison avec la base <i>FRGC</i>	108
5.2.3	Scénarios des expérimentations	109
5.2.4	Critères de performances	110
5.3	Résultats obtenus et discussions	111
5.3.1	Algorithme basé sur <i>ICP</i>	112
5.3.2	Algorithme basé sur notre variante <i>R-ICP</i>	115
5.3.3	Algorithme basé sur le calcul de cartes géodésiques	121
5.3.4	Comparaison des algorithmes <i>ICP</i> , <i>R-ICP</i> et cartes géodésiques	124
5.4	Conclusion	128
6	Conclusion et perspectives	129
	Bibliographie	133

Chapitre 1

Introduction

1.1 Contexte et problématiques

Depuis les attentats du 11 septembre 2001 aux États-unis, et la médiatisation qui en a été l'objet, les restrictions sur les frontières se sont élargies. Les pays, notamment européens et les états américaines, souhaitent mieux connaître et s'assurer de l'identité des voyageurs, surtout en matière de flux migratoires. Cependant, leurs politiques de déploiement de la sécurité des citoyens et des services n'ont pas attendu ses événements pour émerger. En effet, avec la croissance exponentielle des communications, tant en volume qu'en diversité (déplacement physique, transaction financière, accès aux services, etc.), les préoccupations internationales ont fait ressortir ces nécessités en vue de lutter contre la fraude, la crime et le terrorisme. Face à ces responsabilités politiques, d'une part, et à ce besoin de protection civile d'autre part, les systèmes d'identification et d'authentification de personnes connaissent un intérêt grandissant et suscitent un vif intérêt dans différents domaines, tels que le secteur bancaire, les services publics, ainsi que dans le domaine de transport et notamment de voyages. Leur dénominateur commun, est d'offrir des moyens simples, pratiques, fiables, pour vérifier l'identité d'une personne, sans l'assistance d'une autre personne.

Afin de résoudre ces problèmes et de répondre à ces besoins liés à la sécurité, la biométrie se présente comme une technologie puissante. En effet, les différents moyens biométriques sont basés sur des caractéristiques comportementales et/ou physiologiques spécifiques à chaque personne. Ces caractéristiques présentent l'avantage qu'elles sont universelles, uniques, permanentes et qu'elles ne peuvent être falsifiées à la différence des moyens classiques tels que les mots de passe et les badges qui peuvent être utilisés frauduleusement par d'autres individus. Empreintes digitales, numérisation de l'iris, géométrie de la main ou analyse comportementale, plusieurs de ces mesures ont été techniquement prouvé et commercialisées. Cependant, ils exigent, généralement, la coopération des utilisateurs qui les trouvent intrusifs. Ce caractère contraignant oblige la personne à se plier aux exigences techniques du procédé, ce qui constitue généralement un gage d'efficacité et d'acceptation.

Dans ce contexte, la reconnaissance faciale s'avère bien adaptée pour faire face à cette limitation puisque la coopération ou l'apprentissage des participants n'est pas exigée. Elle apparaît donc comme une alternative appropriée, à condition que des taux de reconnaissance suffisants soient garantis. Une raison pour laquelle cette modalité biométrique a suscité récemment une attention substantielle de la part de l'ensemble des acteurs du monde de la reconnaissance de visage, laboratoires de recherche mais aussi industriels.

Les premières études théoriques, dans ce domaine, remontent au début des années 1970 et le passage vers des produits commerciaux a reçu une impulsion décisive à partir des années 1994-1996 grâce à la mise en oeuvre du programme FERET (Face Recognition Technology), organisé par le ministère de la Défense américaine (Department of Defense, DoD). À l'issue de ces tests d'évaluation du programme FERET, on disposait d'une base d'images de référence. Depuis la fin du projet en 1996, le grand changement est l'apparition sur le marché de produits commerciaux. La grande compétitivité du marché a fait éclore un grand nombre d'algorithmes de reconnaissance de visage ou de variantes, dont la plupart n'étaient même pas présents lors des tests d'évaluation FERET. Le DoD décida alors de lancer les programmes FRVT successivement en 2000 et en 2002 (Facial Recognition Vendor Test 2000/2002) dont l'objectif était d'évaluer les performances des produits commerciaux. Les résultats publiés, à l'issue de ces campagnes, ont montrées que les algorithmes développés souffrent encore de plusieurs problèmes posés par les changements dans l'apparence des individus dû à l'éclairage, à la posture, aux expressions faciales, à l'âge, à la coiffure et au maquillage, aux accessoires portés, etc.

Comme tout autre problème de reconnaissance des formes, les problèmes liés à la reconnaissance de visages peuvent être catégorisés en deux catégories : (a) *la variabilité intra-classes* et (b) *la similarité inter-classes*. La première catégorie présente, d'abord, les variations intrinsèques au visage dus aux changements d'expressions faciales, de la physionomie du visage, au phénomène de vieillissement et aux interventions chirurgicales. Cette variabilité englobe, également, les variations environnementales qui sont dus aux changements des conditions d'éclairage, l'orientation du visage de la personne et à l'occlusion par des accessoires ou par d'autres objets. La seconde catégorie présente la similarité entre classes d'individus. Ici, les apparences sont similaires dus à des liens familiaux (exemples : jumeaux, jumelles, père/fils, mère/fille, etc.) ou plus généralement l'appartenance à la même race.

Ces limitations, qui freinent le niveau de maturité de cette technologie biométrique, ont entraîné le développement d'autres formes d'acquisitions telles que : les cartes de profondeurs (range images), les images infrarouge (IR images) et les images de hautes résolutions en vue de surmonter ces problèmes difficiles à résoudre avec des algorithmes de traitement classiques agissant sur l'information texture. C'est dans ce contexte que s'inscrivent les travaux de recherche de cette thèse. Nous essayons de traiter les problèmes liés à la reconnaissance faciale par l'introduction d'une nouvelle représentation du visage par sa forme tridimensionnelle. Notre principale motivation est l'insuffisance de l'information texturale 2D, très sensible aux variations, essentiellement, d'illumination et de pose.

Nous nous basons donc, dans notre étude sur les surfaces faciales comme supports de reconnaissance, en vue de calculer les scores de similarités entre les visages. Par ailleurs, outre

le problème de reconnaissance, nous avons abordé l'aspect acquisition de la forme 3D du visage. Le problème, ici, est de reproduire le plus fidèlement possible un modèle numérique du modèle réel. Devant le coût très élevé des scanners lasers 3D considérés comme le moyen le plus adapté pour avoir des bonnes reconstructions, d'une part, et l'imprécision des mesures des approches de reconstruction passives, d'autre part, nous essayons de proposer, également, des contributions à ce sujet.

1.2 Objectifs et contributions

La mise en oeuvre des programmes d'évaluation, et l'apparition des bases de données, appelées de référence, accompagnées de leurs protocoles d'évaluation appropriés, a énormément contribué au développement des algorithmes de reconnaissance faciale. Ainsi, l'alimentation de la communauté par des rapports, des discussions menées et des conclusions tirées, à l'issue des évaluations des différents approches et systèmes commerciaux, a aidé à aborder cet axe de recherche à différents niveaux. C'est, également, le cas de nos travaux qui s'appuient sur ces rapports et les considère comme le point de départ, notamment le plus récent celui du projet FRVT 2002.

L'objectif principal de cette thèse est, donc, d'analyser l'existant, en se référant à ce type de documents, et de proposer des solutions aux problèmes. La principale innovation réside dans la substitution de l'image de texture par celle de forme tridimensionnelle du visage. En réalité, cette nouvelle modalité, utilisée seule ou couplée avec l'image de texture, a reçu une vraie impulsion, au sein de la communauté, au même moment que nous avons débuté le présent travail. Cependant, à la différence de la majorité des travaux présents dans la littérature, nous avons abordé les deux problématiques d'acquisition et de reconnaissance faciale 3D. Ces études menées d'une façon couplée, nous ont permis d'acquérir des idées complémentaires et une compréhension plus élaborée en matière de données 3D. Les principales contributions présentées touchent donc à ces deux axes de recherches :

En premier lieu, une chaîne de reconstruction 3D, de bout en bout, dédiée aux visages humains a été mise en place. Cette technique est basée sur la stéréo-vision active (utilisation d'une lumière texturante projetée sur le visage) et profite des outils mathématiques de la modélisation géométrique 3D pour proposer un nouveau schéma hybride plus efficace que l'existant. Ce mariage entre la vision active et la modélisation graphique nous a permis de résoudre le problème de coût très élevé des scanners 3D à base de balayage laser tout en garantissant une bonne qualité géométrique de reconstruction. Nous nous intéressons, plus particulièrement dans cette partie, au problème de mise en correspondance stéréo, par la formulation mathématique du problème et sa résolution moyennant une technique d'optimisation adaptée, à savoir la programmation dynamique (*DP : Dynamic programming*). Un passage en dessous du pixel dans les mesures a été, également, étudié et proposé pour plus de précision en matière d'information de profondeur. En outre, nous étudions les techniques d'interpolation géométrique en vue de compléter le schéma de reconstruction. Cette partie de nos travaux a fait l'objet de plusieurs publications présentant la démarche scientifique de notre approche [BAC05a][BAC05c][BAC06b][BAC05b].

En second lieu, une approche de matching de visages 2 1/2D vs. 3D a été mise en place pour calculer le score de similarité entre deux visages via leurs surfaces faciales. Elle est basée sur l'algorithme du plus proche voisin itéré (*ICP : Iterative Closest Point*). Cette première approche fait face aux problèmes de variations de conditions d'éclairage et de posture [BOAC05][BAC06c]. En effet, la disponibilité d'un modèle complet 3D de visage dans la galerie de reconnaissance permet d'effectuer le matching quelque soit l'angle de prise de vue de l'image requête. En plus, une abstraction est faite sur l'information texturale ce qui permet d'éliminer le problème de changement d'éclairage. Cette approche est améliorée, par la suite, pour plus d'invariance aux changements d'expressions faciales et tout changement dans la physiologie du visage (Barbe, moustache, etc.). Cette amélioration consiste à introduire une nouvelle métrique orientée région dans l'algorithme de matching 3D [BAC06a]. Elle consiste à segmenter le visage en régions qui ont des influences variables sur la déformation de la forme 3D du visage. Par conséquent, elle attribue plus de poids aux régions statiques qu'aux régions moins statiques dans le visage.

Enfin, nous avons développé l'une des premières bases de visages 3D et son protocole pour l'évaluation de nos approches et d'autres. Nous avons essayé de répondre aux besoins, en matière d'évaluation et de mesures de performances, d'un spectre large d'approches en mettant en place une base de données variées (2D, 2 1/2D et 3D) permettant l'évaluation des algorithmes développés et leur invariance par rapport aux challenges de la technologie. Le développement de cette base de données a commencé au sein du projet européen BioSecure par une première campagne d'acquisition menée durant le workshop résidentiel organisé à Paris en août 2005 et prend suite dans le cadre du projet national technovision *IV*² (Identification par l'Iris et le Visage via la Vidéo)¹. Nous avons mis au point, dans le cadre de ce projet et en partenariat avec des acteurs académiques et industriels un nouveau protocole d'évaluation prenant en compte l'évolution des travaux dans ce domaine au niveau national et international.

1.3 Organisation de la thèse

Le présent manuscrit s'articule autour de quatre parties essentielles, outre la présente introduction et la conclusion, qui seront présentées de la manière suivante :

Afin de poser correctement les problématiques traitées dans cette thèse, la première partie est consacrée à l'étude de l'état de l'art des deux domaines qui nous intéressent : la modélisation 3D de visages et leurs reconnaissances. Cet état de l'art présente un survol sur les principales méthodes de numérisation tridimensionnelles et les méthodes de modélisation faciales qui en découlent ainsi que les méthodes récentes d'identification/authentification de personnes par visage 3D. Elle dévoile les différentes techniques proposées au sein de la communauté en soulignant les limites et les avantages de chacune.

La deuxième partie de cette thèse est consacrée à la description de notre approche hybride de reconstruction 3D de visage par vision active et modélisation géométrique. Nous

¹<http://lsc.univ-evry.fr/techno/iv2/>

abordons dans cette description toute la chaîne de modélisation allant de la calibration des caméras à la modélisation géométrique 3D en mettant l'accent sur les contribution scientifique du travail effectué. Nous achevons cette partie par présenter une procédure de validation et une discussion traitant les évaluations quantitatives et qualitatives des résultats de l'approche proposée.

La troisième partie de cette thèse porte sur l'exploitation de la forme 3D du visage comme un moyen biométrique en vue de reconnaître ou authentifier des personnes. Nous présentons, dans ce chapitre, un nouveau paradigme efficace en vue de surmonter les variations intrinsèques et extrinsèques au visages considérées comme sources d'imprécision des systèmes existants. Nous présentons, essentiellement, un nouvel algorithme de matching de surfaces faciales orientée région, que nous appelons *R-ICP*, et qui dérive de l'algorithme du plus proche voisin itéré : *ICP*. Une bonne partie est consacrée à l'étude de l'anatomie du visage en vue de mener une compréhension sur cet objet très complexe et de concevoir un schéma de segmentation unifié et robuste.

En fin, dans la quatrième partie, nous présentons, dans un premier temps, la nouvelle base de visages 3D *ECL-IV²*, puis, les résultats obtenus en matière d'identification et d'authentification avec une discussion traitant tous les aspects liés au mesure de performance des systèmes biométrique. Cette partie comporte, également, une comparaison de des techniques développées avec une technique à base de calcul géodésique effectué sur les surfaces faciales après matching.



Chapitre 2

État de l'art en modélisation et reconnaissance faciales 3D

2.1 Introduction

Initialement motivée par des applications robotiques, la vision par ordinateur (*computer vision*) voulait doter les machines d'une vision semblable à celle de l'homme [FP02, TV98]. En effet, l'espace qui nous entoure a une structure tridimensionnelle (3D). Lorsque l'on demande à une personne de décrire ce qu'elle voit, elle n'éprouve aucune difficulté à identifier les objets. Et pourtant l'information qui est réellement disponible sur la rétine de ses yeux n'est qu'une collection de points, en chaque point il y a tout simplement une information qui donne une indication quant à la quantité de lumière et la couleur qui proviennent de l'espace environnant qui ont été projetées à la rétine. Guidé à la fois par des informations codées dans la rétine et par ses propres connaissances, le processus visuel construit des percepts. De plus, il n'y a pas de correspondance terme à terme de l'information sensorielle et la réponse finale (une scène avec des objets 3D).

Aujourd'hui, les applications dans ce domaine ne sont plus seulement robotiques mais couvrent un spectre plus large comme la reconstruction 3D, la reconnaissance des formes, le contrôle industriel, l'imagerie médicale (Voir figure 2.1 [SW95]), l'imagerie satellite et plus récemment la réalité virtuelle, l'archéologique et l'encodage de la vidéo. Ces domaines d'applications nécessitent le plus souvent l'extraction des indices 3D à partir des projections 2D issues d'une ou de plusieurs caméras et donc conduisent à une reconstitution de la structure tridimensionnelle des objets tels que des terrains, des villes, des bâtiments, des objets de la vie quotidienne et même des corps et des visages humains. Les algorithmes développés visent, par conséquent, l'interprétation non ambiguë d'une scène réelle 3D. Les caméras effectuent une opération de projection qui transforme le monde tridimensionnel en une représentation bidimensionnelle. Les approches ainsi développées tentent à répondre à la question suivante : Comment peut-on, dans ces conditions, retrouver la dimension perdue ?

Récemment, plusieurs techniques de reconnaissance de formes veulent étendre leurs

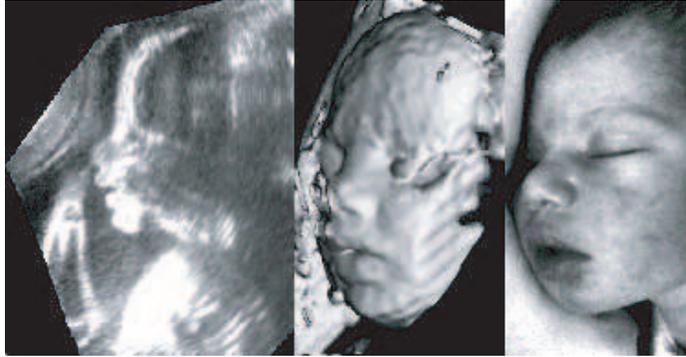


FIG. 2.1 – Reconstruction ultrasonore d'un fœtus (gauche), de la surface 3D (milieu) et la photo du bébé 24h après l'accouchement (droite).

systèmes de vision 2D aux systèmes 3D. Leurs motivations se manifestent dans le complément d'information fourni par ces capteurs. En effet, la disponibilité de l'information 3D permet de palier à plusieurs problèmes rencontrés en 2D grâce à une interprétation plus complète des formes sujettes de la reconnaissance. La reconnaissance faciale, un des sujets de reconnaissance des formes très connus et largement abordés dans la littérature, tente de migrer, ces quelques dernières années, d'une représentation 2D à une représentation 3D. Son objectif est de faire face aux problèmes dus à l'utilisation d'un modèle 2D (image couleur ou d'intensité), qui reste l'handicap majeur de ne pas atteindre des taux de reconnaissance élevés. Dans ce contexte, plusieurs techniques de numérisation tridimensionnelle ont été utilisées dans le cas du visage et d'autres techniques dédiées ont été élaborées pour ce type de problèmes et d'autres.

Dans ce chapitre nous donnons un état de l'art couplé entre la modélisation 3D du visage et la reconnaissance basée sur cette nouvelle représentation. Dans la première partie, nous nous intéressons au problème de numérisation 3D et son application dans le cas du visage. Dans la deuxième partie, c'est le problème de reconnaissance/authentification via les données 3D qui est abordé. Dans chacune des deux parties, nous donnons une taxonomie des approches, des techniques, des algorithmes et des systèmes industriels existants. Ces taxonomies présentent, en quelque sorte, l'organisation des deux parties de ce chapitre. Nous finirons par donner un résumé et un bilan de l'existant en vue de démarquer le travail réalisé dans cette thèse par rapport à l'existant.

2.2 Technologies de numérisation tridimensionnelle

La photographie 3D est l'un des domaines clés en vision par ordinateur et en modélisation graphique vu qu'il les combine dans un même framework. Ce domaine trouve ses champs d'application en inspection industrielle, en rétro-projection (*reverse engineering*), en archéologie,

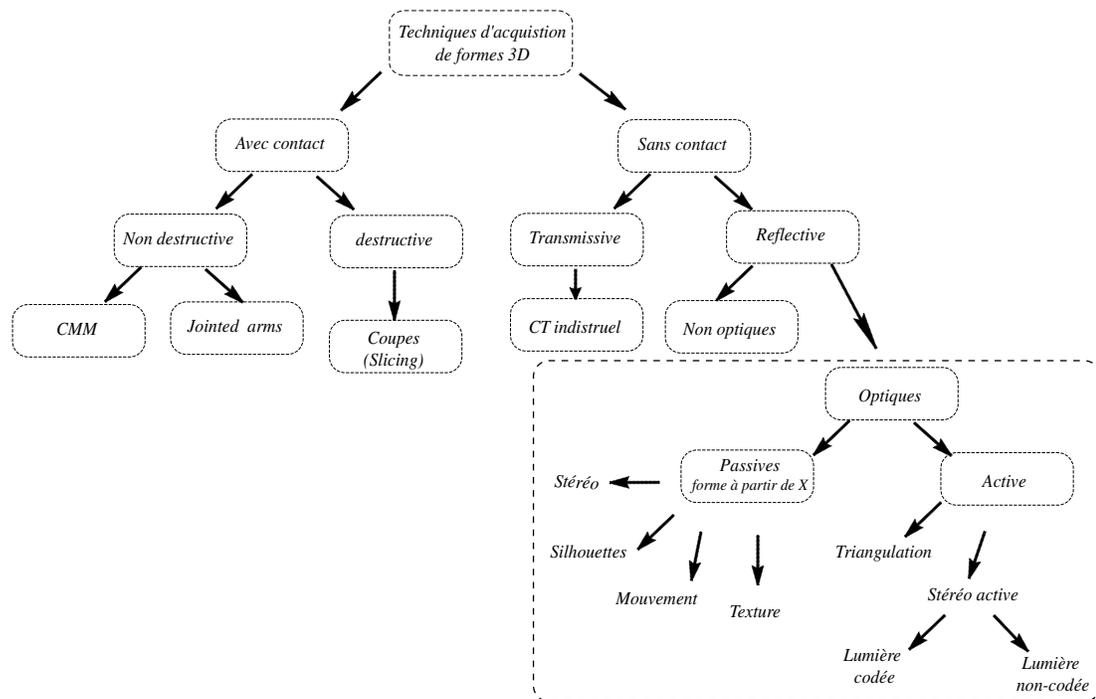


FIG. 2.2 – Taxonomie des techniques d'acquisition de formes 3D.

en animation d'objets 3D, etc. Plusieurs approches ont été élaborées pour répondre aux besoins de chaque type d'application et selon le type des objets à numériser (statique/dynamique, déformable/rigide, etc.). La figure 2.2 donne une classification quasi-complète des méthodes existantes dans le domaine d'acquisition de formes 3D. Elle distingue les méthodes avec contact (destructive, non destructive) des méthodes sans contact (transmissive et reflective). Dans le présent survol, nous restreindrons notre étude aux méthodes dites optiques (faisant partie des méthodes sans-contact avec une approche reflective). Elle nous semblent être la plus adaptée à nos besoins en terme de reconstruction faciale 3D en vue de la reconnaissance. Les méthodes de cette catégorie peuvent être classées elles même en deux sous-catégories qui sont : *les méthodes actives* et *les méthodes passives*. Les deux sections qui suivent dévoileront les techniques développés au sein de ces deux sous-catégories en citant les avantages et les limitations de chacune d'elles. Nous abordons également l'adaptation de ces approches dans le cas du visage.

2.2.1 Techniques de reconstruction actives

La vision active consiste à combiner un ou plusieurs capteur(s) photographique avec une source lumineuse, afin de mesurer les coordonnées tridimensionnelles des points sur la surface d'un objet. Dans cette catégorie d'approches, deux types de sources lumineuses sont généralement employées : le laser [Bla04] ou un schéma de lumière structurante [SPB04]. Le laser, une lumière monochromatique, qui en passant à travers une lentille cylindrique, crée un plan lumineux éclairant une tranche de la scène. Ceci permet de n'observer par une caméra que la partie de la scène ainsi éclairée (voir figure 2.3 (a)). Connaissant l'équation du plan lumineux dans l'espace de la scène et ses coordonnées dans l'image, on peut alors calculer la position dans l'espace de la tranche de la scène captée par la caméra associée au dispositif laser. Ce principe appelé *triangulation active* est exploité dans plusieurs produits industriels dont les plus connus sont ceux de *Minolta*¹ (en Europe) et de *Cyberware*² (en Amérique du Nord) appelée en anglais (*laser range finder*). Ces équipements dotés d'un miroir rotatif associé au émetteur de plan laser, effectuent un balayage en vue de mesurer l'information 3D sur la totalité de l'objet observé. Ces caméras 3D sont les plus utilisées dans les domaines comme la rétro-projection (*reverse engineering*) et le contrôle de qualité (*quality control*) effectués sur des pièces mécaniques, Voir figure 2.3 (b). Elles ont remplacé les techniques d'acquisition avec contact telle que la machine de mesure de coordonnées (*CMM*) qui étaient très utilisées en contrôle industriel de pièces mécaniques. Ils présentent des bonnes performances d'acquisition en terme de qualité de rendu reconstruit, de densité et de précision de mesures, mais souffrent de quelques faiblesses. Ces faiblesses sont dues à la discontinuité de la réflectance sur la surface de l'objet, la variation de la forme (sa pose), les occlusions qui peuvent empêcher le trait laser à arriver à la surface de l'objet ou par simplement l'absorption du trait laser par la surface de l'objet à numériser. Ces équipements souffrent également de leurs prix très élevés et donc restreints à des applications pointues et très particulières.

¹<http://www.konicaminolta.com/>

²<http://www.cyberware.com/>

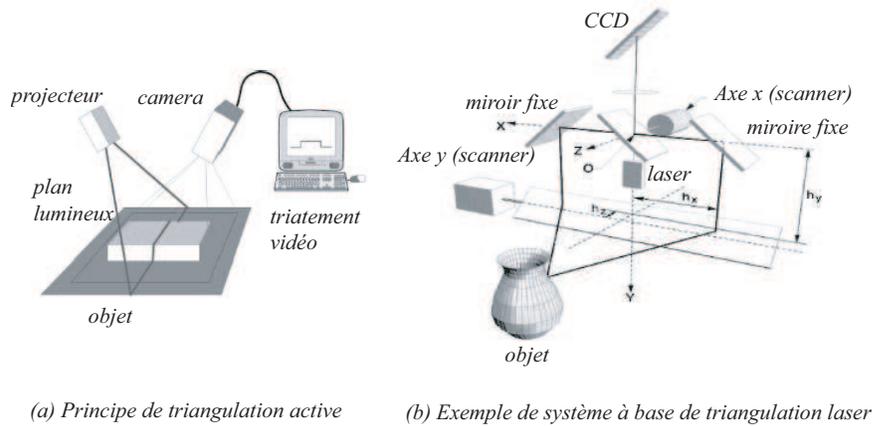
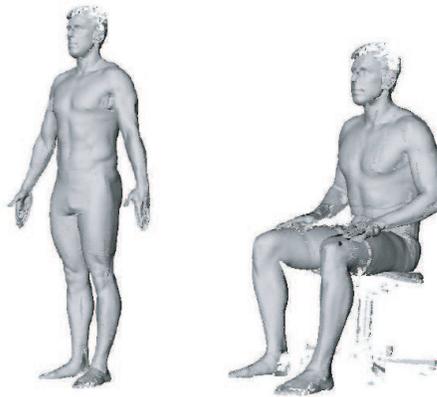


FIG. 2.3 – Principe d’acquisition 3D par triangulation active.

De nouveaux champs d’applications de ces méthodes ont vu le jour très récemment, comme l’étude anthropométrique d’une population et la reconnaissance de visages 3D. En effet, une base d’acquisition 3D de corps humain a été développée par NIST³, appelée *CAESAR anthropometric database* pour une étude plus complète. La figure 2.4 présente un extrait de cette base. Elle contient plus de 2000 corps d’hommes et de femmes âgés entre en 18 et 65 ans en vue de mesurer les variabilités anthropométriques de la population d’amérique du nord et d’Europe.

FIG. 2.4 – Extrait de la base de corps humains 3D *CAESAR anthropometric database*.

³National Institute of Standards and technology

Le même type de capteur, à base de triangulation laser, a été utilisé récemment pour reconstruire des visages humains pour des applications différentes. Par exemple, dans [BV03] les auteurs présentent une base de visages 3D d'une centaine d'individus acquise moyennant un scanner Cyberware. Cette base a été utilisée pour synthétiser un modèle 3D générique et déformable. Elle 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 à la suite d'une étape de mise en correspondance dense de tous les visages 3D de la base avec un visage référence moyennant le flot optique. Ce modèle unifié a été exploité par la suite dans la reconnaissance des personnes [BRV02, BV03]. Une phase d'adaptation se fait par l'intermédiaire d'une approche d'analyse par synthèse. Elle consiste à chercher les paramètres de forme et de texture correspondants à un modèle 3D qui ressemble le plus à l'image 2D introduite du visage test. Les visages présentés dans leurs approches sont de très bonne qualité et en forme et en texture.

Une deuxième base de visages a été publiée par NIST⁴ dans le cadre du projet FRGC⁵ [PFS⁺03] comportant des vues frontales des visages et intégrant les deux canaux de forme et de texture pour chaque sujet. L'objectif du projet est l'évaluation des approches de reconnaissance faciales 2D et 3D sur un corpus d'images de taille significative et riche de variations d'éclairage et d'expressions faciales. L'appareil utilisé pour l'acquisition est le modèle VI900/910 de KonicaMinolta⁶. Ces appareils spécialisés produisent une numérisation de bonne précision mais ils nécessitent une étape de prétraitement (intervention manuelle ou automatique) pour éliminer le bruit présenté par des pics et de boucher les trous au niveau des maillages résultants. Ces phénomènes sont présents surtout dans les régions chevelues du visage (sourcils, barbes, moustache) et les régions des yeux où le plan laser envoyé est très peu réfléchi. Le même type d'appareil (VI300) est utilisé dans le cadre du projet national technovision *IV*² pour l'acquisition 3D des visages en vue d'évaluer les algorithmes de reconstruction et de reconnaissance. La figure 2.10 présente deux extraits des bases de données *IV*² et FRGC V.2, elle montre les problèmes d'occlusion, de bruit d'acquisition et d'absorption du trait laser par quelques régions du visage traduit par des trous sur le maillage obtenu.

La seconde grande famille de méthodes actives utilise pour la numérisation un motif lumineux qui substitue le laser. Ces méthodes projettent, sur la scène entière, un motif structurant qui est souvent une grille (ou des nœuds espacés) ou des bandes de lumière parallèles non-codées ou codées différemment [PA82]. La géométrie de la scène est obtenue en analysant les déformations du motif lumineux, capturé par la caméra, par rapport au motif d'origine ou en dé-codant la succession de lumières envoyées sur l'objet à numériser, respectivement. Par ailleurs, dans le cas de la projection d'un motif simple sur la scène (trait noir sur un fond blanc), il est relativement simple de calculer les coordonnées 3D à partir de la déformation du trait sur la scène. En effet, le problème de mise en correspondance, entre le motif d'origine et le motif résultant, n'est pas présent puisque une et une seule primitive, si elle existe, est considérée pour la triangulation pour chaque ligne de l'image capturée.

⁴National Institute of Standards and Technology

⁵Face recognition Grand Challenge

⁶<http://www.konicaminolta.com/>

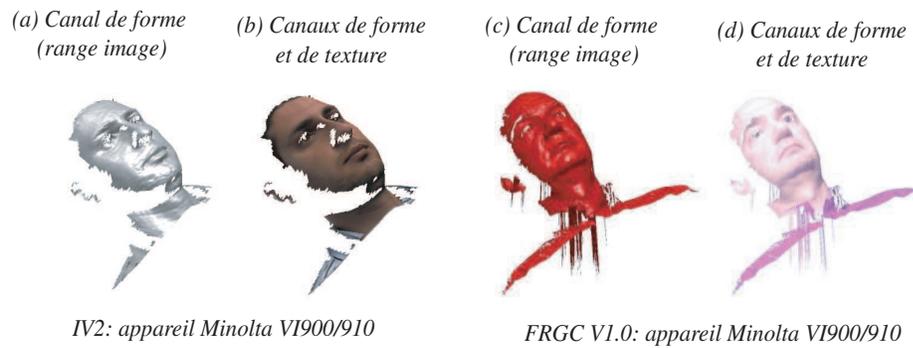


FIG. 2.5 – Extrait des bases de visages IV2 et FRGC V1.0.

Pour couvrir la totalité de la scène observée par le capteur, il suffit de répéter la procédure dans le temps en modifiant à chaque fois la position du trait dans le motif lumineux utilisé [PA82]. La deuxième alternative consiste à concevoir un motif plus complexe qui peut être une répétition dans l'espace du trait dans la direction verticale, horizontale ou dans les deux directions (grille) [GDD01]. Dans ce cas le problème n'est plus simple et consiste à mettre en correspondance les primitives du motif d'origine avec les primitives localisée sur l'objet. Une stratégie d'appariement doit donc être mise en place en prenant en considération les phénomènes d'ambiguïté et d'occlusion, sources de complexité de ce problème.

Un spectre très large de travaux basés sur la projection de motifs lumineux a été élaboré depuis les années 80 jusqu'à présent. Chaque approche propose un schéma de lumière couplé à une méthode de mise en correspondance en vue de reconstruire un objet ou une scène donnée. Dans ce survol, nous présentons quelque uns de ces travaux qui ont marqué le développement de ces approches dans la littérature. Le lecteur pourra se reporter à [SPB04] pour un état de l'art plus complet sur la vision active et les différents schémas utilisés jusqu'à l'année 2002. Cet article est doublé d'une bonne comparaison expérimentale des résultats de reconstructions effectuées sur un buste de cheval⁷.

Une ces techniques qui a marqué le parcours de développement de plusieurs approches dans la littérature est la projection d'une succession de lumières complémentaires (bandes noires sur fond blanc). Un motif envoyé sur l'objet permet d'assigner aux pixels de l'image, dans un premier temps, des bits binaires (exemple : pixel éclairé : bit=1, pixel sombre : bit=0). Puis, un mot binaire est formé pour chaque pixel par projections de plusieurs schémas de lumière espacées dans le temps (*temporal coding*). La projection de n lumières, successivement, produit 2^n codes, par conséquent la résolution de la reconstruction augmente en fonction du nombre de lumières projetés. Posdamer & Altschuler [PA82] sont les premiers à avoir proposer de projeter n modèles de lumières pour coder 2^n franges.

⁷<http://eia.udg.es>

Ce principe était très apprécié par la communauté et le même schéma a été repris par plusieurs chercheurs comme dans [BBK05]. Dans [BBK05], l'approche est appliquée aux visages humain dans un contexte de reconnaissance faciale 3D. Une amélioration de ce schéma a été proposée par Güring dans [Güh01]. L'objet est d'abord subdiviser en huit régions par projection successive de trois modèles de lumières par le schéma de Posdamer & Altschuler, puis, un nouveau modèle périodique est introduit et bougé dans une direction donnée (*phase shifting*) dans le but d'augmenter la résolution du système d'acquisition. Le principe de base étant de minimiser le pas d'échantillonnage, une dé-codification plus robuste est donc déduite pendant le passage de la deuxième lumière. Le problème majeur de ces approches est le nombre de motifs lumineux à projeter sur l'objet pour obtenir de bonnes résolutions de modèles numérique. Pour cela, ce type de codage est utilisé pour des objets statiques ou en général des scènes statiques et non déformable.

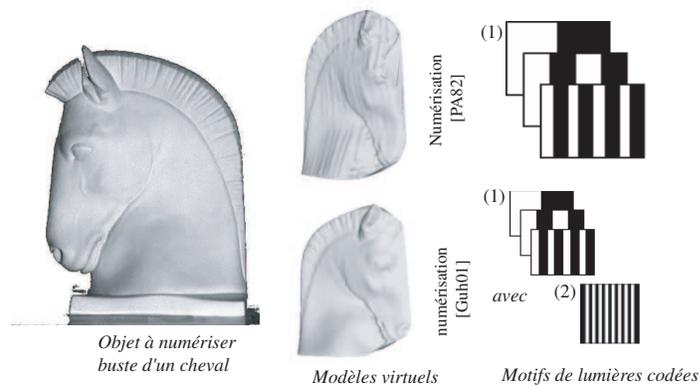


FIG. 2.6 – Deux exemples de codage temporel et les résultats respectifs de reconstruction.

Une deuxième forme de codage est employée surtout pour les scènes ou les objets dynamiques, c'est le codage spatial (*spatial coding*). Le but étant de réduire le nombre de motifs envoyé sur l'objet par l'utilisation d'un seul modèle comportant lui même plusieurs motifs élémentaires. Le code d'un pixel est donc obtenu à partir de son voisinage dans l'image. Nous citons dans ce type de codage trois codifications employés : (a) *Non-formal codification*, (b) *De Bruijn sequences* et (c) *M-arrays*. Ces approches sont basées sur la division du motif en régions, l'information contenue dans chaque région est capable de générer un mot code sans l'utilisation d'un codage mathématique. Par exemple, en 1993 Maruyama et al. [MA93] ont conçu un nouveau motif binaire avec des traits noirs composées de segments aléatoires. Chaque trait est identifié par la projection de points, elles aussi, aléatoires sur la scène. Cette méthode est destinée aux objets de surfaces lisses. Les séquences de De Bruijn engendrent elles aussi un algorithme structuré par le biais de son fondement théorique basé sur les graphes de De Bruijn. Ces graphes sont utilisés pour décoder les séquences du motif projeté, à l'image de l'approche décrite dans [VO90].

D'autres méthodes plus variées ont été également proposées, elles substituent les codes

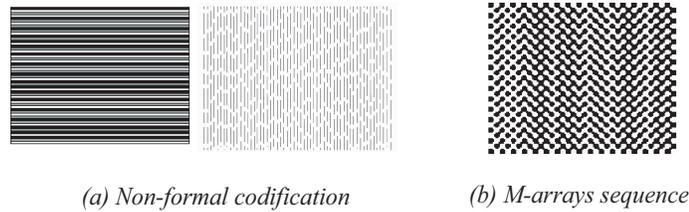


FIG. 2.7 – Exemples de motifs pour une stratégie de codage spatial.

binaires avec d'autres codes plus riches en niveau de gris et en couleurs. L'idée ici est d'exploiter la multitude de codes offerts par ces plattes de niveau de gris [CH85] ou de couleurs [TI90][SI85][ZCS02] projetées sur l'objet à reconstruire. Pour l'appariement, il suffit de retrouver le code du pixel dans l'image capturée par sa couleur ou son niveau de gris puis de calculer la disparité par rapport au modèle initial. la figure 2.9 présente différents motifs lumineux couleurs utilisés dans littérature.

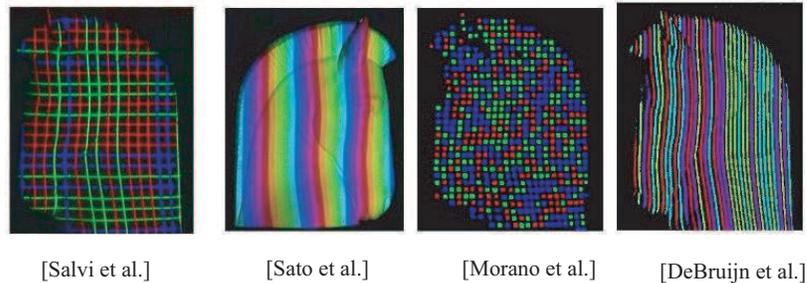


FIG. 2.8 – Exemple de motifs de lumières en couleurs envoyés sur l'objet à numériser.

Dans leur approche d'acquisition en vue de l'authentification de personnes via leurs faces 3D, C. Beumier et M. Achery [BA01a][C.03] ont utilisé le principe de projection de motif structurant pour construire leur base de test appelée *3D.RMA*. Cette base a été développée dans le cadre du projet *M2VTS (Multimodal Verification for Teleservices and Security Applications)*⁸ et deux prototypes ont été réalisés l'un d'entre eux passe par une lumière en niveau de gris alors que le second considère une succession de bandes de couleurs. Les primitives à mettre en correspondance sont les centres des franges de lumières projetées. La base comporte les visages de 120 personnes qui sont présentés par des points 3D et la vérification d'identité se fait par comparaison des surfaces formées en comparant des coupes longitudinales de ces surfaces [BA01a]. Pour un bon moment, cette base est restée la seule base publique même si la qualité n'était pas supérieure.

⁸<http://www.sic.rma.ac.be/~beumier/M2VTS/m2vts.html>

Une seconde base multi-modale appelé *XM2VTSDB (Multi Modal Verification for Teleservices and Security applications)*⁹, parmi ses données des visages 3D, propriétaire au CVSSP¹⁰ a été acquise par projection d'un motif lumineux structurant [GSBC⁺03]. Cette base comporte les visages de 295 sujets utilisées dans des campagnes d'évaluation de différents algorithmes de reconnaissance/vérification de personnes par le visage. Les images 3D écrites en VRML (*Virtual Reality Modeling Language*) issu de cette base sont lisses et présentent les structures grossières des visages sans présenter les détails au niveau des régions de détails comme les lèvres, les yeux, etc.

2.2.2 Techniques de reconstruction passives

La vision passive, contrairement à la vision active, n'utilise aucune structuration particulière de la scène. Les seules données disponibles sont, une ou plusieurs images prises à partir de plusieurs angles de vue (par une ou plusieurs caméras). Il existe de nombreuses familles d'approches pour modéliser des formes à partir d'images (*shape from X*). Certaines méthodes n'utilisent qu'une seule image, à condition de disposer d'information et de connaissances a priori supplémentaires. L'information apportée par une seule image 2D ne peut sinon aboutir à la construction d'un modèle 3D, si rien n'est fait pour lever l'ambiguïté fondamentale de projection due à la perte de l'information de profondeur. Parmi celles utilisées dans les méthodes existantes, il y a les propriétés de réflectance d'une surface (*shape from shading*), la texture des surfaces (*shape from texture*). Dès que plusieurs images d'une même scène sont disponibles, divers mécanismes de détection de cohérence et de triangulation permettent d'apporter une réponse géométrique à la question du contenu de la scène. Il est par exemple possible, à partir des disparités entre niveaux de gris de deux images prises de points de vue très proches, de retrouver l'information de profondeur pour chaque pixel de ces images (*shape from stereo*). Nous nous limitons dans notre état de l'art aux deux grandes catégories faisant partie des approche multi-images : les méthodes basées silhouettes et les méthodes basées sur la stéréovision.

Dans le cas des méthodes basées sur l'extraction de silhouettes dont nous citons les travaux [Zhe94][MFK99][XJF01], l'acquisition est faite moyennant plusieurs caméras placées devant l'objet, dans ce cas l'objet est fixe. Une deuxième alternative consiste à faire pivoter l'objet (une table rotative est utilisée) alors qu'une seule caméra est utilisée pour l'acquisition des images [Her04]. Ces images sont ensuite utilisées pour extraire les silhouettes de l'objet (les contours extérieures). En cherchant l'intersection des cônes formées par ces contours et le centre de chaque image, un modèle grossier de l'objet est ainsi produit (voir figure 2.9). Les détails du modèle produit augmentent donc avec le nombre d'images utilisées. Plusieurs améliorations de cette technique d'acquisition 3D ont été proposées, notamment dans [HS04], en combinant l'approche décrite avec une approche multi-stéréo pour le raffinement. En effet, les auteurs proposent de construire, dans un premier temps, un modèle grossier de l'objet, puis de le raffiner en proposant une nouvelle formulation de la contrainte

⁹<http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/>

¹⁰Centre for Vision, Speech and Signal Processing

des silhouettes et une approche de diffusion multi-résolution du vecteur gradient pour la composante stéréo. De même dans [MFK99] où les auteurs utilisent une approche hybride qui part d'un modèle initial formé par une extraction de silhouettes puis raffiné par vision active. Dans [LMT00], un modèle générique est reconstruit à partir de modèles 3D qui ont été acquise par scanner 3D. Le modèle initial est une combinaison linéaire de vecteurs propres obtenus par application de l'analyse en composante principale sur les visages de la base. Le cœur de cette approche consiste à coller au mieux les silhouettes extraites à partir de plusieurs images au modèle générique synthétisé. Les améliorations ainsi proposées permettent de retrouver les détails du visage non préservé par les méthodes basées silhouettes, néanmoins elles sont coûteuses en terme de temps de calcul et s'orientent plus vers des applications d'acquisition d'objets statiques et rigide comme les objets muséologiques (sites archéologiques, statues, etc.).



FIG. 2.9 – Construction de l'enveloppe visuelle par intersection des cônes définis par les centres optiques des caméras et les silhouettes de l'objet [Her04].

Dans le cas des approches basées sur la stéréovision passive, c'est l'estimation de la disparité (déplacement dans les images entre les projections d'un même point 3D) qui permet d'en déduire l'information profondeur [LTM96][CFFB06][VTC05]. Cette géométrie multi-images permet, par l'intermédiaire de l'intersection de plusieurs rayons optiques, de retrouver la troisième dimension, ambiguë sur un seul rayon. Le défi relevé par ces techniques réside principalement dans la résolution du problème de la mise en correspondance dense ou la mise en correspondance de primitives dans les images. Pour un bonne formulation de la problématique et un état de l'art complet des ces approches, le lecteur pourra se référer au [Kos02]. Dans la littérature, un spectre très large de travaux relatif à ce problème de mise en correspondance dense ou de primitives a été développé. Le problème majeur des ces approche réside dans la définition d'un schéma de mise en correspondance stéréo robuste. En effet, ce problème est plus complexe dans le cas du visage du fait que la majorité de ses regions (celles de peau) sont similaires. Ceci entraîne des ambiguïtés et des complexité supplémentaires pour les algorithmes d'appariement. La précision des modèles reconstruits dépendent essentiellement de ce facteur. Une solution ciblant les visages humain a été proposée dans [Nic03, Nic02], elle considère un capteur multi-images composé de cinq caméras montées sur la même hauteur en face de la personne, chaque ensemble de trois caméras reconstruisent un modèle partiel du visage. Ici, une nouvelle approche de calcul de correspondant est mise en œuvre, néanmoins, deux lumières aléatoires servant à texturer les

différentes parties du visage sont introduites. Les résultats présentés montrent des visages lisses par filtrage des points 3D obtenus d'où une perte au niveau des détails. Dans [PJ06], les auteurs proposent un schéma hybride de reconstruction de visages en vue de la reconnaissance. Ce schéma consiste à calculer l'information profondeur de quelques primitives du visage et l'adaptation d'un modèle générique en vue de retrouver la structure 3D du visage requête.

Un autre ensemble d'approches de reconstruction 3D à partir d'un flux vidéo continue qui capture l'objet a été développé. Alors que en stéréovision, la solution se trouve dans la combinaison de deux ou trois images prises de points différents, en perception du mouvement tridimensionnel, la solution se trouve dans la combinaison de plusieurs images prises, cette fois-ci, à des instants différents. Dans le cas de formation d'une image 3D du visage à partir de ce type d'approche, l'introduction d'un modèle générique reste l'un des moyens les plus utilisés pour désigner une solution initiale proche. En effet, en passant par un modèle générique [CC03] comme modèle initial, certains algorithmes raffinent cette solution, par une étape d'optimisation. Il s'agit de déformer le modèle de référence pour le rapprocher aux mesures réelles extraites par flot optique. C'est ce qui a été décrit aussi dans [IY96] et [AAM03], où les méthodes consistent à donner une approximation du visage, à partir de deux vues statiques orthogonales : une vue frontale et une deuxième de profil. L'adaptation du modèle se fait moyennant des points d'intérêts du visages (position des yeux, de la bouche, etc.). Dans les deux méthodes, la modélisation est rapide, néanmoins, les résultats présentés restent étroitement liés au modèle de référence pris comme solution initiale.

Ce paragraphe a fait l'objet d'un survol des approches de reconstruction tridimensionnelle. Leurs principes sont parfois appliqués dans le cas des visages humains. Plusieurs catégories ont été distinguées vu le large spectre de méthodes qui ont été proposées dans ce domaine. En plus de ces approches, d'autres moins récentes basées sur la texture où sur les variations globales de luminosité sont censées décrire les variations des normales à une surface observée. Dans ce cas, on extrait alors la forme générale des objets à partir des variations photométriques observées. Dans la section suivante, nous nous penchons sur un deuxième survol des approche de reconnaissance faciales en particulier celles qui exploitent une numérisation tridimensionnelle du visage.

2.3 Identification/vérification de personnes par visages 3D

La biométrie est la discipline de l'analyse des caractéristiques biologiques, morphologiques et comportementales d'une personne. Elle est destinée à déterminer son identité de manière irréfutable. Par ailleurs, elle attire, de plus en plus de chercheurs de différentes disciplines tels que : le traitement d'images, la reconnaissance des formes, les réseaux de neurones, la vision par ordinateur et la psychologie. La reconnaissance de visages est l'un des axes de recherche, dans cette discipline, qui a connu un gain d'intérêt considérable ces 10 dernières années, vu son caractère d'acquisition peu intrusif. Ce problème peu être formulé de la manière suivante : disposant d'une image fixe ou de séquence vidéo du visage (ou toutes autres formes de représentation), le problème consiste à identifier ou a authentifier la

ou les personnes de la scène en utilisant une base de visages enregistrés. En identification, il s'agit de reconnaître un visage inconnu parmi les visages, d'identités connus, de la base. C'est par conséquent un problème de matching 1 à plusieurs (*one-to-many*). En revanche, l'authentification consiste plutôt à confirmer l'acceptation ou le rejet d'une identité présenté au système. C'est, par conséquent, un problème de matching 1 à 1 (*one-to-one*).

Au début des années 1970, la reconnaissance par le visage était principalement basée sur des attributs faciaux mesurables comme l'écartement entre les yeux, les sourcils, les lèvres, la position du menton, la forme, etc [Kel73]. Depuis les années 1990, les différentes techniques utilisées exploitent les évolutions effectuées dans le domaine du traitement d'images (filtrage, transformations, normalisation, etc.), et plus récemment encore les techniques d'analyses et de classification de données. La majorité de ces techniques agissent sur des images de couleurs ou d'intensité du visage. Cependant, vu qu'elles sont basées sur des données de nature pixellique, ces méthodes ont montrées leur insuffisance et le manque de robustesse pour des applications de haute sécurité. En effet, des programmes d'évaluation des approches et des produits commerciaux, depuis 1994, ont été mis en place pour voir le niveau de maturité de la technologie et identifier les pistes de recherches futures : FERET [PMRR00], FRVT [PGM⁺03], XM2VTS [GSBC⁺03], FRGC [PFS⁺03]. En vue de s'affranchir des limitations imposées par l'écrasement sur un plan image, par l'effet de projection, un ensemble plus récent d'approches utilisant, de différentes manières une modélisation tridimensionnelle (partielle ou complète) du visage, a vu le jour. Ces techniques, via leur introduction de la troisième dimension a permis de profiter de l'information profondeur dans certains travaux et de sa flexibilité pour adapter, d'une manière plus réaliste le modèle d'illumination en vue de retrouver les conditions d'éclairage réelles de prise de vue, dans d'autre travaux.

Dans la suite de ce paragraphe, nous donnons un survol sur les approches de reconnaissance faciales, mais nous nous intéressons plus particulièrement aux méthodes utilisant des images 3D. D'abord, nous donnons une taxonomie des approches selon les représentation des visages et les algorithmes utilisés (voir figure 2.10). Puis, nous dévoilons les approches de reconnaissance faciales de chacune des catégories de la taxonomie dégagée. Ici, une attention plus approfondi est attribuée aux techniques 3D et 2D+3D vu que cette thèse s'intéresse plus particulièrement à ce type d'approches. Finalement, nous donnons un bilan de l'existant et nous introduisons nos approches de reconnaissance faciale 3D.

2.3.1 2D vs. 2D

Ces techniques concernent la reconnaissance/authentification à partir d'images frontales. Elles sont appliquées sur des images couleurs, d'intensité et plus récemment des images prises avec des capteurs thermiques (*IR images*). Au début, elles se basaient sur des caractéristiques géométrique des visages (*geometric features*), puis elles ont reçu une évolution remarquable par l'introduction des approches d'analyse de données appliquées sur les vecteurs formés de pixels des images pour la recherche de bases de représentation plus restreintes et plus pertinentes. Outre ces dernières approches, d'autres basées sur des techniques plus avancée d'analyse d'image comme les filtres de Gabor et de matching de

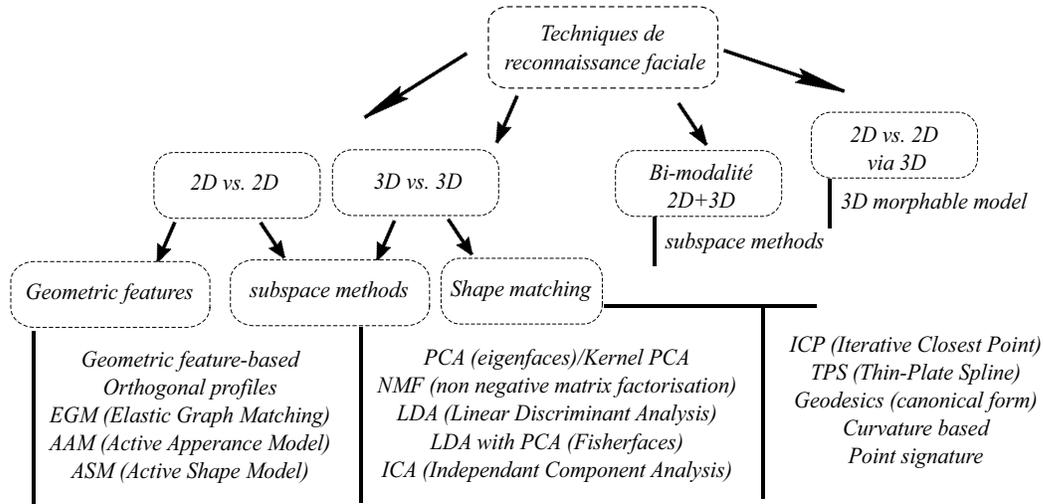


FIG. 2.10 – Taxonomie des approches de reconnaissance de visages 2D, 2D/3D, 3D et 2D+3D.

graphe ayant comme sommets les points caractéristiques du visage ont été développé. Nous classons toutes ces approches en deux sous catégories à savoir : (a) les méthodes de sous espace (*subspace methods*) et (b) les approches à base de caractéristiques géométrique (*geometric feature-based methods*). Dans la suite, nous détaillons ces deux sous catégories et les approches qui en découlent. Il est à noter que quelques unes de ces approches ont été appliquée sur des images de profondeurs, profitant ainsi du développement mathématiques considérable que les approches 2D ont gagnées, ces quelques dernières années.

2.3.1.1 Méthodes de sous-espace (*subspace methods*)

Ces techniques opèrent sur des images 2D et plus précisément sur des pixels. Une image de taille $N \times M$ est considérée comme un vecteur de haute dimension de taille $N \times M$. Ce dernier constitue le format avec lequel les techniques d'analyse statistique analysent les images. Le but ici est de déceler une représentation efficace (espace caractéristique) facilitant les calculs de similarité et permettant l'emploi de techniques d'apprentissage pour l'analyse et la synthèse des images. Parmi ces techniques, l'approche des visages propres (*eigenfaces*) [TP91] et celle des visages de Fisher (*Fisherfaces*) [YY01] sont désormais les plus connues.

La méthode de visages propres (*eigenfaces*) consiste à une analyse en composante principale (ACP) appliquée sur les vecteurs de pixels des différentes images 2D [TP91]. Une partie de la base de visages est utilisée pour produire un nouveau sous espace de projection, appelé base d'apprentissage. En reconnaissance, une image du visage à identifier ou à authentifier est présentée. Ce nouveau vecteur est projetée sur le nouvel espace de représentation en vue d'extraire ses nouvelles coordonnées. Elles sont comparées aux coordonnées des visages de

la bases de reconnaissance via une distance euclidienne : C'est le score de dissimilarité entre les visages. Cette méthode est rapide et elle permet de projeter les visages dans un sous espace de taille plus réduit. Cependant, elle ne permet pas de distinguer entre les classes et donc elle est optimale pour représenter et non pour discriminer. Plusieurs variantes de cette méthodes ont été proposée dans littérature, en utilisant différentes bases de visages avec différentes configurations d'apprentissages et de test.

La méthode de visages de Fisher (*Fisherfaces*) est fondée sur le principe de l'analyse linéaire discriminante (LDA). Cette dernière définit différentes classes avec différentes statistiques. Ainsi, elle permet de regrouper les images dans leurs classes correspondantes permettant la maximisation de la dispersion inter-classes et minimisant celle intra-classes. La méthode des visages de Fisher offre une reconnaissance faciale plus efficace que celle des visages propres grâce à son principe de classification [BHK97]. Dans [YY01, PV03], cette technique est appliquée sur le résultat produit via la méthode visages propres. La première étape est destinée pour réduire la dimension et par conséquent produire une nouvelle base de projection, la seconde étape vise la discrimination des primitives produits par la première pour une meilleure discrimination inter-classe.

D'autres méthodes ont été également proposées dans cette sous catégorie comme la méthode ICA (*Independent Component Analysis*) pour une meilleure représentation, les SVM (*Support Vector Machine*) comme classifieur de primitives plus robuste et des approches hybrides combinant différentes méthodes de représentation et de classification. Pour un survol plus complet des approche 2D vs. 2D voir [ZCPR03].

2.3.1.2 Méthodes basées sur des attribues géométriques (*geometric feature-based methods*)

A la différence des approches de sous espace (*subspace methods*), cette catégorie de méthodes ne nécessite pas de phase d'apprentissage, uniquement une photo par personne dans la base de visages est exigé. Une des approches les plus connues est celle basée sur l'appariement de graphe élastiques (*Elastic Graph Matching*). A chaque visage est associé un graphe libellé en se basant sur les transformation en ondelettes de Gabor. Le graphe d'un nouveau visage est extrait moyennant un appariement de graphe élastique et peut être comparé par une simple fonction de similarité. Dans [ZYL97], les auteurs ont montré la robustesse de cette approche par rapport aux approches par visages propres et par réseaux neuronaux. Plusieurs variantes de cette approche ont été développé, depuis. Nous citons MEGM pour (*Morphological elastic graph matching*), où les caractéristiques obtenues par la transformation par ondelettes de Gabor sont remplacées par des caractéristiques obtenues via l'application d'opérateurs morphologiques sur l'image du visage dans un contexte multi-résolution [JD96]. Nous trouvons, également, l'introduction des techniques de discrimination linéaire appliquées sur les vecteurs de caractéristiques en vue de sélectionner les caractéristiques les plus discriminantes [TKP02, JD96, KTP00].

Une deuxième technique appelée Modèle Actif d'Apparence AAM *Active Appearance Model* est aussi bien connu dans la littérature. 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 [CET98]. 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. Pour construire le modèle d'apparence, les images de la base seront modifiées de façon à faire coïncider les points de la face 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 reconnaissance d'un visage se fait par une méthode d'optimisation comme la méthode de descente de gradient effectué sur l'ensemble des paramètres de déformation.

2.3.2 2D vs. 2D via 3D

Cette sous catégorie d'approches considère un modèle tridimensionnel du visage comme étape intermédiaire pour la comparaison de deux images faciales 2D. En effet, un modèle 3D est plus flexible et plus réaliste qu'une image 2D pour caractériser les conditions d'éclairage ou de les modifier. Dans ce cas, c'est une étape de normalisation de l'illumination entre les images qui est visée. En outre, l'introduction d'un modèle 3D peut permettre de corriger la pose de la personne par mappage de texture sur le modèle puis l'obtention d'une vue frontale. L'approche la plus connue dans cette catégorie est le modèle déformable (*morphable model*). Introduite par l'équipe de l'université de Basel [BV99, BRV02, BV03], l'approche proposée adopte une technique d'analyse par synthèse. Elle 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 référence moyennant un algorithme à base de flot optique. Le modèle générique est donc produit pour servir comme modèle déformable en fonction de paramètres de formes, d'apparence, de pose et d'illumination. Ces paramètres produits à la suite de l'étape d'adaptation du modèle 3D par rapport à l'image 2D du visage à reconnaître sont conservés pour la reconnaissance. La même approche a été utilisée en vue de produire un modèle frontal à partir d'une image sous n'importe quel angle de vue. Cette approche a été testée pendant le programme d'évaluation FRVT 2002 [PGM⁺03], elle a affiché des améliorations par rapport aux approches existantes.

Dans d'autres approches, le modèle déformable 3D, est utilisé pour synthétiser un nombre important d'images pour chaque sujet en vue de produire un corpus d'apprentissage plus important englobant différentes variations de pose [?]. D'abord, un modèle 3D du sujet est obtenu par adaptation d'une image 2D du sujet sur le modèle déformable. Le synthèse d'images 2D est effectué après avoir obtenu le modèle tridimensionnel du visage du sujet sous différentes orientations et différentes conditions d'éclairage artificielles, mais, réalistes.

2.3.3 3D vs. 3D

Avec l'essor du 3D et des approches d'acquisition de formes 3D, l'information forme a attiré une attention substantielle du fait qu'elle n'est pas influencée par une variation d'illumination ni de pose et qu'elle permet de déceler des descripteurs de formes importants, essentiellement de courbure et de segmentation en régions homogènes. Ainsi, un tas de nouvelles approches agissant soit sur des images de profondeur (*range images*) soit directement sur la surface faciale a vu le jour.

2.3.3.1 Méthodes de sous espace (*subspace methods*)

Une des premières approches a été proposée par Heshner et al. dans [HSE03]. Elle fait recourt à une méthode de réduction de dimensionnalité (ACP) sur des images de profondeur du visage au lieu des images intensité. Leur base renferme 222 images de profondeur associées à 37 personnes. Six expressions faciales différentes ont été collectés pour une même personne, dans la base de test. Les images de profondeur sont normalisées en pose moyennant la détection du sommet nasal et son alignement avec l'axe des Y . Un espace propre de dimension réduite est ensuite créé à partir des images de profondeur normalisées et sur lequel sont projetées les images en question. Un taux de reconnaissance de 83% est obtenu. L'analyse en composantes principales a été utilisée aussi, moins récemment, par Achermann et al. [AJB97] couplée avec les Modèles de Markov cachés. Leur approche consiste à construire un espace propre à partir de 5 poses chacune pour 24 personnes. Leur méthode a été testée sur 5 poses différentes d'un même sujet. Les auteurs ont reportés un taux de reconnaissance de 100% sur leur base moyennant ACP avec 5 images d'apprentissage par personne. Pan et al. [PW05] expérimentent deux approches de reconnaissance de visages 3D utilisant les deux techniques de visages propres et la distance de Hausdorff. La base de visage utilisée est celle de XM2VTS [GSBC⁺03] et les résultats obtenus sont respectivement 5-7% et 3-5% comme *EER* (*Equal Error Rate*). La méthodes des visages propres a été utilisée aussi comme algorithme de base (*baseline*) des évaluations faites au sein du programme d'évaluation *FRGC 2005*. Cette méthodes affichent des résultats moins intéressant que son application aux images d'intensité des mêmes personnes [PFS⁺03]. Cependant, la fusion des deux modalités 2D et 3D, utilisant le même algorithme *eignefaces*, améliore le résultat [PFS⁺03].

2.3.3.2 Méthodes basées sur l'alignement de surfaces

Les méthodes appartenant à la première catégorie étendent les techniques de sous espace, notamment ACP, sur des images de profondeurs. L'évaluation au sein de FRGC 2005 [PFS⁺03], de cette approche, a montré que son application en 2D est plus intéressant que son application en 3D. D'autres méthodes basées sur l'appariement des surfaces faciales, initialement utilisées dans des problèmes d'alignement de modèles partiel ou de reconnaissance d'objet 3D, ont été adaptée pour la cas du visage.

La méthode la plus connue est celle du plus proche voisin itéré (*Iterative Closest Point*)

qui a été proposée par Besl et McKay [BM92] pour le recalage de nuages de points 3D. Étant donné deux surfaces 3D à recaler, cette méthode traite une de ces deux surfaces en tant que modèle référence et l'autre en tant que surface test. Elle vise le recalage rigide de la surface test, itérativement, jusqu'à ce qu'elle soit le plus proche possible de la surface modèle. Cette méthode emploie l'heuristique du plus proche voisin au sens euclidien à fin d'établir une correspondance entre les points des deux surfaces modèle et test. En d'autres termes, pour chaque point de la surface test, elle calcule le point de la surface modèle qu'elle considérera comme son point correspondant. Ces points correspondants serviront par la suite à la détermination d'une transformation rigide assurant le mouvement de la surface test pour minimiser au plus la distance séparant les points correspondants. Une erreur au sens des moindres carrées est calculée. Elle servira comme critère d'arrêt de l'itération. En effet, l'algorithme s'arrête dès que la valeur de cette erreur chute au dessous d'une valeur seuil.

Medioni et al. [MW03] ont exploité l'algorithme *ICP* pour le recalage et le calcul de similarité entre les surfaces faciales 3D. L'acquisition d'une base de 700 modèles 3D (100 sujets) est effectuée par le biais d'un capteur stéréo passif. Ils reportent un taux de reconnaissance de 98%. Plus récemment, Lu et al. ont utilisé le même principe pour recaler des surfaces faciales [LJ05]. Ils ont employé un algorithme basé caractéristique suivi d'une version hybride d'*ICP* qui alterne dans ses itérations successives entre la version de Besl et McKay [BM92] et celle de Chen et Medioni [CM92]. Dans ce sens, ils ont profité des avantages des deux méthodes. Leur algorithme hybride a été testé sur une base de 18 sujets différents avec des images test présentant des variations de pose et d'expressions. Une image test est recalée avec les 18 images référence et le meilleur alignement est celui ayant la plus faible erreur résiduelle. Avec cette erreur résiduelle uniquement, un taux de reconnaissance de 79.3% a été reporté. Dans leurs travaux, ils ont pu améliorer ce taux jusqu'à 84% en profitant en plus de l'information texture et de courbure. Plus loin, une deuxième amélioration a été introduite par l'application de l'algorithme TPS (*Thin-plate Spline*) pour le recalage non-rigide des surfaces faciales après un premier recalage rigide via *ICP*. Une base de 100 sujets a été utilisée pour l'évaluation et un taux de reconnaissance de 89% est obtenu utilisant les données 3D et 91% utilisant les données 2D et 3D fusionnées. Une amélioration, est aussi présentée dans [CBF05] par Chang et al., se manifeste dans l'application d'*ICP* sur des régions du visage. Ces régions de différentes tailles sont sélectionnées autour du nez, les expérimentations sont effectuées sur la base FRGC contenant 4000 modèles de 400 personnes avec différentes expressions faciales. Ils affichent un résultat de 92% comme taux de reconnaissance.

2.3.3.3 Méthodes basée sur le calcul de courbure

Les propriétés de surface comme les courbures gaussiennes, principales, minimales et maximales permettent la segmentation de la surface en régions de concavité, de convexité. Ainsi, elles offrent une bonne discrimination de l'information pour des finalités de reconnaissance d'objets. Tanaka et al. [TIC98] calculent des cartes de courbures principales minimales et maximales à partir des cartes de profondeur des visages. A partir de ces cartes

de courbure, ils extraient les lignes de sommets et celles de vallées sur la surface faciale. Les premiers constituent un ensemble de vecteurs correspondant aux maxima locaux sur les valeurs de courbure principale minimale. Les seconds sont les vecteurs correspondant aux minima locaux sur les valeurs de courbure principale maximale. A partir des lignes de sommets et celles de vallées, ils construisent des images gaussiennes étendues (*EGI*) pour le visage en mappant chacun des vecteurs de courbure principale sur deux sphères unités, une pour les lignes de sommets et une pour celles de vallées. La mise en correspondance entre l'image de profondeur référence et celle test s'effectue moyennant la corrélation sphérique de Fisher entre les *EGI* de sommets et de vallées respectifs. Notons que cette mesure de similarité est invariante à la rotation. Cet algorithme a été testé sur 37 images de profondeur de sujets différents. Un taux de reconnaissance de 100% a été obtenu. L'extraction des lignes de sommets et de vallées nécessite le seuillage des cartes de courbures. Vu l'absence de règle explicite d'obtention de seuillage idéal, la localisation des lignes en question est très sensible à la valeur de seuillage choisie ce qui constitue un inconvénient incontestable.

Lee et Miliotis [LM90] ont obtenu des régions convexes à partir de la surface faciale moyennant la courbure à fin de représenter les différentes régions faciales. Chaque région convexe est représentée par une *EGI* moyennant un mappage un à un entre les points de ces régions et ceux sur une sphère unité qui ont une même normale à la surface. La similarité entre deux régions convexes est évaluée par la corrélation de leurs images gaussiennes étendues. Pour établir la correspondance entre deux visages, il suffit de procéder par une mise en correspondance de graphe en tenant en compte uniquement les régions convexes. On a supposé que les régions convexes du visage sont moins sensibles aux changements des expressions faciales que celles non-convexes. Ainsi, leur approche a un certain degré d'invariance aux expressions faciales. Mais, leur algorithme a été testé uniquement sur 6 personnes et aucun résultat n'a été explicitement affiché. Plus récemment, cette approche a été améliorée par l'introduction des SVM pour la classification appliqués sur des vecteurs de huit valeurs de courbures. Le taux de reconnaissance obtenu est de 96% pour une base de 100 sujets.

Un inconvénient majeur de ces méthodes est que la localisation des points caractéristiques ainsi que les centroides des régions faciales est très susceptible au bruit surtout puisque la courbure est une dérivée seconde. Ceci mène à des erreurs de localisation des caractéristiques qui peuvent être aggravées par des occultations résultantes d'un changement de pose.

2.3.3.4 Méthodes basées sur le calcul géodésique (*bending-invariant canonical form*)

Une idée assez innovante a été proposée par Kimmel [BBK05] et ses deux étudiants jumeaux Bronstein. Elle met en avant la nature isométrique de la surface du visage (préservant la longueur entre les points dans le sens d'une distance géodésique mais pas euclidienne). L'idée consiste à modéliser le visage comme un nuage de points menu d'une distance géodésique et à projeter ce nuage de points vers un espace de dimension réduite menu d'une distance euclidienne ce qui conduit à représenter la surface faciale par une surface isométrique appelée forme canonique (*canonical forms*). Cette projection a été effectuée

moyennant l'algorithme MDS (*Multidimensional Scaling*) qui est une technique de réduction de dimensionnalité préservant la distance. Notons que la nouvelle représentation est, par hypothèse, invariante aux expressions faciales puisque la distance géodésique est invariante à toute sorte de transformation non-rigide subit par cette surface. Ces formes canoniques sont par la suite alignées et interpellées sur une grille cartésienne donnant naissance à des images canoniques. Finalement, un espace propre est créée à partir de ces images canoniques d'apprentissage. Les images test sont assujetties aux mêmes traitements avant de subir la mise en correspondance. D'après leurs résultats, leur algorithme de reconnaissance sait distinguer entre les frères jumeaux qui l'ont développé et un taux de reconnaissance égal à 100% est obtenu.

2.3.4 Bi-modalité 2D+3D

Alors que la recherche en reconnaissance faciale 3D a débuté avant 1990, la combinaison 2D+3D a vu le jour au début de l'année 2000. Depuis, plusieurs approches visent la fusion des deux modalités au niveau des scores de similarité obtenus indépendamment. Après l'approche purement 3D décrite dans [BA01b] qui consiste à comparer les profils de modèles 3D après leurs recalage, Beumier et al. proposent dans [BA00] un nouveau schéma de reconnaissance bi-modale par l'introduction de comparaison de profils 2D, combinée à la première approche. La fusion est effectuée moyennant une somme pondérée des scores élémentaires. L'approche 2D+3D affiche un score plus élevé (EER = 1,4%) par rapport à l'approche utilisant seulement les données 3D. Dans [BBGK04], les frères Bronstein proposent un schéma de décomposition utilisant ACP bimodal 2D+3D. En 3D, c'est la transformation isométrique (*canonical forms*) qui a été utilisée, alors qu'en 2D c'est une approche de visages propres appliquée sur les image de texture associé au visage 3D (*flattened textures*). Une autre expérimentation réalisée par Chang et al. dans [CBF03], consiste à appliquer une approche de visages propres sur des images de profondeurs et des images d'intensité de 200 sujets. Les scores de reconnaissance obtenus sont respectivement, 89% pour l'approche 2D seule, 94% pour pour l'approche 3D seule et 99% pour l'approche combinée 2D+3D. De même, Tsalkanidou et al. dans [TTS03] expérimentent l'approche de visages propres sur des images de profondeur (3D) et sur des images couleurs (2D) de 40 sujets de la base XM2VTS. Un taux de reconnaissance est obtenu par une approche multi-modale 2D+3D autour de 99%, plus élevé que les taux obtenu par l'approche appliqué sur des données 2D ou 3D seules. Papatheodorou et al. appliquent dans [PR04] une approche multi-modale 2D+3D via une généralisation de l'algorithme de matching *ICP* en 4D (x, y, z, intensité). A la différence des autres approches, ils combinent les deux modalités au niveau des données et non au niveau des scores indépendants. Leur tests sont été effectué sur une base de 62 sujets dans la galerie. Les images de test incluent des variations de pose et d'expressions faciales. Le taux de reconnaissance obtenu est autour de 99% pour des expressions neutres et une vue frontale.

Des companies qui commercialisent des systèmes de reconnaissance faciale ont aussi expérimenté la combinaison des deux modalités 2D et 3D. Dans [HBGdM05], une description de l'approche *Viisage* est donnée. Le matching 3D est effectué moyennant un appariement

de graphe hiérarchique, déjà utilisé dans leurs système 2D commercialisé. Cette approche est plus rapide que les approches itérative de style *ICP*. La fusion est faite au niveau des scores de similarité. Leur approche affiche un score de vérification 93% correspondant à un taux de fausse acceptation égal à 0,1%. D'un autre coté, l'approche *Geometrix*, décrite dans [MGM⁺05] fait appel à l'algorithme de matching 3D développé dans [MW03] et à l'algorithme de matching 2D¹¹. Un taux de vérification autour de 87% est affiché par la combinaison des deux approches correspondant à un taux de fausse acceptation égal à 0,1%. Les évaluations des deux approches sont effectuées sur le corpus FRGC V2.0 qui contient 4007 scans 3D.

la plupart des approches multi-modales 2D+3D confirment l'intérêt de fusion des deux modalité 2D et 3D en vue d'obtenir des taux de reconnaissances et d'authentification plus élevés que les approches exploitant les données 2D ou 3D seules. Dans [PFS⁺03], cette hypothèse a été aussi confirmé, cependant les taux de reconnaissances 3D sont moins intéressants que les taux obtenus via les données 2D. Ceci est du à l'algorithme de base utilisé (*eigenfaces*) qui exploite les données 3D comme un vecteur de haute dimension, et donc en faisant abstraction sur le potentiel que peut présenter ces données par apports aux données 2D.

Dans la présente section nous avons donné un état de l'art sur les approches de reconnaissance faciales selon une taxonomie que nous avons élaborée en fonction du support de reconnaissance utilisée : 2D, 3D ou 2D via 3D. Dans la conclusion de ce chapitre nous mettons l'accent sur les orientations de nos travaux de recherche en fonction d'un bilan de l'existant en matière d'acquisition et de reconnaissance faciales 3D.

2.4 Conclusion

Nous nous intéressons, dans cette thèse, aux deux problématiques d'acquisition et de reconnaissance tridimensionnelles de visages. L'objectif est de palier aux différentes limitations en vue de proposer des algorithmes plus précises et plus efficace. Nous avons présenté, ainsi dans ce chapitre, un survol des approches proposées dans littérature. Très peu d'entre elles proposent un schéma d'acquisition et de reconnaissance complet, la plupart propose un algorithme de reconnaissance utilisant des techniques d'acquisition existantes (scanners lasers, lumière structurant, etc.). Nous avons donné, également, une taxonomie des approches sur lesquelles sont organisées les deux parties de notre état de l'art.

Dans la première partie, deux grandes catégories d'approches d'acquisition tridimensionnels ont été distinguées : *les approches actives* et *les approches passives*. Peu d'approches de reconnaissance faciales 3D qui s'est intéressés aux approches passives malgré leurs contexte peu contraignant (pas de lumière projetée sur le visage). Ceci est dû au manques de précision des mesures et la production de modèles non-denses ou grossiers. C'est pour ces raisons, que la vision active (introduction de motif lumineux envoyé sur le visage) a pris un intérêt

¹¹www.nevenvision.com

plus important. Moyennant la triangulation active, ces approches sont plus précises et produisent des modèles plus denses couvrant la géométrie 3D du visage, malgré l'introduction de lumière projetée sur le visage.

Dans les chapitres qui suivent, nous présentons des contributions à l'acquisition et à la reconnaissance tridimensionnelle de visages. Dans un premier temps, nous nous intéressons à une approche de stéréovision active en vue de reproduire la totalité ou une partie du visage. Puis, nous abordons le problème de reconnaissance faciale 3D par la mise en place d'un schéma de matching de surfaces faciales 3D. Des contributions à ces deux domaines seront présentées, d'une part, en vue de reproduire le plus fidèlement possible la structure 3D du visage et d'autres part pour faire face aux problèmes d'éclairage, de changement de pose et d'expressions faciales, et bien d'autres problèmes classiques de la reconnaissance faciale.

Chapitre 3

Reconstruction faciale 3D par vision active et modélisation géométrique

3.1 Introduction

Pendant les quelques dernières années, la communauté de recherche s'intéressant à la reconnaissance faciale a manifesté des intérêts pour passer de la vision 2D à la vision 3D en vue de doter les systèmes de reconnaissance par des algorithmes plus précis. En effet, les données bi-dimensionnelle ont montré leur insuffisance pour obtenir des taux de reconnaissance et d'authentification suffisamment élevés malgré les progrès en matière de techniques d'analyse de données et de classification utilisées. Par ailleurs, l'information profondeur employée, seule, ou couplée à celle de texture peut présenter une nouvelle modalité biométrique intéressante à étudier. Ci-après un extrait du texte descriptif du projet américain FRGC¹ (*Face Recognition Grand Challenge*), successeur des programmes FERET et FRVT, décrivant les potentiels de cette représentation du visage (*3D Face*).

"...Three-dimensional (3D) face recognition algorithms identify faces from the 3D shape of a person's face. In current face recognition systems, changes in lighting (illumination) and pose of the face reduce performance. Because the shape of faces is not affected by changes in lighting or pose, 3D face recognition has the potential to improve performance under these conditions..."

Le présent travail s'inscrit dans ce contexte de reconnaissance/authentification d'identité moyennant la forme 3D du visage. Nous essayons dans la première partie, de reconstruction tridimensionnelle, de répondre aux besoins, en matière d'acquisition, du procédé de reconnaissance 3D décrit dans le chapitre suivant. En effet, notre technique d'acquisition permet d'obtenir, d'une part, un modèle partiel (2 1/2D) qui sera présenté comme image

¹<http://www.frvt.org/FRGC/>

requête dans le procédé de reconnaissance. D'autre part, le même procédé associé à d'autres opérations d'association de maillage 3D permet d'obtenir les modèles 3D complets qui sont incorporés, dans une première étape, à la base de donnée pour la reconnaissance (la galerie 3D). Dans cette approche nous abordons, outre le problème de reconstruction lui-même, les problèmes suivants : la précision des mesure de profondeur, la fidélité du modèle produit ainsi que sa résolution et le coût du capteur pour être intégré dans une application biométrique. Le prototype réalisé, ainsi que toutes les étapes de notre procédé sont amplement décrits dans ce chapitre. Le module de reconnaissance fera l'objet du chapitre suivant, où nous présenterons l'approche complète dédiée à la reconnaissance moyennant le matching de surfaces faciales.

3.2 Vue d'ensemble de l'approche proposée

L'approche de reconstruction 3D que nous proposons s'insère dans la catégorie des méthodes actives. Elle est destinée principalement à la reconstruction des visages humains pour des finalités de reconnaissance/authentification de personnes. Le choix d'une méthode active s'explique par le fait que les méthodes passives présentent peu de précision et c'est généralement une structure non-dense de l'objet réel qui est présentée alors que l'objectif est de reconstruire, le plus fidèlement possible, une partie ou la totalité du visage. Pour se faire, une source de lumière a été introduite dans le but de créer de la texture artificielle sur le visage. En effet, le visage présente dans la plupart de ses régions, essentiellement celles de peau, une texture lisse et homogène et par conséquent très similaire. Ceci augmente la complexité du problème de mise en correspondance considéré comme problème clé des approches par stéréo-vision. Cependant, et comme on vient de le mentionner dans notre état de l'art, le choix de la lumière à projeter, dans les méthodes actives, est primordial car il agit directement sur les performances du système en terme de précision, densité, rapidité et fiabilité. Pour cela, nous étions rigoureux dans nos choix et la conception de la lumière employée est complétée par d'autres modules de modélisation en vue d'obtenir des modèles 3D de qualité satisfaisante.

A la différence des approches de codification, notre modèle de lumière est employée uniquement comme élément texturant de la scène, dans notre cas le visage. Sa projection vise la discrimination et la localisation de primitives et aucune codification n'a été employée. En effet, il s'agit de projeter successivement et dans un laps de temps très réduit deux modèles de lumières complémentaires. Ceci permet de générer deux courbes sinusoïdales sur les profils d'intensité de chaque ligne épipolaire des images stéréo. La première étape revient donc à trouver les points de frontières de franges avec une bonne précision. Ici, nous présentons un modèle amélioré de détection sous-pixélique de primitives en proposant un nouveau schéma d'interpolation plus adaptée que les anciennes méthodes. Le problème de mise en correspondance est traité moyennant la programmation dynamique adaptée aux données du problème.

En ce qui concerne la modélisation 3D, une première reconstruction grossière du visage

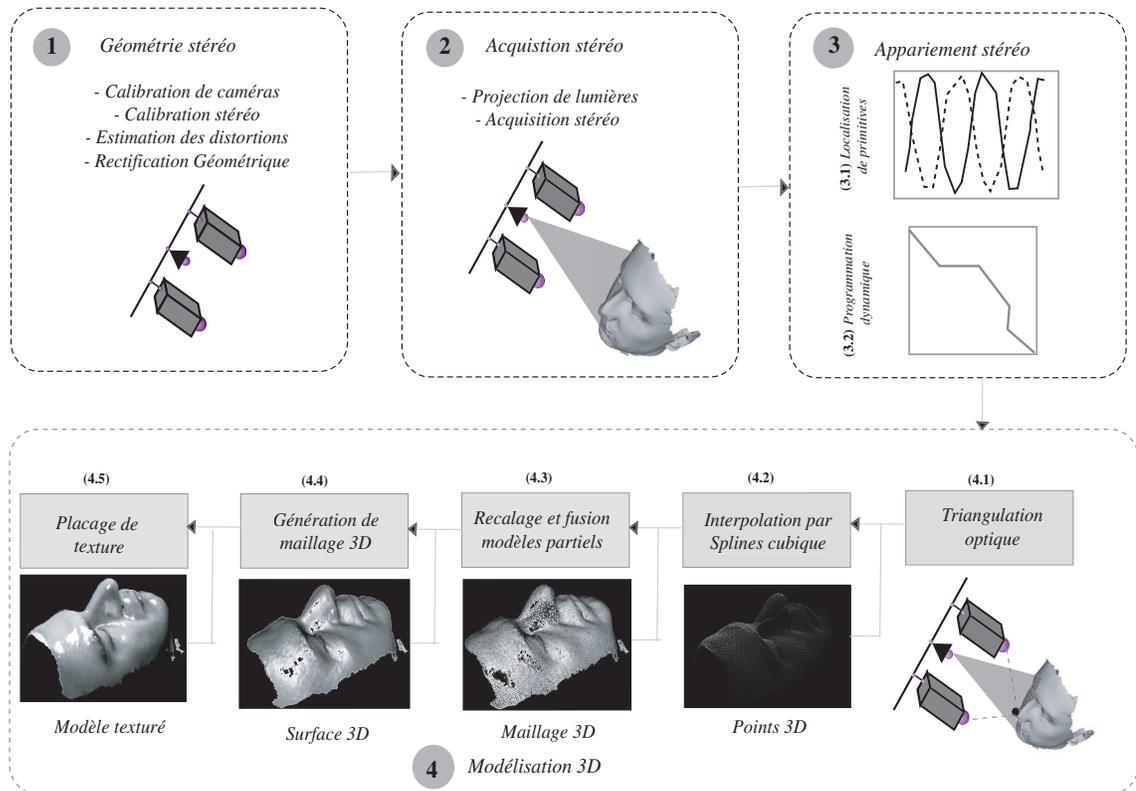


FIG. 3.1 – Pipeline des étapes de notre approche de reconstruction 3D de visages.

est d'abord obtenue moyennant la triangulation optique. L'information profondeur correspondant aux couples de points résultants à l'appariement est retrouvée par triangulation. En fin, une étape d'interpolation dédiée permet de raffiner le résultat obtenu. Cette dernière étape permet d'avoir une reconstruction dense malgré le temps de projection de lumières réduit (seulement deux modèles de lumières sont projetés successivement). L'importance d'un tel mariage entre la vision active et la modélisation géométrique se manifeste dans la complémentarité apportée. Les résultats obtenus montrent la qualité des rendus obtenus par notre approche. Ceci se manifeste surtout au niveau des régions riches en détails sur le visage. Les résultats présentés sont comparables à ceux obtenus par des scanners lasers en terme de qualité de rendus et de précision de mesures.

Plus concrètement, nous passons d'abord, par une étape d'étalonnage qui permet d'estimer les paramètres intrinsèques et extrinsèques ainsi que la géométrie épipolaire du capteur stéréo (figure 3.1, Étape 1). Puis, l'application d'une transformation géométrique dérivée de la géométrie épipolaire, appelée rectification, sur les images nous permet d'obtenir une nouvelle configuration virtuelle du capteur stéréo. Dans cette dernière les axes des caméras sont parallèles et les lignes épipolaires conjuguées sont aussi parallèles. Ceci réduit le problème de recherche de correspondant, d'un problème bi-dimensionnel à un problème mono-dimensionnel. D'autres contraintes d'ordre, d'unicité et de différence de disparités sont aussi introduites, pour restreindre l'espace de recherche de correspondance gauche/droite. L'appariement des primitives ainsi localisées par la technique de projection de lumières complémentaires (figure 3.1, Étape 2), permet de passer à la triangulation optique qui n'est autre que l'intersection des rayons optiques issus de chaque caméra et passant par les couples de points formés (figure 3.1, Étape 3). Finalement, un pipeline d'étapes d'interpolation, de maillage, de recalage des modèles partiels obtenus et de placage de texture finalisent le modèle 3D et apporte du réalisme au visage reconstruit (figure 3.1, Étape 4).

Notre plateforme d'acquisition est composée de deux caméras montées en réseau avec un ordinateur servant à gérer les flux vidéo et à commander le processus de capture d'images de manière synchrone. A l'ensemble nous associons un vidéo-projecteur multimédia employé comme source de lumière texturant à projeter sur le visage du sujet placé en face. Il est à noter que cette configuration trinoculaire (2caméras + 1vidéo-projecteur) est plus flexible que la configuration binoculaire (1caméra + 1vidéo-projecteur) fréquemment utilisée dans les travaux réalisés en vision active. En effet, le système où on substitue la deuxième caméra par une source de lumière, nécessite la calibration du projecteur ainsi que l'établissement de correspondances lumière-image qui sont des tâches plus compliquées à effectuer [Tro95], d'une part, et moins précises, d'autre part. Pour cela, nous avons opté à la structure trinoculaire dans le but d'atténuer les ambiguïtés autour du problème d'appariement qui est sans doute une des étapes clés de notre approche. C'est donc un problème classique d'appariement image-image qui est traité. En outre, la procédure de calibration concerne seulement la paire de caméras qui est un problème classique et souvent abordée dans la littérature.

Dans ce paragraphe nous avons présenté une vue d'ensemble des différentes étapes de notre chaîne de reconstruction 2 1/2D et 3D appliquée aux visages. Les sections suivantes serviront à décrire, en détail, toutes ces étapes de calibration, de projection de lumière, de

mise en correspondance et de modélisation 3D.

3.3 Calibration du capteur stéréo

Bien entendu la première partie de notre chaîne de reconstruction 3D est la calibration du capteur en vue d'extraire ses paramètres servant comme base de la triangulation 3D. Dans ce paragraphe nous allons brièvement décrire le modèle géométrique associé au processus de capture d'images à l'aide d'une caméra. Ce modèle est caractérisé par un certain nombre de paramètres que nous allons estimer par étalonnage (ou calibration, ou calibrage). Nous allons ensuite étudier différentes approches de calibration, ainsi que leur classification dans le but d'effectuer le choix de la méthode la plus adaptée. Enfin, nous allons décrire un capteur comportant deux caméras (capteur stéréoscopique) : le type de capteur utilisé dans notre approche de reconstruction 3D de visages. Nous allons également étudier quelques propriétés de ce système, en particulier la géométrie épipolaire qui nous permettra de faciliter la tâche de mise en correspondance et qui fera l'objet du paragraphe suivant intitulé stratégie de mise en correspondance stéréo.

Calibrer une caméra consiste à déterminer de manière analytique, la fonction qui associe à un point de l'espace tridimensionnel, sa projection dans l'image obtenue par la caméra. Il s'agit bien entendu d'un problème essentiel de la vision artificielle, dès qu'on s'intéresse à la géométrie dans l'espace de la scène observée à travers une ou plusieurs caméras. Puisque dans ce cas, on cherche la position dans l'espace des points dont on connaît uniquement l'image par la projection par la ou les caméras. Le cas qui nous intéresse plus particulièrement dans cette étude est celui du calibrage d'un système stéréoscopique, c'est-à-dire, d'un ensemble de deux caméras rigidement liées et observant la même scène.

3.3.1 Modèle mathématique de caméra

Le modèle d'une caméra est une formulation mathématique de l'ensemble des paramètres physiques qui la caractérisent. Ce modèle comporte des propriétés internes (optiques et géométriques) appelés paramètres intrinsèques et des paramètres caractérisant la position et l'orientation de la caméra appelés paramètres extrinsèques. La plupart des méthodes de calibrage reposent sur le modèle projectif linéaire proposé par Hall [EJCF82]. En effet, non seulement il permet de modéliser la plupart des capteurs projectifs, mais en plus, il permet de simplifier les outils mathématiques mis en jeu pour l'estimation des paramètres du modèle ; le but étant de trouver la relation entre le point 3D de la scène et le point 2D correspondant par projection .

La figure 3.2 illustre les différents stades de la projection du point 3D de la scène observée par la caméra sur l'image. Ces étapes sont les suivantes :

1. La première étape consiste à relier le point P_w du système de coordonnées monde, au système de coordonnées caméra, c'est la transformation monde/caméra.

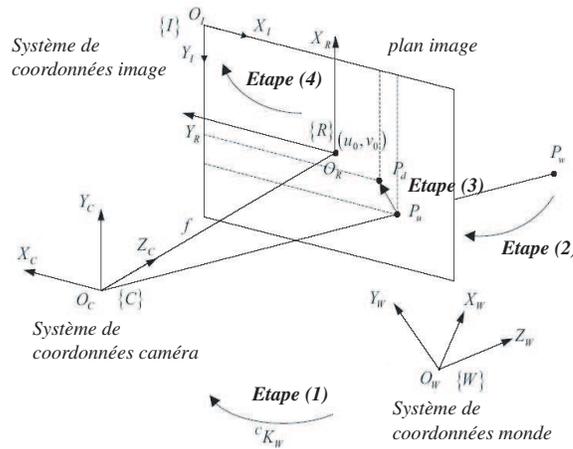


FIG. 3.2 – Relation entre un point 3D et sa projection 2D.

2. La deuxième étape consiste à projeter le point P_w exprimé dans le repère caméra sur le plan image pour obtenir P_u , en utilisant la projection 3D-2D.
3. La troisième étape, illustre le phénomène de distorsion introduit par le capteur au moment de l'acquisition. Le point P_u est alors transformé au point P_d en ajoutant un déplacement $\delta = (\delta_x, \delta_y)$.
4. Finalement, l'étape quatre consiste à changer le système de coordonnées image de telle façon que l'origine du repère soit placée en haut à gauche de l'image.

Dans la suite de cette sous-section, nous allons détailler ces transformations et citer quelques travaux existants rapportant à ce sujet. Soit ${}^a P_b$ un point donné à une étape de sa transformation entre l'étape de départ qui est le monde 3D et l'étape finale qui est sa projection 2D. Ici, a représente le système de coordonnées et b l'étape de transformation avec :

$$a \in \{c = \text{camera}, i = \text{Image}, w = \text{monde}\} \text{ et} \\ b \in \{w = \text{monde}, d = \text{distorsions}, u = \text{sansdistorsions}\}$$

3.3.1.1 Transformation monde/camera

Cette transformation permet de changer le système de coordonnées monde en système de coordonnées image. C'est une transformation rigide composée d'une rotation ${}^c R_w$ et d'une translation ${}^c T_w$. Un point ${}^w P_w = ({}^w X_w, {}^w Y_w, {}^w Z_w)$ exprimé dans le repère monde est transformé dans le repère caméra en ${}^c P_w = ({}^c X_w, {}^c Y_w, {}^c Z_w)$ par l'équation (1) :

$${}^c P_w = {}^c R_w \cdot {}^w P_w + {}^c T_w \quad (3.3.1)$$

Cette étape est souvent ignorée car on confond en général le repère monde avec le repère caméra, sauf dans des applications robotiques où le robot (la ou les caméra) est en mouvement.

3.3.1.2 Projection 3D-2D

Cette étape caractérise la transformation du monde réel 3D au monde projectif associé à la caméra. Soient un point O_r dans le plan image appelé point principal et une droite perpendiculaire au plan image passant par O_r , appelé l'axe optique. Soit un point O_c placé sur l'axe optique à une distance f du plan image. Le point O_r est le centre de projection de la caméra et f est la distance focale. Le point cP_w est projeté sur le plan image le long d'une droite passant par lui-même et par O_c . Le point ${}^cP_u = ({}^cX_u, {}^cY_u)$ est lié au point cP_w par les équations (2) :

$${}^cX_u = f \frac{{}^cX_w}{{}^cZ_w} \quad , \quad {}^cY_u = f \frac{{}^cY_w}{{}^cZ_w} \quad (3.3.2)$$

3.3.1.3 Prise en compte des distorsions

Cette étape est basée sur la modélisation des distorsions introduite par le capteur optique. Les équations suivantes transforment le point cP_u , sans distorsions, en un point cP_d , avec distorsions, δ_x et δ_y représentent les déplacements appliqués :

$${}^cX_d = {}^cX_u + \delta_x \quad , \quad {}^cY_d = {}^cY_u + \delta_y \quad (3.3.3)$$

Chaque méthode de calibrage propose une façon de modéliser ces distorsions dont voici quelques exemples :

- Le modèle de caméra proposé par Faugeras-Toscani [FT87] ne présente pas de distorsions, donc les point cP_u et cP_d sont les mêmes. Dans ce cas :

$$\delta_x = \delta_y = 0 \quad (3.3.4)$$

- Le modèle de Weng [WCH92] considère deux types de distorsions : radiales et tangentielles, le déplacement d'un pixel s'écrit sous la forme :

$$\delta_x = \delta_{xr} + \delta_{xt} \quad (3.3.5)$$

- Dans son modèle, Tsai [Tsa87] ne considère que la distorsion selon la première composante radiale (k_1). En effet, ses études ont montrés que ce coefficient est beaucoup plus important que les distorsions selon les composantes tangentielles et les autres coefficients radiaux. Le déplacement est exprimé par l'équation (6) :

$$\delta_x = k_1 {}^cX_d ({}^cX_d^2 + {}^cY_d^2) \quad , \quad \delta_y = k_1 {}^cY_d ({}^cX_d^2 + {}^cY_d^2) \quad (3.3.6)$$

- Le modèle de Zhang [Zha00] considère les deux premiers coefficients de distorsions radiales, soit les équations suivantes :

$$\delta_x = {}^c X_d [k_1 ({}^c X_d ({}^c X_d^2 + {}^c Y_d^2)) + k_2 ({}^c X_d ({}^c X_d^2 + {}^c Y_d^2))], \quad (3.3.7)$$

$$\delta_y = {}^c Y_d [k_1 ({}^c X_d ({}^c X_d^2 + {}^c Y_d^2)) + k_2 ({}^c X_d ({}^c X_d^2 + {}^c Y_d^2))] \quad (3.3.8)$$

3.3.1.4 Transformation camera/image

Dans le repère bidimensionnel (O_I, X_I, Y_I) associé à l'image, les points image sont mesurés en pixels. Afin de pouvoir écrire la matrice de transformation du repère caméra au repère image, nous devons introduire les paramètres suivants : u_0, v_0 qui sont les coordonnées de O_r (point principal), dans le repère image (mesurées en pixels), k_u le facteur d'échelle vertical (pixel/mm) et k_v le facteur d'échelle horizontal. En effet, les pixels d'une caméra sont rarement carrés. Le modèle de caméra proposé par Faugeras-Toscani [FT87] et Weng [WCH92] utilise l'équation pour caractériser cette transformation, le point ${}^i P_d$ est retrouvé en transformant le point ${}^c P_d$:

$${}^i X_d = -k_u {}^c X_d + u_0 \quad , \quad {}^i Y_d = -k_v {}^c Y_d + v_0 \quad (3.3.9)$$

3.3.1.5 Transformation globale

Toutes les transformations citées ci haut sont caractérisées par une seule matrice de transformation Ω qui exprime le modèle projectif d'une caméra. Cette matrice s'écrit sous la forme suivante (10) :

$$\begin{pmatrix} s {}^i X_d \\ s {}^i Y_d \\ s \end{pmatrix} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \Omega_{14} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & \Omega_{24} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & \Omega_{34} \end{pmatrix} \begin{pmatrix} {}^w X_w \\ {}^w Y_w \\ {}^w Z_w \\ 1 \end{pmatrix} \quad (3.3.10)$$

Cette notation utilise les coordonnées homogènes qui permettent de caractériser la totalité de ces transformations. Dans cette équation, s est un scalaire choisi arbitrairement, ${}^w P_w$ dont les coordonnées homogènes sont $({}^w X_w, {}^w Y_w, {}^w Z_w, 1)$ représente le point 3D exprimé dans le référentiel monde, en appliquant à cette matrice de transformation on obtient le point 2D ${}^i P_d$ de coordonnées homogènes $({}^i X_d, {}^i Y_d, 1)$ exprimé dans le référentiel image.

3.3.2 Au-delà d'une caméra

Nous sommes maintenant en mesure d'étudier un capteur stéréoscopique. Il s'agit tout simplement d'utiliser deux caméras qui observent la même scène. On récupère ainsi deux projections de chaque point de la scène. Soit P est un point de la scène et p et p' sont les projections dans les images de gauche et de droite.

3.3.2.1 Calibrage stéréo

Le calibrage stéréoscopique consiste à déterminer la matrice de transformation entre le repère caméra gauche et le repère caméra droite. Soit Π_s cette matrice qui est composée d'une rotation et d'une translation. La démarche à suivre pour calibrer un tel capteur, c'est-à-dire la détermination de la matrice π_s , est la suivante :

- On commence par calibrer chaque caméra par rapport à une mire unique. Ceci fournit les paramètres intrinsèques et extrinsèques de chaque caméra.
- A l'aide des paramètres extrinsèques, on construit deux matrices B et B' ; la première étant la transformation du repère mire au repère de la caméra droite et la deuxième étant la transformation du repère mire au repère de la caméra gauche. Π_s est obtenue en couplant ces deux transformations. En effet, Π_s est calculée par l'équation (13) :

$$\Pi_s = \Pi'_r \Pi_l^{-1} \quad (3.3.11)$$

La matrice Π_s s'écrit donc comme (14) :

$$\Pi_s = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} r_1 & t_x \\ r_2 & t_y \\ r_3 & t_z \\ 0 & 1 \end{pmatrix} \quad (3.3.12)$$

où $b = (b_x, b_y, b_z)$ représente le vecteur de translation allant de O_c à $O_{c'}$. Il représente les coordonnées du point O_c dans le repère de la caméra droite. La norme de ce vecteur, c'est-à-dire la distance entre les deux caméras, représente la ligne de base B .

$$B = \|\vec{t}\| = \sqrt{t_x^2 + t_y^2 + t_z^2} \quad (3.3.13)$$

3.3.2.2 Relation droite/gauche

Nous voulons maintenant établir une relation entre un point de l'image gauche et son correspondant dans l'image droite. Soit $p = (x, y, 1)$ la projection de $P = (X, Y, Z)$ dans l'image gauche et $p' = (x', y', 1)$ la projection de P dans l'image droite ; p et p' sont liés par les équations (16) :

$$x' = \frac{Zr_1p + t_x}{Zr_3 + t_z}, \quad y' = \frac{Zr_2p + t_y}{Zr_3 + t_z} \quad (3.3.14)$$

On peut donc exprimer la position d'un point de l'image droite en fonction de son correspondant dans l'image de gauche, des paramètres du capteur et de la profondeur Z du point P : c'est la contrainte épipolaire.

3.3.3 Méthode de calibration

En vision par ordinateur, il est souvent utile de connaître un modèle de la caméra pour pouvoir obtenir des informations 3D à partir des images. Dans la littérature, ce sujet est

étroitement abordé et plusieurs approches ont été conçu selon l'application envisagée. Nous essayons dans ce paragraphe d'étudier certaines méthodes de référence ainsi que leurs classification en vue de choisir la méthode la plus adaptée pour estimer le modèle mathématique de notre capteur stéréo.

3.3.3.1 Classification des méthodes de calibration

Le calibrage de caméra consiste donc à déterminer cette matrice de projection Ω donnée par l'équation (3.3.10). Les techniques utilisées sont diverses possédant un point commun qui est exprimé par le modèle projectif proposé par Hall [EJCF82]. Ces méthodes peuvent être classées [Zha00] selon les critères suivants :

- *Calibrage fort/Auto-calibration* : Le principe du calibrage dit fort consiste à utiliser une grille ou tout autre objet de calibrage (appelé mire de calibration), pour lequel les positions de points marqués sur l'objet, appelés points de contrôle, sont connus. Ces points de contrôle peuvent être des coins, des points, des intersections de droites ou toutes autres primitives pouvant être facilement extraites à partir d'images numériques. Les plus anciennes méthodes, utilisent une mire 3D pour effectuer l'estimation des paramètres. Néanmoins, il y a une nouvelle génération de méthodes qui utilisent simplement un modèle planaire [Zha00], en considérant que le repère mire est lié au plan d'équation : $Z = 0$. Ces techniques sont évidemment plus flexible à mettre en œuvre. L'auto-calibration (self calibration) [Y.99], contrairement au calibrage dit fort, n'utilisent que quelques propriétés géométriques de la scènes (points fuites,...)[Y.99], c'est pour cela qu'on l'appelle auto-calibrage. Ces méthodes sont plus flexibles mais beaucoup plus difficiles à mettre en œuvre et elles ont bien plus de paramètres inconnus à définir. Finalement, l'obtention de résultats n'est pas toujours garantie, elle dépend de la géométrie de la scène. Ces méthodes sont bien évidemment moins précise que les méthodes de calibration par rapport à une mire dont on connaît la géométrie 3D.
- *Méthodes linéaires/non-linéaire* : Les méthodes linéaires, dont les méthodes de Hall [EJCF82] et de Faugeras-Toscani [FT87], ne tiennent pas compte des distorsions introduites par le système d'acquisition, donc, le problème revient à résoudre un système linéaire. Les paramètres optimaux sont obtenus en minimisant une fonction qui représente une erreur quadratique de la forme (3.3.15) :

$$\xi = \sum_{i=0}^n \sum_{j=0}^m \|m_{ij} - \hat{m}(A, R_i, t_i, M_j)\| \quad (3.3.15)$$

où :

- n : le nombre de positions de la mire, m : nombre de point de contrôle
- m_{ij} est le point de contrôle j détecté sur l'image i
- \hat{m} est la projection du point M_j sur l'image i à l'aide des paramètres estimés

- (R_i, t_i) est la transformation rigide de la mire dans sa position i par rapport à la caméra (paramètres extrinsèques)

En revanche, les méthodes non-linéaires modélisent ces distorsions, et par la suite, le problème n'est plus linéaire. Dans ces méthodes, on fait recours à des techniques d'optimisation non linéaire (Newton-Raphson, Levenberg-Marquard, descente de gradient, ...) pour minimiser une erreur quadratique de la forme suivante (3.3.16) :

$$\xi = \sum_{i=0}^n \sum_{j=0}^m \|m_{ij} - \hat{m}(A, R_i, t_i, k_1, \dots, k_h, Mi)\| \quad (3.3.16)$$

- n : le nombre de positions de la mire, m : nombre de point de contrôle
- m_{ij} est le point de contrôle j détecté sur l'image i
- \hat{m} est la projection du point M_j sur l'image i à l'aide des paramètres estimés
- (R_i, t_i) est la transformation rigide de la mire dans sa position i par rapport à la caméra (paramètres extrinsèques)
- le vecteur (k_1, k_2, \dots, k_h) représentent les coefficients de distorsions à estimer.

3.3.3.2 Méthode de calibration

Il est clair que l'utilisation d'une méthode de calibration linéaire est moins précis qu'une méthode non-linéaire vue que cette dernière prend en compte les distorsions optiques introduite réellement par le capteur d'acquisition. Néanmoins, le résultat des approches non-linéaires dépend du choix des valeurs initiales des paramètres. Une solution à ce problème consiste donc à évaluer tous les paramètres du capteur (saufs les coefficients de distorsions) en utilisant une des techniques linéaires décrites auparavant comme solution initiale où les distorsions sont initialisées à un vecteur nul. Cela permet une convergence plus rapide et garantie vers la solution souhaitée. Ces modèles dont ceux de Tsai [Tsa87], Faugeras-Toscani [FT87] avec distorsions, Weng [WCH92], Zhang [Zha00], Bouguet [Y.99], diffèrent par leurs modélisation des distorsions.

Dans cette étape de notre travail, l'étalonnage des caméras est effectué en appliquant une technique utilisant une mire de calibration et elle est non-linéaire (i.e. tient en compte des distorsions). Dans cette approche, nous calculons, dans un premier temps, une solution initiale par une méthode linéaire (*closed-form solution*) [Zha00], puis, nous utilisons un algorithme d'optimisation itératif (descente de gradient) pour effectuer le raffinement des résultats. Cette méthode permet donc de tenir compte les distorsions radiales et tangentielles à la fois, et donc estime quelques coefficients de ces vecteurs. Elle fait partie des techniques flexibles qui utilisent un modèle de calibration planaire. Une implémentation de cette technique de calibration fait partie de la librairie OpenCV² mis à disposition par Intel sur Internet. Les routines de cette boîte à outils sont écrit en *Matlab* appelé d'ailleurs *Camera Calibration Toolbox for Matlab*³. La calibration d'un capteur binoculaire est également

²<http://sourceforge.net/projects/opencvlibrary/>

³www.vision.caltech.edu

disponible dans la même boîte à outils (voir figure 3.4). Les caméras sont d'abord calibrées par rapport à une mire unique, puis, l'association des paramètres extrinsèques permet de calculer les transformations rigides (matrice de rotation R et vecteur de translation t) entre elles.

Par étalonnage, nous estimons les modèles de projection de chaque caméra, ainsi que la relation spatiale qui les relie. Les paramètres à estimer sont alors les suivants :

- *Distance focale (Focal length)* : f_c exprimé en pixels (taille : 2x1).
- *Point principal (Principal point)* : cc (taille : 2x1).
- *Coefficient d'angle (Skew coefficient)* : S_{coef} ce coefficient définit l'angle entre les axes u et v du référentiel liée à l'image : α_c (taille : 1x1).
- *Distorsions (Distortions)* : coefficients de distorsion de l'image (distorsions radiales et tangentielles) k_c (taille : 5x1).
- *La transformation spatiale (R, t)* : Rotation + translation entre les deux caméras.

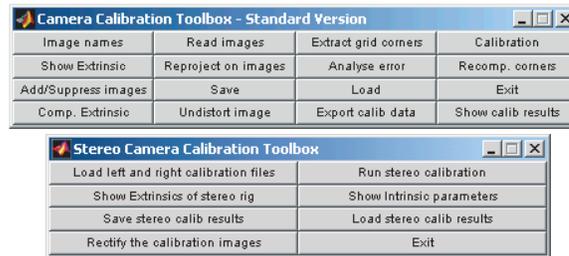


FIG. 3.3 – Interface de la boîte à outils de calibration utilisée pour la calibration de notre capteur stéréo.

La figure 3.4 illustre la disposition spatiale des caméra, gauches et droites, et les positions de la mire utilisées pour la calibration (mire plane sous forme de damier). Une fois notre capteur binoculaire est calibré, c'est à dire que nous disposons de tous les paramètres intrinsèques et extrinsèques, nous pouvons passer à la seconde étape de notre chaîne de reconstruction 3D, à savoir l'appariement stéréo.

3.4 Stratégie de mise en correspondance stéréo

La question à la quelle nous essayons de répondre dans cette section est la suivante : quelle stratégie d'appariement à mettre en œuvre pour décider qu'une paire de primitives gauche/droite se correspondent et représente le même point dans l'espace ? En essayant de répondre à cette question nous allons développer deux stades de traitements qui ensemble constitueront notre stratégie d'appariement. Le premier vise la discrimination de primitives et le second permet leurs mises en correspondance. Des critères intrinsèques à géométrie stéréo ainsi que des critères dérivés de la géométrie du visage sont également exploités pour aider à la résolution de ce problème.

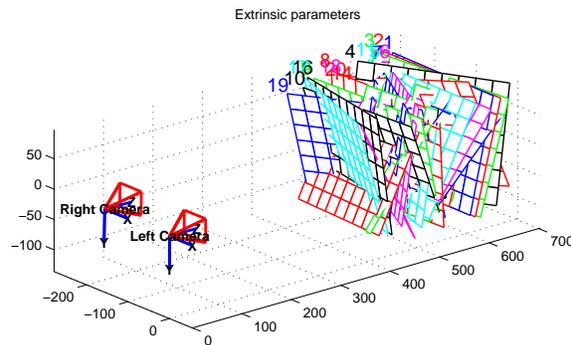


FIG. 3.4 – Disposition des caméras par rapport à la mire en mouvement.

3.4.1 Géométrie épipolaire et rectification géométrique

La géométrie épipolaire est notamment la contrainte la plus intéressante à introduire du fait qu'elle ramène le capteur utilisé à une configuration virtuelle simple. Dans cette dernière, les caméras sont parfaitement parallèles et sont situés à la même hauteur. Ceci permet de limiter la combinatoire du problème de mise en correspondance dans la mesure où il le ramène d'un problème de recherche bidimensionnel à un problème de recherche monodimensionnel. Après estimation de la géométrie épipolaire et l'application d'une transformation dérivée, appelé rectification géométrique, les lignes épipolaires conjuguées sont parallèles et les primitives qui se correspondent ont nécessairement la même ordonnée.

En pratique, il y a deux méthodes qui permettent de calculer la géométrie épipolaire d'un couple de caméras : Soit en calibrant les deux caméras par rapport à une mire commune, soit en estimant la matrice fondamentale à partir d'un certain nombre d'appariements gauche/droite. Dans notre approche, c'est la calibration forte qui a été utilisée et donc c'est la première solution qui a été retenue pour la calibration stéréo. Cette méthode est plus précise pour calculer les paramètres des caméras et la transformation rigide entre elles. De même, il y a plusieurs méthodes destinées à la rectification des images dont la plus connue est de les projeter sur un plan unique, appelé plan de rectification. En effet, on fait subir à chaque image une transformation de façon à obtenir une paire d'images stéréo coplanaire et parallèles à la droite passant par les centres de projection des deux caméras.

Jusqu'à présent nous avons essayé de réduire la complexité du problème de mise en correspondance mais il n'est pas encore totalement résolu. En effet, il reste encore une dimension à balayer pour rechercher le correspondant d'une primitive donnée. Dans le but de réduire encore l'ambiguïté de correspondance, nous avons opté dans un premier temps, à la projection d'une lumière en vue de texturer le visage à numériser. Puis, par une technique

d'extraction introduite dans la section suivante, nous localisons avec haute précision des primitives qui se correspondent. Finalement, nous adaptons le principe de programmation dynamique pour traiter le problème d'alignement de séquences stéréo et qui fait face au phénomène d'occultation qui est habituellement la source majeure de complication.

3.4.2 Lumière et détection sous-pixélique de frontières de franges

La précision des méthodes passives de reconstruction tridimensionnelle est généralement limitée à un niveau pixélique [Tro95]. La projection d'une lumière hétérogène sur la scène, permet en plus d'une atténuation considérable du problème d'ambiguïté, de s'affranchir de cette limitation. En effet, la détection de primitives peut se faire moyennant une dérivée effectuée dans le sens orthogonal aux raies projetées ou en localisant les centres de franges de couleurs projetées. Cette étape de localisation de primitives est cruciale dans la formation de l'image de profondeurs vue la précision apportée.

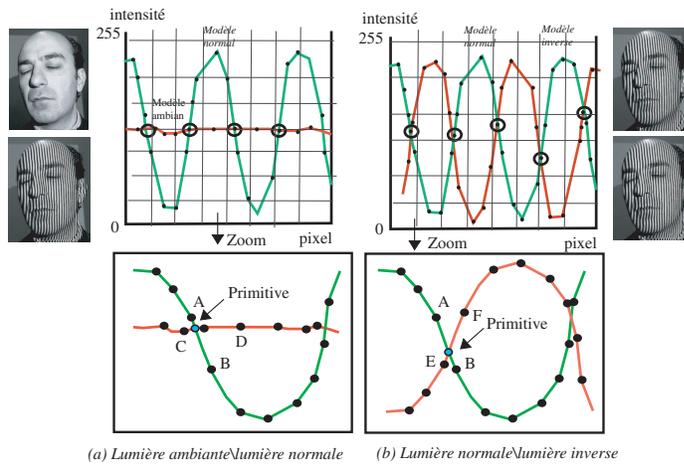


FIG. 3.5 – Principe de projection de lumière et de détection sous-pixélique de primitives.

Dans [Tro95], deux moyens ont été présentés pour effectuer la détection sous-pixélique de primitives pour une reconstruction plus précise. Le premier étant la projection d'une série de franges noires et blanches (dans un même modèle de lumière) sur l'objet à numériser. La détection de points de contour se fait par localisation des intersections avec une image obtenue sans lumière structurante (voir figure 3.5 (a)). Une méthode similaire a été déjà proposée par Yé dans [Ye88]. Comme illustré dans la figure 3.5 (a), c'est l'intersection des segments AB et CD qui détermine la position des primitives. Le deuxième schéma propose la projection de deux lumières complémentaires qui s'intersectent en un ensemble de primitives (voir figure 3.5 (b)). Dans [Tro95], l'auteur affirme l'efficacité de ce dernier schéma et son invariance par rapport aux changements de contrastes. Par conséquent, l'utilisation du premier modèle seul peut engendrer une certaine déviation dans la position des contours.

Dans les deux cas, le modèle d'interpolation entre les pixels d'un profil d'intensité proposé est linéaire. D'autres modèles de localisation de primitives ont été présentés dans la littérature. Parmi eux, nous pouvons citer la localisation des centres des raies projetées comme ce qui a été présenté dans [BA00]. Ces centres sont représentés par les pics des sinusoïdes générées, sur les profils d'intensité, suite à la projection de telles lumières. De même, les méthodes qui émettent sur les objets des points en couleur [MOC⁺98] ou en niveaux de gris et qui estime la surface d'intensité de ce point par une surface gaussienne et d'en déduire le pic ou le maxima.

Le modèle que nous proposons, dans notre approche, est conçu comme une succession de deux lumières complémentaires constituées d'ensembles de raies noires et blanches (comme illustré dans la figure 3.5 (b)). Néanmoins, le modèle d'interpolation utilisé se présente comme des tranches locales de polynômes de troisième ordre continus et de dérivés continues sur les extrémités des intervalles : c'est le modèle des *splines cubiques*. L'utilisation d'un tel modèle nous semble plus adapté que l'utilisation d'un modèle linéaire dans le cas des images. En effet, ces méthodes d'interpolation sont les plus utilisées pour augmenter les résolutions des images digitales. Ils génèrent des courbes lisses et continues en passant par les points d'origines. Ils ont montré également leur efficacité par rapport aux transformations géométriques effectuées sur les images, notamment les rotations, ce qui n'est pas le cas du modèle linéaire qui calcule simplement la moyenne entre les deux extrémités d'un intervalle donné. En outre, les courbes issues d'une interpolation linéaire présentent des discontinuités au niveau des points de contrôle.

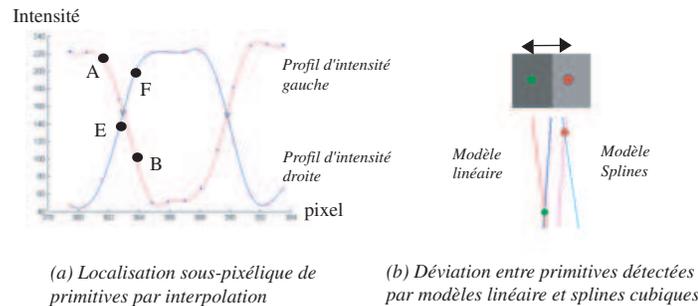


FIG. 3.6 – Déviation entre intersection par modèle linéaire et modèle de splines cubiques.

Le problème de détection sous-pixélique de primitives est ramené en quelques sortes à une étape d'augmentation de résolutions des images d'origines puis à une localisation des contours sur les images résultantes de hautes résolutions. La figure 3.6 (a) illustre l'aspect lisse introduit par ce modèle d'interpolation ainsi que la précision sous-pixélique avec laquelle sont détectées ces primitives. Dans cette figure, les points en gras par lesquelles passent les courbes d'interpolation représentent les valeurs d'intensité des pixels d'un même segment épipolaire. Les courbes qui les joignent sont des polynômes d'interpolation de degrés

3 par morceaux. Les points de contours recherchés sont donnés par les intersections des segments AB et EF . Nous pouvons voir également sur la figure 3.6 (b) les améliorations apportées par l'introduction des modèles de splines cubiques pour optimiser les déviations au niveau des positions des frontières de franges. Les modèles de splines cubiques semblent plus logiques à introduire, les résultats de cette amélioration appliquée à la procédure de détection de primitive se traduisent par une continuité verticale parfaite des points 3D reconstruits. Ceci valide ainsi la substitution du modèle linéaire classique par des modèles de splines cubiques plus adaptés.

Dans la figure 3.7, une illustration de projection de modèles de lumières complémentaires effectuées en stéréo sur le visage est donnée. Des extraits de profils d'intensité des images avec les points d'intersections sont également présentés. Comme affiché sur cette figure, le modèle d'interpolation employé permet de générer des sinusoïdes continues dont les valeurs sont continues. Les intersections sont identifiées par soustraction des deux signaux l'un de l'autre et la comparaison par rapport à un seuil fixé. La présente méthode assure, comme décrit au par avant, la continuité verticale (ou continuité figurale) des primitives détectées. La différence par rapport aux techniques classiques de binarisation est bien entendu ce phénomène de discontinuité. En effet, en se limitant à un niveau pixelique, un effet d'escalier va se présenter sur les images de primitives. Ceci est dû à la restriction faite sur des entiers dans le traitement et génère, par conséquent, des discontinuités verticales sur les primitives résultats qu'on retrouve forcément sur le modèle 3D reconstruit. Un petit gain de précision sur les images stéréo mais qui peut être beaucoup plus significatif sur le modèle reconstruit a été introduit par l'emploi des splines cubiques pour l'interpolation.

3.4.3 Appariement par programmation dynamique

Nous disposons à cette étape, et pour chaque paire de lignes épipolaires conjuguées, de deux ensembles de primitives à mettre en correspondance, sachant qu'il peut y avoir des occlusions et que certaines parties dans l'image gauche n'apparaissent pas dans l'image droite et inversement. Pour résoudre ce problème en tenant compte de ces phénomènes dus à la projection, nous faisons appel au concept mathématique de programmation dynamique. Cette méthode d'optimisation a été employée pour la première fois en vue de comparer deux signaux en reconnaissance de la parole. Elle a été introduite en stéréovision pour la première fois par Baker et Binford en 1981 [BB81] pour être réutilisée plus tard par Otha et Kanade en 1985 [OK85]. Ces derniers l'ont exploitée pour mettre en correspondance d'une manière dense des images aériennes. Ils ont introduit, en plus, une contrainte de continuité figurale (*inter-scanline constraint*) qui fournit une contrainte supplémentaire entre les lignes épipolaires traitées indépendamment dans une première étape (*intra-scanline constraint*). Le principe de la programmation dynamique est largement utilisé pour résoudre des problèmes d'optimisation et surtout d'alignement de séquences de caractères ou en biologie pour l'alignement de séquences d'ADN ou de protéines.

Dans ce qui suit, nous donnons une formulation mathématique du problème ainsi que l'adaptation de ce concept pour résoudre le problème d'appariement dans notre technique

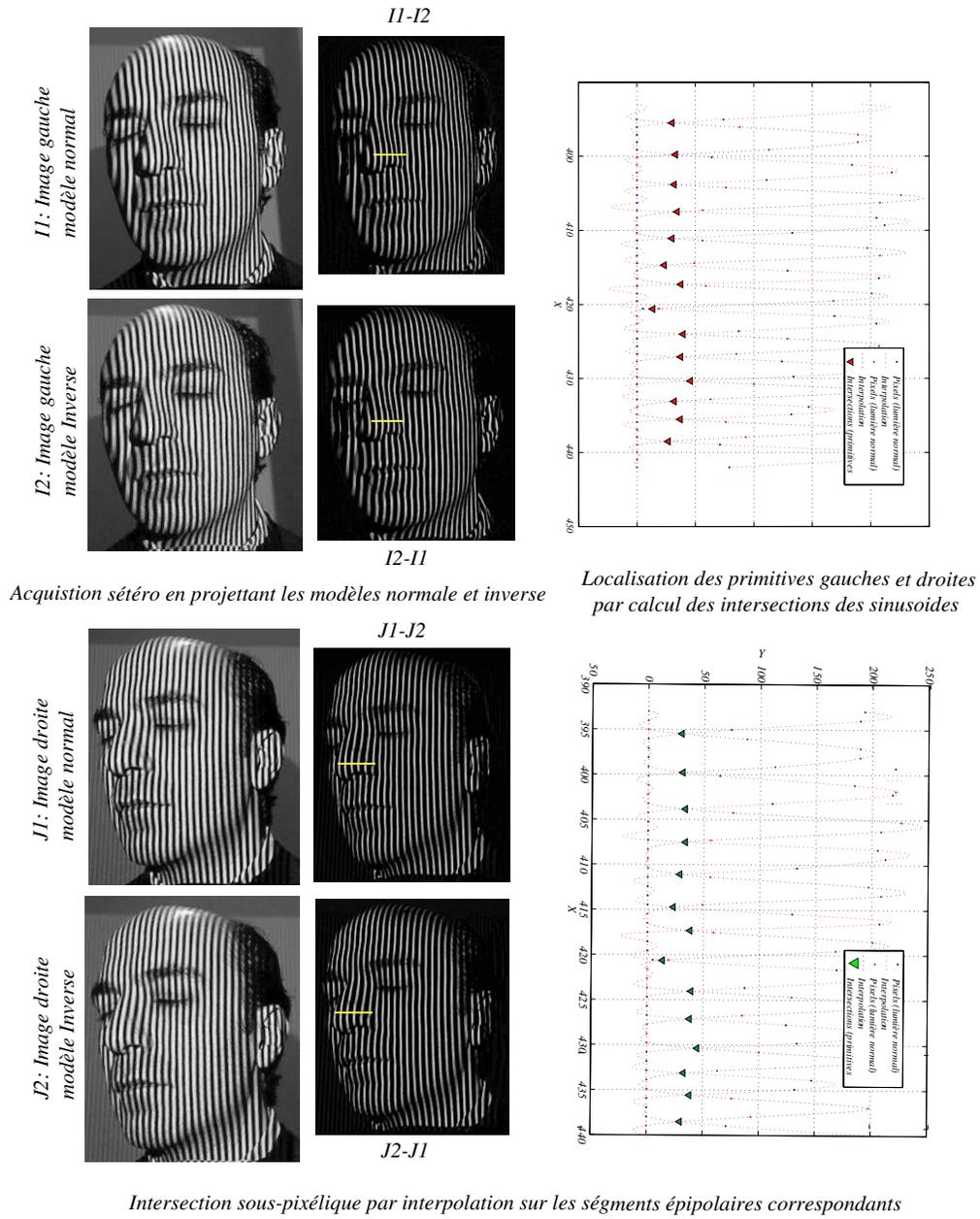


FIG. 3.7 – Principe de projection de lumière et de détection sous-pixélique de primitives.

active. L'approche par PD que nous proposons est décrite amplement dans la suite. Des arguments par rapport à l'exploitation de certains critères introduits dans cette approche sont aussi présentés notamment les contraintes d'ordre, d'unicité et de différence de disparité.

3.4.3.1 Formulation mathématique du problème

Soit $A = a_1, a_2, \dots, a_M$ et $B = b_1, b_2, \dots, b_N$ deux ensembles de primitives localisées sur deux lignes épipolaires conjuguées et à mettre en correspondance. La séquence A contient M primitives alors que la séquence B contient N primitives. Chaque primitive, que nous désignons aussi par symbole, est caractérisée par une valeur associée (son voisinage d'intensité dans l'image). Le but maintenant est d'aligner ses deux séquences provenant d'images stéréo et par conséquent similaires sachant qu'il peut y avoir des occlusions et que certaines primitives n'apparaissent pas dans les deux prises de vue.

Selon plusieurs auteurs comme Myer et Miller [Myer and Miller 1989], un tel alignement est équivalent à un problème de recherche du chemin le plus court dans un graphe donné. Le graphe d'alignement entre A et B , noté $G_{A,B}$ peut être représenté comme suit : Les noeuds de $G_{A,B}$ sont les pairs (i, j) avec $i \in [1, M]$ et $j \in [1, N]$ alors que les arcs sont labellisés par les mesures de similarité entre des primitives de A et de B . Le problème consiste donc à déterminer pour chaque couple (i, j) des sommets, le plus court chemin, s'il existe, qui joint la primitive a_i de A à la primitive b_j de B .

Le principe de résolution par programmation dynamique consiste à construire progressivement une matrice, appelé matrice de programmation dynamique dans laquelle chaque case contient la distance cumulative entre les sous séquences $A_i = a_1, a_2, \dots, a_i$ et $B_j = b_1, b_2, \dots, b_j$. Pour obtenir le meilleur alignement entre A_i et B_j il suffit d'explorer cette matrice et de chercher le plus court chemin joignant les primitives gauches/droites. Le chemin global et optimal est l'association des chemins partiels qui peuvent être :

- De (a_{i-1}, b_{j-1}) à (a_i, b_j) noté $(\phi_{i,j})$ si
 $\max(dp(a_i, b_j), dp(a_{i-1}, b_j), dp(a_i, b_{j-1})) = dp(a_i, b_j)$.
- De (a_{i-1}, b_{j-1}) à (a_{i-1}, b_j) noté $(\phi_{i-1,j})$ si
 $\max(dp(a_i, b_j), dp(a_{i-1}, b_j), dp(a_i, b_{j-1})) = dp(a_{i-1}, b_j)$.
- De (a_{i-1}, b_{j-1}) à (a_i, b_{j-1}) noté $(\phi_{i,j-1})$ si
 $\max(dp(a_i, b_j), dp(a_{i-1}, b_j), dp(a_i, b_{j-1})) = dp(a_i, b_{j-1})$.

Quel score de similarité entre les primitives devons nous définir pour pouvoir décider quel chemin doit-on emprunter ? La réponse à cette question et l'application de la PD à l'appariement stéréo actif constituent l'objet de la sous section suivante.

3.4.3.2 Application à l'appariement stéréo

Dans notre cas d'application, la programmation dynamique permet de déterminer une solution optimale comme le plus court chemin (de coût maximal) dans l'espace des primitives. Les séquences à matcher seront représentées par les primitives gauches/droites détectées sur chaque paire de ligne épipolaires conjuguées. Un chemin partiel consistera en une séquence de paires de primitives, chaque paire correspond à un élément de la matrice des coûts. Il sera défini en fonction du maximum entre trois valeurs. Le coût d'un chemin partiel sera le score de similarité entre deux primitives calculé à partir d'une mesure de corrélation centrée et normalisée sur les voisinages des primitives. La solution globale (chemin optimal) est défini par le coût global et cumulatif des coûts des chemins partiels (3.4.1) :

$$\sigma(\Phi_{i,j}^*) = \begin{cases} 0, & \text{si } i = j = 0, \\ \max \left\{ \begin{array}{l} \sigma(\Phi_{i-1,j-1}^*) + score(e_i, q_j) \\ \sigma(\Phi_{i-1,j}^*) + occlusion \\ \sigma(\Phi_{i,j-1}^*) + occlusion \end{array} \right\}, & \text{sinon} \end{cases} \quad (3.4.1)$$

Où $\sigma(\Phi_{i,j}^*)$ est le coût cumulatif optimal, $\Phi_{i,j}^*$ est le chemin optimal recherché. Pour en tenir compte du problème d'occlusion, un terme pénalisant la solution globale a été introduit et est modélisé par la constante *occlusion*. Le terme $score(q_i, e_i)$ est le score de similarité calculée pour la paire de primitives (q_i, e_i) . Ainsi, la fonction coût est une matrice dont les lignes sont indexées par le segment épipolaire gauche et les colonnes par le segment épipolaire droite (voir figure 3.9).

L'application de la programmation dynamique permet l'exploitation de quelques critères mais aussi doit vérifier d'autres sur la scène à reconstruire :

- *La contrainte d'ordre* : La contrainte épipolaire nous a permis de réduire l'espace de recherche de correspondant à un espace monodimensionnel. Une deuxième contrainte qui permet de réduire encore la combinatoire du problème : c'est la contrainte d'ordre ou de monotonie qui dépend de la nature de la scène à numériser. Si on considère un certain nombre de points (A, B, C, D) placés sur le visage réel, les projections de ces points par les caméras gauche (a, b, c, d) et droite (a', b', c', d') sont rangées en ordre complètement inverse, comme illustre la figure 3.8.

La programmation dynamique implémente d'une manière intégrée cette contrainte de respect d'ordre de primitives ou de monotonie. En effet, les primitives sont placées par ordre croissant de leurs positions dans les images et les seuls déplacements autorisés dans la recherche du plus court chemin dans la matrice des coûts ne contre dit pas cette hypothèse et permet de suivre cet ordre. L'ordre est par conséquent respecté.

- *La contrainte d'unicité* : Ce critère impose le fait que chaque point de l'image gauche doit avoir un seul correspondant, s'il existe, dans l'image droite et inversement. Cette contrainte est également présente dans la procédure d'appariement moyennant la programmation dynamique. En effet, dans la matrice des coûts cumulative chaque primitive de l'image gauche est matchée avec une et une seule primitive de l'image droite.

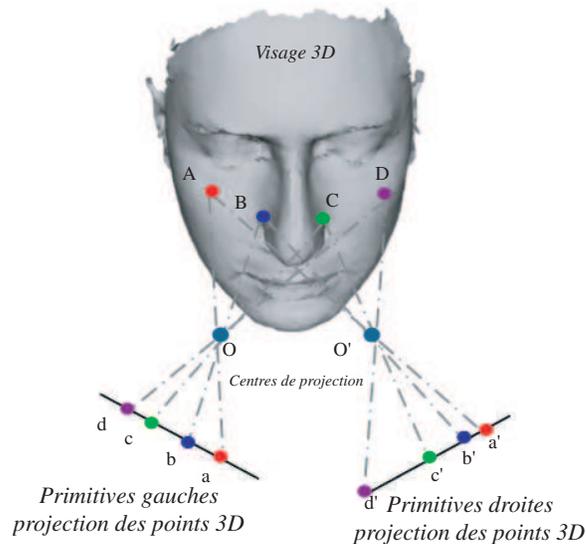


FIG. 3.8 – Contrainte d'ordre (monotonie) : cas du visage.

Si la primitive est considérée comme occultée d'une des deux cotés gauche ou droite, l'algorithme barre cette primitive en allant voir la suivante. Par conséquent, chaque élément d'une séquence n'intervient qu'une seule fois dans cette procédure d'appariement.

- *La contrainte de différence de disparité* : Une procédure d'optimisation de l'algorithme de la programmation dynamique consiste à restreindre l'espace de recherche de correspondances aux cases situées sur le champ diagonal de la matrice des coûts. L'exploration de cette matrice se limite donc aux termes diagonales, les autres termes ne sont pas calculés et sont initialisés à des valeurs de minimum de similarité. En effet, la profondeur d'un point est associée à sa disparité et par conséquent la variation de la profondeur est liée directement à la variation de la disparité. En allant loin des correspondances diagonales, les valeurs de disparités augmentent au fur et à mesure et d'une manière exponentielle. Le rejet de ces cases lointaines de la diagonale de la matrice est due à la cohérence des valeurs 3D sur la surface faciale. En effet, deux points 3D physiquement voisins sur le visage ont nécessairement des valeurs de disparités correspondantes proches. L'algorithme de programmation dynamique présente une complexité de $O(MN)$ où M et N représente le nombre de primitives gauches et droites respectivement. A l'issue de cette optimisation, la complexité est de $O(M * \rho * \rho)$ où $[-\rho, \rho]$ présente la taille, en nombre de primitives, de la fenêtre de calcul des scores de similarité sur les termes diagonales.

La version optimisée de l'algorithme décrite dans la section précédente a été mis au point et implémenté. Plusieurs configurations figures comme résultats de notre application de l'algorithme de programmation dynamique selon la présence de l'occultation à droite,

à gauche ou l'absence dans la plupart des cas. Malheureusement, cette technique comme d'autres destinées à la résolution du problème de mise en correspondance n'est pas à cent pour cent fiable. Il y a, en effet, quelques défauts liés surtout à la présence de l'occultation et la mesure de similarité entre les symboles. Le problème principal lié à l'emploi de la programmation dynamique est la propagation des faux appariements issus des paires de lignes épipolaires conjuguées. En effet, la solution globale dépend des solutions partielles et par conséquent si un faux appariement généré en cours de route le reste des appariements y dépendent est donc ils sont aussi faussement effectuées. A ce stade nous n'apportons pas de solution mais nous agissons plutard dans l'étape de modélisation 3D. Les lignes faussement appariées présentent ainsi des valeurs de profondeurs non cohérentes avec le reste et là la solution et de supprimer toute la ligne 3D correspondantes.

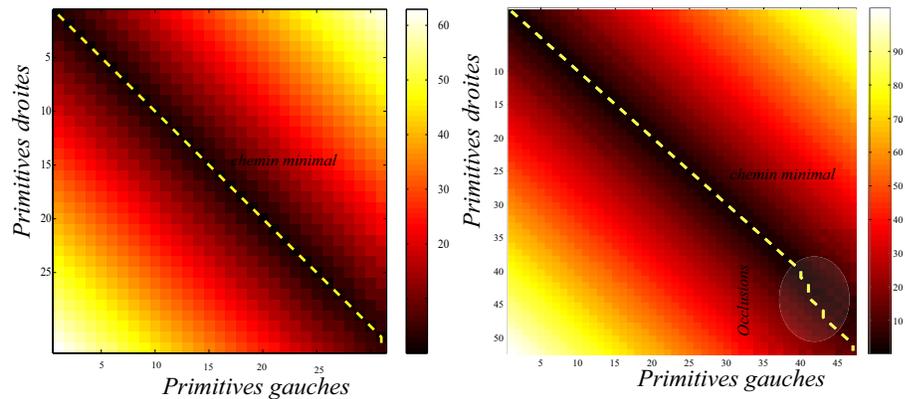


FIG. 3.9 – Exemple de calcul de matrice des coûts et du chemin optimal : (A) cas d'absence d'occlusion, (B) cas de présence d'occlusion.

3.5 Modélisation partielle du visage

A ce stade nous disposons de plusieurs paires de points qui se correspondent et qui représentent chacune deux différentes projections d'un point 3D situé sur le visage à numériser. La première étape de la modélisation tridimensionnelle que nous proposons s'effectue par triangulation optique. Les rayons optiques issus des centres de projections des deux caméras et joignant la paire de primitives appariées se croisent au point 3D recherché. La deuxième étape consiste à prendre les points 3D produit par l'étape précédente et à appliquer un modèle d'interpolation dans le but d'améliorer la résolution et la qualité du rendu résultat. En effet, la reconstruction issue de la triangulation optique est grossière. Elle est dense sur l'axe des ordonnées, néanmoins elle est moins dense sur l'axe de l'abscisse vu la conception de la lumière projeté au début. Nous faisons appel donc à ce stade à un modèle d'interpolation très connu en modélisation graphique qui est le modèle des Splines cubiques. Ces modèles

sont les plus adaptés du fait que leurs courbes d'interpolation possède un caractère lisse et continue tout en passant par les points d'origine obtenus par triangulation optique. En plus, contrairement aux courbes d'interpolation d'ordre élevé, ces polynômes par morceau d'ordre 3 ne présente pas de fortes oscillations et s'adapte parfaitement.

Dans cette section, nous présentons donc toutes les phase nécessaire pour produire un modèle partiel du visage y compris l'étape de triangulation stéréo, d'interpolation par splines cubiques et de maillage des points en vue de générer une forme cohérente. La modélisation entière du visage sera traitée dans la section suivante et est fondée sur cette étape pour produire différents modèles partiels ou 2 1/2D.

3.5.1 Triangulation optique

Notre capteur stéréoscopique est calibré et nous disposons de tous les paramètres intrinsèques de chaque caméra ainsi que la matrice de transformation épipolaires qui les relie. Nous disposons également des paires de primitives qui ont été mises en correspondance moyennant notre algorithme de programmation dynamique. Nous somme donc au mesure de calculer la position dans l'espace du point 3D correspondant en résolvant les deux ensemble d'équations qui relie les points dans les images stéréo $p(u, v)$ et $p'(u', v')$ respectivement pour les images gauche et droite au point 3D $P(X, Y, Z)$ (équations (3.5.1) et (3.5.1)).

$$\begin{aligned} u &= \frac{\Omega_{11}X + \Omega_{12}Y + \Omega_{13}Z + \Omega_{14}}{\Omega_{31}X + \Omega_{32}Y + \Omega_{33}Z + \Omega_{34}} \\ V &= \frac{\Omega_{21}X + \Omega_{22}Y + \Omega_{23}Z + \Omega_{24}}{\Omega_{31}X + \Omega_{32}Y + \Omega_{33}Z + \Omega_{34}} \end{aligned} \quad (3.5.1)$$

$$\begin{aligned} u' &= \frac{\Omega'_{11}X + \Omega'_{12}Y + \Omega'_{13}Z + \Omega'_{14}}{\Omega'_{31}X + \Omega'_{32}Y + \Omega'_{33}Z + \Omega'_{34}} \\ V' &= \frac{\Omega'_{21}X + \Omega'_{22}Y + \Omega'_{23}Z + \Omega'_{24}}{\Omega'_{31}X + \Omega'_{32}Y + \Omega'_{33}Z + \Omega'_{34}} \end{aligned} \quad (3.5.2)$$

On peut également reconstruire le point $P(X, Y, Z)$ dans le repère de la caméra de gauche en utilisant l'équation. Dans ce cas, les coordonnées de P sont données par :

$$\begin{aligned} x' &= \frac{(r_{11}x + r_{12}x + r_{13}x)Z + t_x}{(r_{31}x + r_{32}x + r_{33}x)Z + t_y} \\ X &= x \frac{Z}{f} \\ Y &= y \frac{Z}{f} \end{aligned} \quad (3.5.3)$$

Finalement on peut constater que lorsque les images ont été rectifiées la matrice décrivant la transformation gauche droite est réduite à une translation et on aura (3.5.4) :

$$\frac{B + x' - x}{Z - f} = \frac{B}{Z} \quad (3.5.4)$$

Le point P peut alors être reconstruit dans le repère rectifié de la caméra gauche grâce aux équations suivantes :

$$\begin{aligned} X &= x \frac{Z}{f} \\ Y &= y \frac{Z}{f} \\ Z &= f \frac{B}{x' - x} \end{aligned} \quad (3.5.5)$$

3.5.2 Interpolation via des modèles de *splines cubiques*

Les points reconstruits par triangulation optique seront exprimés dans le repère d'une des deux caméras. Comme la localisation de primitives dans les images est dense dans la direction Y et est semi-dense dans la direction X des images stéréo, nous essayons de remédier à ce problème par prédiction des points intermédiaires entre les primitives selon l'axe X . Notre modèle de prédiction repose sur des polynômes d'interpolation locaux de degrés 3 vérifiant certaines conditions de continuité et de dérivabilité sur les points des extrémités des segments : c'est les splines cubiques. Nous retenons cette solution vu la bonne adaptation de ses fonctions d'interpolation, dans le cas d'une surface lisse et continue entre autre le visage.

Dans ce qui suit nous présentons d'abord la méthode d'interpolation par splines cubiques et son intérêt dans la modélisation graphique. Ensuite nous présentons l'application des ces modèles dans notre cas et nous illustrons par quelques résultats qui valident ce choix. Cette solution de prédiction substitue en quelques sortes un déplacement (shifting) du modèle lumineux dans l'espace qui peut être effectué pour densifier l'acquisition sur l'axe X . Cette dernière opération nous semble inadéquate dans la mesure où l'objet à numériser peut bouger pendant le temps de l'acquisition ; sans oublier la gestion compliquée des flux vidéo et la correspondance planning/timing de l'acquisition qui comportera dans ce cas plusieurs paires d'images.

La théorie de l'interpolation consiste à trouver une fonction dont la courbe passe par tous les points d'une liste déterminée à l'avance appelés points de contrôle. Il existe plusieurs modèles qui ont été proposés dans la littérature pour tel que : les splines, l'interpolation polynômiale de Lagrange, l'interpolation d'Hermite et d'autres qui ont été employés dans différents domaines d'interpolation de données. Dans le cas des splines, on considérera une série de $(n + 1)$ points dont les coordonnées sont (x_i, y_i) . On cherchera alors une fonction

(un polynôme) pour chaque intervalle $[x_i, x_{i+1}]$ reliant les points i et $i + 1$: il faut ainsi déterminer n fonctions, afin de former la fonction d'interpolation. Il existe plusieurs types de splines : les splines linéaires qui sont simplement les droites reliant chacun des points d'arrêt, les splines quadratiques qui sont des polynômes de degré 2, et les splines cubiques que nous développerons plus profondément dans la suite de la section. L'intérêt des splines quadratiques et cubiques par rapport aux splines linéaires est d'obtenir une fonction d'interpolation lisse : elle est continûment dérivable.

Plus particulièrement, les splines cubiques font intervenir, pour chaque intervalle, un polynôme du troisième degré, la courbe d'interpolation est donc elle aussi lisse (dérivée première continue) et la dérivée seconde est également continue. Les courbes splines cubiques simulent le comportement d'une latte de bois, à laquelle est imprimée une certaine forme à l'aide de piquets (les points d'arrêt). La déformation de la latte entre deux piquets est représentée par une fonction dont la quatrième dérivée s'annule, c'est à dire par un polynôme du troisième degré. Nous avons donc n polynômes cubiques $(f_i)_{1 \leq i \leq n}$ que nous allons écrire sous la forme (3.5.6) :

$$f_i''(x_i) = f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i, \quad x_{i-1} \leq x \leq x_i, \quad i \in \{1, 2, \dots, n\} \quad (3.5.6)$$

Nous avons défini les splines cubiques comme ayant leur seconde dérivée continue ; pour déterminer les coefficients nous allons donc utiliser cette condition de continuité, qui s'écrit ainsi avec la notation adoptée (3.5.7) :

$$f_i''(x_i) = f_i'', \quad f_i''(x_{i-1}) = f_{i-1}'', \quad i \in \{1, 2, \dots, n\} \quad (3.5.7)$$

Le polynôme recherché étant du troisième ordre, sa dérivée seconde est linéaire, elle doit prendre les valeurs f_1', f_2', \dots, f_n' .

Ces modèles d'interpolation sont très désirés dans le domaine de la modélisation graphique. En effet, ils s'adaptent aux objets lisses par leurs courbes elles même lisses et continues sur les extrémités des intervalles d'actions. Nous pouvons espérer une bonne interpolation si les points de contrôle ne sont pas très espacés dans l'espace.

Dans notre cadre d'application, nous introduisons ces fonctions d'interpolations en vue d'enrichir la résolution du modèle reconstruit dans une seule direction. Celle où les primitives sont moins denses, c'est-à-dire la direction des abscisses. La figure 3.10 illustre la richesse de primitives à reconstruire sur l'axe Y et leurs présence moins dense sur l'axe X de l'image. La même figure montre aussi le rendu après interpolation en maillage et en surface 3D. Les résultats valident l'emploies de ces fonctions d'interpolation et par la suite la substitution d'un déplacement du modèle lumineux sur la direction X qui est inadapté dans le cas du visage. Néanmoins, pour garantir une interpolation fidèle du modèle numérique au modèle d'origine il est indispensable de partir d'un ensemble significatif de points d'appuis (ou de contrôle). En conséquent, les franges projetées sur le visage doivent être bien serré dans le but d'avoir le maximum de points 3D issus de la triangulation optique.

A ce stade de traitement, un nuage de points est obtenu en deux étapes. La première recherche d'intersections de rayons optiques issues des points correspondants alors que la

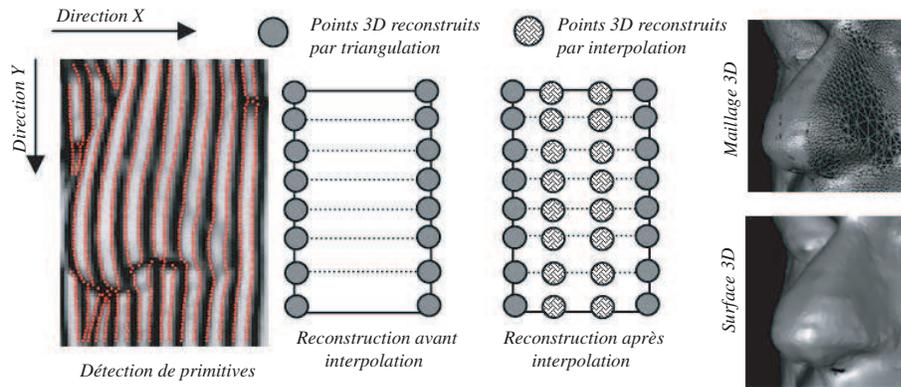


FIG. 3.10 – Interpolation par Splines cubiques pour densifier la reconstruction.

deuxième interpolate entre ces points pour enrichir la résolution du modèle tout en restant fidèle au modèle initial grossier. Pour produire une surface représentative du visage (partielle ou complète), nous faisons appel à des techniques de maillage 3D et en particulier la dualité : diagramme de Voronoï / triangulation de Delaunay qui fera l'objet de la section suivante.

3.5.3 Génération de la forme 3D

Cette partie concerne la génération de forme 3D à partir d'un nuage de points 3D. En effet, les algorithmes de reconstruction permettent d'obtenir un ensemble de points auxquels il faut associer une forme plus réaliste. Néanmoins, à partir d'un nuage de points quelconque, il est possible de définir un grand nombre de maillages différents. Nous avons choisi le maillage par triangulation entre les nœuds (points 3D) pour répondre à cette question. C'est la méthode de maillage la plus connue et la plus traitée en littérature de modélisation d'objets 3D.

L'objectif étant d'approximer l'altitude de nœuds qui se situent dans les triangles, il est souhaitable de choisir une triangulation qui minimise la longueur des facettes et qui par conséquent garanti une meilleure répartition des angles dans un triangle donné. Autrement dit, l'interpolation sera optimale si les formes des facettes se rapprochent du triangle équilatéral. Ce critère est vérifié par la triangulation de Delaunay, c'est pourquoi les algorithmes utilisés sont principalement basés sur cette technique.

La triangulation de Delaunay repose sur la partition de l'espace par les diagrammes de Voronoï qui consistent en un découpage de cet espace en régions définies autour des points. C'est l'étape qui permet de réaliser facilement par la suite la triangulation. Ensuite, le calcul de l'interpolation linéaire est simple. Pour chaque nœud de la grille, il suffit de trouver le triangle concerné. L'altitude du nœud est calculée à l'aide de l'équation du plan portant la facette triangulaire. Dans ce qui suit nous donnons un recul mathématique sur la dualité

diagramme de Voronoï/triangulation de Delaunay ainsi que le résultat obtenu avant et après application sur les points 3D du visage reconstruit.

1. *Diagramme de Voronoï* : Le diagramme de Voronoï est un outil pour positionner un ensemble de points les uns par rapport aux autres. Ce diagramme permet notamment de connaître les voisins d'un point et de quantifier la proximité de ce point par rapport à chacun de ses voisins.

Soit S un nuage de points $(s_i)_{i=1,\dots,n}$, en n dimensions, la construction de la triangulation de Delaunay de l'enveloppe convexe de ce nuage peut s'établir en utilisant le fait que celle-ci est le dual du diagramme de Voronoï construit sur S .

Le diagramme de Voronoï est l'ensemble des cellules ou sites, V_i , définis par (3.5.8) :

$$V_i = \{S \text{ telque } d(S, s_i) \leq d(S, s_k), \quad \forall k \neq i\} \quad (3.5.8)$$

Où $d(,)$ est la distance entre deux points, distance induite par la métrique euclidienne. Une cellule, V_i , est donc le lieu des points plus proches de s_i que de tout autre point de S . Les cellules V_i sont des polygones fermés (en 2D) convexes, et recouvrent l'espace sans chevauchement. Une formulation mathématique des diagrammes de Voronoï de points en métrique euclidienne est présentée sur l'expression ensembliste précédente.

Pour construire le diagramme de Voronoï, une possibilité serait de tracer les plans médiateurs des couples de points voisins ; puis découper ces plans en fonction de leurs intersections, en supprimant les parties qui ne sont pas pertinentes (figure 3.11 (a)).

2. *Delaunay dual de Voronoï* : Par définition, chaque cellule V_i est non vide et est associée à un point s_i de S . A partir de ces V_i on peut construire le dual qui est la triangulation de Delaunay. En deux dimensions, les côtés des cellules sont à mi-chemin des deux points qu'ils séparent et par la suite ne sont autres que les médiatrices des arêtes de la triangulation. Autrement dit, si on joint les sommets s_i appartenant à deux cellules adjacentes, on obtient cette triangulation. En deux dimensions (ou trois), s'il n'existe pas de points cocycliques (cosphériques) alors le dual des cellules de Voronoï associées au nuage de points donné, est la triangulation de Delaunay de l'enveloppe convexe de ce nuage (figure 3.11 (b)).

Le graphe de Delaunay est un graphe planaire noté $T = (S, A)$ dual du graphe de Voronoï défini précédemment, dont les arêtes A sont définies par (3.5.9) :

$$A = (s_i, s_j), \quad s_i, s_j \in S, \quad j \neq i \quad (3.5.9)$$

Les points s_i et s_j sont voisins au sens de Voronoï, c'est-à-dire que les deux cellules qu'ils définissent sont adjacentes.

Quatre propriétés de la triangulation de Delaunay sont importantes :

- Le disque ouvert passant par les trois sommets d'un triangle ne contient aucun autre point de S .
- La triangulation de Delaunay maximise le minimum des angles des triangles.
- L'insertion d'un nouveau point dans S n'entraîne qu'une modification locale du graphe.

Un schéma de triangulation de Delaunay est donc précédé par une partition de Voronoï qui définit les voisins de tous les points. Cette dernière est la subdivision du plan permettant de trouver le point de S le plus proche d'un point de requête q . Le maillage est donc obtenu en reliant les points de S dont les cellules de Voronoï sont voisines. On obtient ainsi une triangulation de S , c'est-à-dire une subdivision de l'enveloppe convexe de S en triangles ayant les points de S pour sommets.

Les sommets du diagramme de Voronoï sont les centres des cercles circonscrits des triangles de la triangulation de Delaunay. Deux sommets sont reliés par une arête s'ils sont les centres circonscrits de deux triangles voisins. Ainsi, il est possible d'obtenir le diagramme de Voronoï à partir de la triangulation de Delaunay. Ceci est particulièrement pratique car il est facile d'insérer un nouveau point dans une triangulation de Delaunay et donc dans le diagramme de Voronoï.

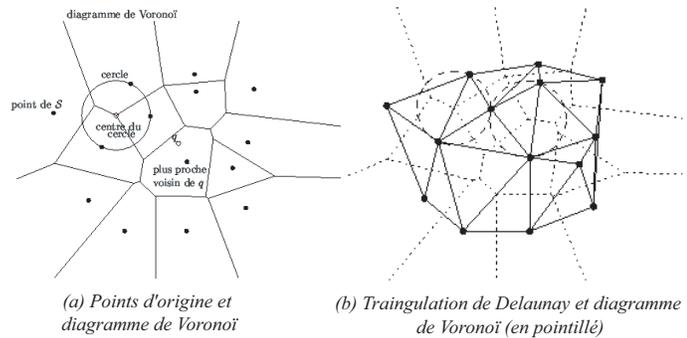


FIG. 3.11 – Dualité : Diagramme de Voronoï/Triangulation de Delaunay.

Dans notre cas, nous appliquons cette technique de triangulation sur les projections des points 3D sur le plan (X, Y) ce qui donne naissance à une grille de points 2D. Nous considérons ainsi le plan 2D formé par ces projections pour appliquer l'algorithme incrémental de triangulation. La décomposition de voronoï est appliquée à ces points du plan par contre la triangulation est générée sur les points 3D en considérant les points 3D du voisinage d'un point donné. La figure 3.12 illustre les résultats d'interpolation et de maillage sur un modèle partiel de visage. Les zooms sur les images montrent la qualité du maillage obtenu est sa cohérence. Ceci est due à ce principe de triangulation qui ne produit pas de larges facettes et se base sur une notion de voisinage définie par la partition de Voronoï.

3.6 Modélisation complète du visage

Une paire d'images stéréo de la scène permet de reconstruire une seule partie du visage. Il s'agit de la partie du visage où le motif lumineux se projette avec un fort angle d'incidence. C'est dans cette zone que les lignes peuvent être détectés aisément sur le visage sans trop de

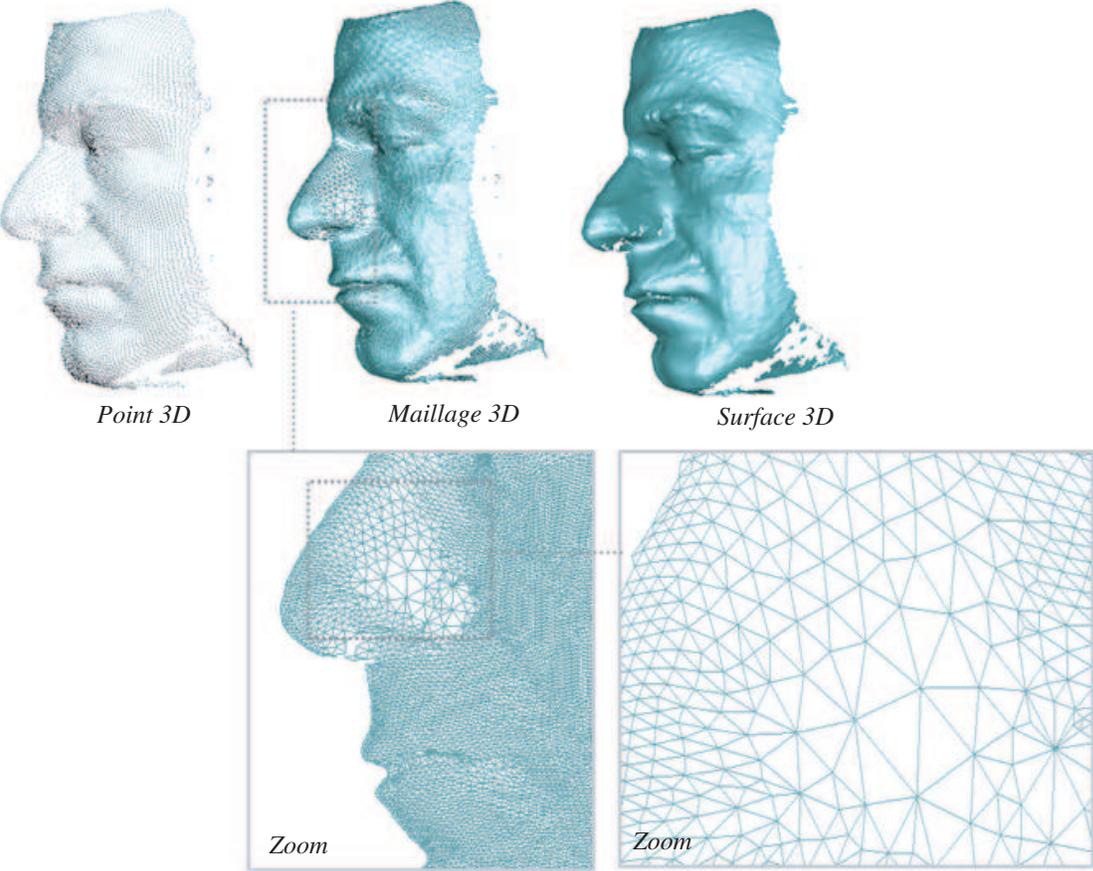


FIG. 3.12 – Formation du modèle 2.5D par interpolation puis par maillage des points.

problèmes occultations ou d'étalement des franges de lumières.

En général une seule de ces parties ne suffira pas pour obtenir un modèle de visage que l'on qualifiera de complet, même si l'acquisition est frontale. En effet, rien que pour le nez il est impossible de voir, sous un angle d'incidence suffisante, les deux flancs gauche et droite. Le visage complet qu'on souhaite obtenir pour l'incorporer dans la base de données de visages à explorer pour vérifier ou reconnaître l'identité d'un client.

De ce fait nous avons opté, dans un premier temps, à la reconstruction partielle de plusieurs parties du visage puis à les associer par des algorithmes spéciaux de recalage en vue d'obtenir un visage 3D plus complet. L'application des étapes de modélisation partielles décrites dans les sections précédentes, sur un ensemble de photos prises dans différentes orientations, produit des modèles partiels appelés également aussi modèles 2 1/2D. Il s'agit dans cette étape de fusionner ces modèles dans le but de construire un modèle 3D complet. Ce problème se décompose en deux parties. D'abord il s'agit de positionner les modèles partiels obtenus les uns par rapport aux autres (recalage grossier). En effet ces modèles sont construits indépendamment à partir de vues indépendantes du sujet (le système d'acquisition est fixe et c'est au sujet de tourner sa tête). Après il faut un recalage plus fin qui permettra l'association plus exacte des modèles. Les recalages grossier et fin entre les modèles sont faits deux à deux.

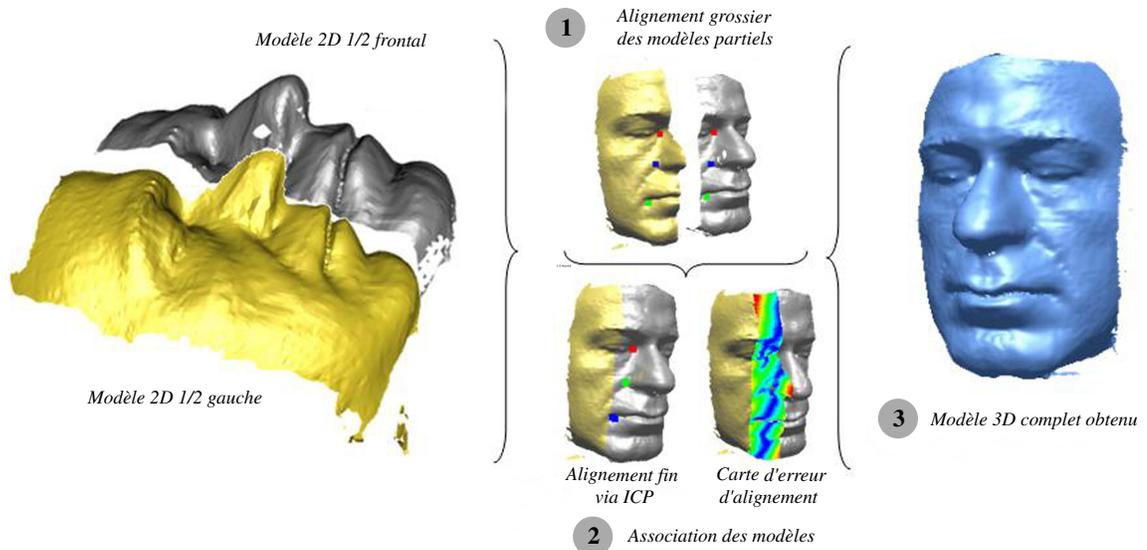


FIG. 3.13 – Recalage et fusion de modèles partiels en vue de produire un modèle 3D complet.

3.6.1 Recalage grossier des différentes vues

La méthode de recalage grossier des modèles partiels consiste à mettre en correspondance des points remarquables déterminés manuellement sur chacun des modèles. Ces primitives sélectionnées sur les parties communes des modèles partiels peuvent être les coins intérieurs des yeux, les coins du nez ou les coins de la bouche. Un minimum de trois paires de points appariées permet, d'abord, de calculer la transformation rigide entre les modèles puis de l'appliquer sur le modèle approprié pour minimiser la distance entre eux. En fin, les modèles sont superposés au niveau de leurs parties communes. Cette étape présente une bonne initialisation à une étape de recalage plus fin qui introduit beaucoup plus de points communs entre les deux surfaces. L'algorithme d'alignement fin est décrit dans la sous-section suivante.

La figure 3.13 illustre la procédure d'alignement grossier entre les différents modèles partiels reconstruits. Le modèle frontal est considéré comme modèle de référence. Les points de même couleurs ont été choisis comme paires de primitives appariées. Le but de cet alignement grossier est de rapprocher la solution à une solution finale où les distances entre les surfaces sont minimales. C'est l'objet de la sous-section suivante qui présente un algorithme itératif qui minimise ces distances jusqu'à converger à une solution finale : c'est l'algorithme *ICP* (Iterative Closest Point) ou le point le plus proche itéré.

3.6.2 Recalage fin moyennant *ICP*

Le recalage grossier permet donc de rapprocher les modèles 2 1/2D obtenus par modélisation partielle et même de les superposer au niveau de leurs parties communes moyennant l'appariement manuel. Il est souhaitable maintenant de raffiner ce recalage et ce en minimisant les distances entre les parties du visage. Pour l'assurer, l'algorithme très connu du plus proche voisin itéré *ICP* (Iterative Closest Point) introduit par Besel en 1992 dans [BM92] a été employé. *ICP* permet à partir des deux ensembles de points $P = \{p_i\}$ et $X = \{y_i\}$, de retrouver la transformation rigide (R, t) qui minimise la distance entre eux. Le principe consiste à déterminer pour chaque point p_i de l'ensemble P , le point le plus proche dans le second ensemble X , au sens d'une distance euclidienne. A partir de ces appariements, une transformation rigide minimisant un critère des moindres carrées (3.6.1) est calculée et appliquée à l'ensemble P :

$$e(R, t) = \frac{1}{N} \sum_{i=0}^N \|(Rp_i + t) - y_i\|^2 \quad (3.6.1)$$

Ce processus est alterné et itéré jusqu'à convergence (stabilité de l'erreur minimale). Par ailleurs, La transformation globale (R, t) est mise à jour de façon incrémentale comme suit : pour chaque itération k de l'algorithme, $R = R_k.R$ et $t = t + t_k$. Le critère à minimiser à l'itération k devient alors (3.6.2) :

$$e(R_k, t_k) = \frac{1}{N} \sum_{i=0}^N \|(R_k(Rp_i + t) + t_k - y_i)\|^2 \quad (3.6.2)$$

Pour plus d'éclaircissements nous fournirons ci-dessous une description de l'algorithme *ICP* et de ses étapes :

- a) On initialise l'algorithme pour $k = 0$ et on pose $S^0 = X = \{y_i\}$
- b) A l'itération k , on cherche pour chaque point p_i^k de P^k , le point de la première entité géométrique qui lui est le plus proche.
- c) La meilleure transformation rigide qui recalage les deux ensembles appariés est estimée à l'aide de l'une des méthodes d'optimisation de transformation rigide.
- d) L'ensemble des points S^{k+1} est alors défini comme étant le transformé de l'ensemble S^k par la transformation rigide estimée à l'itération k .
- e) Si le recalage effectué en d) satisfait un critère de convergence, ce procédé est stoppé, sinon on réitère le même procédé en incrémentant k ($k = k + 1$) et en revenant à b).

Dans notre implémentation de cet algorithme l'estimation de transformation rigide est faite moyennant une décomposition en valeurs singulières qui permet de fournir à chaque itération et au final la transformation (R, t) optimale.

Un pipeline de ces étapes effectué sur trois modèles partiels est illustré par la figure 3.13. L'alignement se fait deux à deux, en passant par le modèle de face (frontal) comme modèle de référence. Des cartes d'erreurs sont également présentées. Pour les modèles profil gauche/modèle frontal, l'erreur moyenne est de $0,49 \text{ mm}$ et l'écart type est égales à $0,31 \text{ mm}$, alors que le recalage des modèles frontal/profil droit, présente une erreur moyenne de $0,43 \text{ mm}$ et un écart type de $0,28 \text{ mm}$. Une fois les surfaces recalées une couture des différents maillages est faite pour l'obtention d'un modèle complet de visage : c'est l'étape de fusion. Le résultat de cette étape est un modèle VRML (*Virtual Reality Modelling Language*) comportant à la fois le maillage 3D et la texture.

3.7 Résultats et procédure de validation

Le but de cette première partie du travail est de proposer une approche de reconstruction partielle et complète 3D dédiée aux visages humain. Cette méthode est destinée à la reconnaissance de visages dans un schéma particulier de matching 2.5D vs. 3D. La reconstruction 3D complète de la surface faciale passe par des reconstructions partielles. En effet, le capteur stéréo utilisé ne permet pas d'acquérir un modèle complet sous une seule angle de prise de vue. Ceci est du au phénomène d'occlusion à l'issu duquel des parties du visages sont cachées par d'autres. Nous souhaitons obtenir des reconstructions (2.5D et 3D) précises et de bonne résolution dans le but de reproduire les détails du visage sur le modèle 3D reconstruit. En réalité, notre défi est d'avoir des acquisitions comparables aux acquisitions faites par des produits industriels existants (basée sur la technologie laser), en terme de précision, densité et de praticabilité. Néanmoins, elle doit être de faible coût

pour son utilisation dans des applications biométriques de reconnaissance/authentification de personnes par le visage.

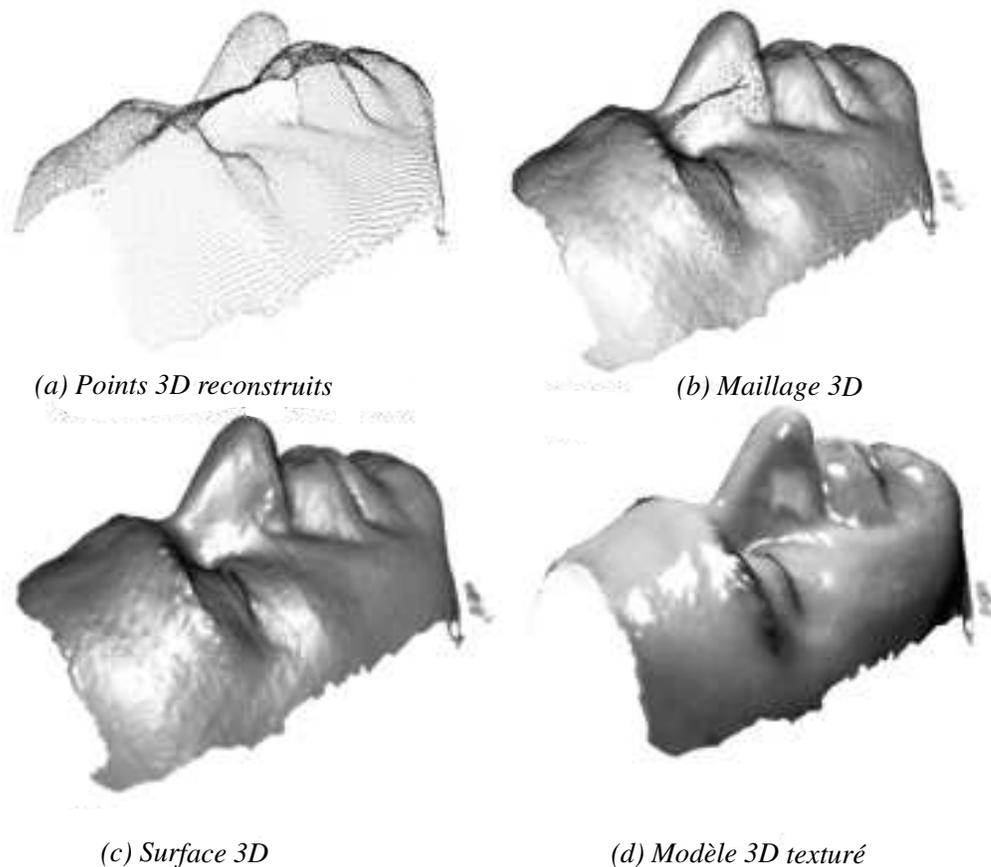


FIG. 3.14 – Résultats finaux de la reconstruction 3D.

La figure 3.14 donne un exemple de visage 3D produit par notre approche de bout en bout. C'est le résultat de l'association de trois modèles partiels un frontal et deux de profils gauche et droite. Le modèle obtenu montre la préservation des détails et la densité du maillage grâce à la combinaison de la vision active et de la modélisation géométrique. Nous avons évité le lissage des détails par la reconstruction active dense sur l'axe des Y avec une précision sous-pixélique. cette première reconstruction grossière permet d'appliquer les courbes d'interpolation sans aucun effet de lissage sur le deuxième axe de reconstruction, X . Dans la figure 3.15, nous donnons un aperçu sur des parties de la forme tridimensionnelle produite et des parties correspondantes dans l'image de texture. Dans ce paragraphe nous étudions, outre ces évaluations qualitatives, des évaluation quantitatives. En effet, nous



FIG. 3.15 – Qualité de la reconstruction 3D : parties forme 3D et texture correspondantes.

avons utilisé un scanner laser 3D (VI300 de Minolta⁴) pour produire des modèles servant de vérité terrain (ou données de référence). La procédure d'évaluation consiste à numériser un visage 3D par notre technique (donnée mesurée) et de numériser le même visage par l'intermédiaire du scanner. Puis, d'apparier les deux surfaces 3D pour calculer la déviation spatiale entre elles. L'appariement de deux formes 3D est fait par le biais de l'algorithme *ICP* (*Iterative Closest Point*), après une étape d'initialisation manuelle (alignement grossier).

La figure 3.16 montre à la fois les deux modèles à comparer (données de référence vis-a-vis données mesurées) et des cartes de déviations signée et absolue, entre eux. Dans la carte de déviation absolu, les couleurs correspondent aux distances mesurées. Les couleurs froides (vers le bleu) représentent les petites distances alors que les couleurs chaudes (vers le rouge) représente les grandes distances. L'erreur moyenne représente 1.65% ($E_{moy} - Z_{min}/Z_{max} - Z_{min}$) de la profondeur du visage calculée par notre approche. Dans la carte de déviation signée, les couleurs correspondent aux déviations spatiales positives et négatives. Les couleurs fortement contrastées représentent les fortes déviations alors que les couleurs faiblement contrastées correspondent aux faibles déviations. La courbe de distribution a une silhouette gaussienne centrée en la valeur nulle de l'ensemble des déviation. La figure 3.16 présente aussi un tableau résumant ces différentes mesures par des valeurs statistiques dont les moyennes et les écarts types.

Déviations/Type	Erreur moyenne (mm)	Ecart type (mm)
Signée	-0.086	2.11
Absolue	2.66(1.65%)	1.62

Les résultats présentés montrent l'efficacité de notre approche pour une reconstruction dense, précise et qui reproduit la structure tridimensionnelle du visage. Elle préserve, également, les détails du visage (les yeux, la bouche, le nez, etc.) sur le modèle numérique.

⁴<http://www.konicaminolta.com/>

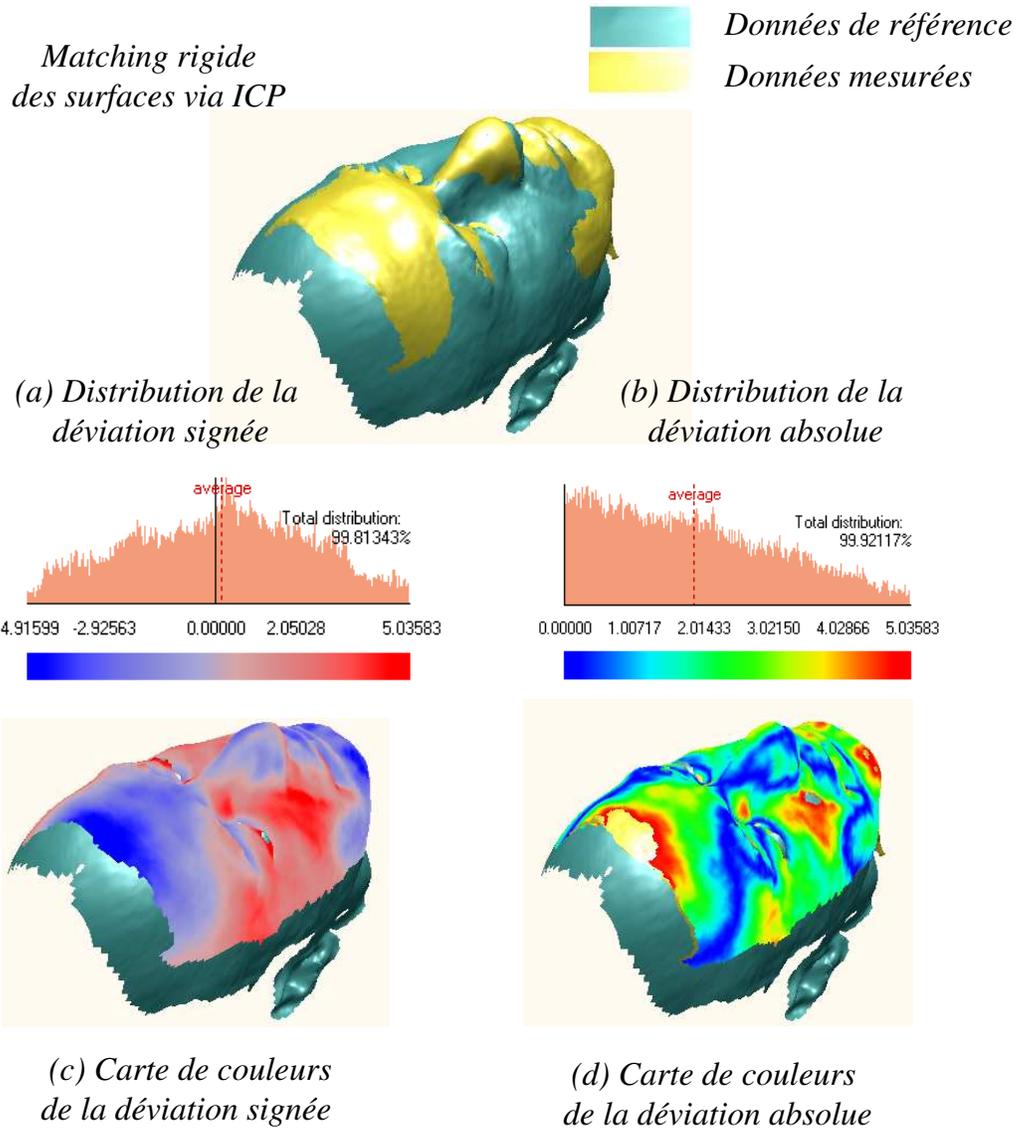


FIG. 3.16 – Déviation spatiale entre les données de références (laser) et les données mesurées (notre capteur).

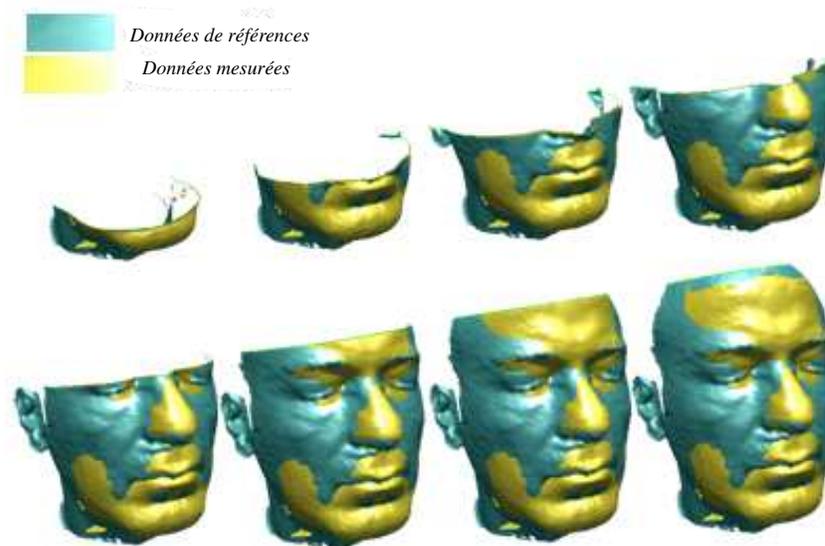


FIG. 3.17 – Différentes sections des données de références (laser) et les données mesurées (notre capteur).

La figure 3.17 montre différentes sections effectuées sur les deux modèles reconstruits par le biais de notre approche et l’approche laser respectivement. Le profil vert est extrait du modèle laser alors que le profil rouge est celui du modèle mesuré par notre approche. Cette figure confirme la ressemblance des deux modèles produits et illustre les petites déviations qui existe entre les deux.

Dans cette procédure d’évaluation des mesures de notre approche d’acquisition 3D appliquée aux visages humains nous devons prendre en considérations certaines considérations. Le but de cette procédure est de monter le mappage entre un modèle produit par un scanner laser et un modèle reconstruit par notre approche. Néanmoins, le scanner dispose lui aussi d’une imprécision qui peut aller jusqu’à 0.75 millimètres. En outre, l’appariement fait par l’algorithme *ICP* essaye de minimiser l’erreur globale entre les deux surfaces et ce engendre des erreurs intrinsèques à l’algorithme qui s’accumulent aux déviations spatiales mesurées. Ceci se manifeste par la non stabilité de ces mesures dans différentes mesures effectuées sur les mêmes données. Une petite différence se présente entre les différents histogrammes calculés dans ces différentes mesures. Sans oublier que le visage peut présenter aussi des petites déformations entre les deux prises de vue par le scanner et l’approche proposée.

3.8 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode de modélisation tridimensionnelle de visage en vue de la reconnaissance. Cette approche est basée sur le mariage entre

la vision active et la modélisation par Splines cubiques, introduite dans le but d'améliorer la qualité des modèles reconstruit avec une plate-forme d'acquisition à faible coût. C'est la principale contribution de cette première partie, d'abord une reconstruction grossière mais précise est obtenu moyennant notre capteur actif. Puis, une reconstruction plus dense est obtenue via l'introduction de la méthode d'interpolation, ses points d'appuis sont obtenu par la première étape de reconstruction grossière. Les résultats obtenus ainsi que la procédure de validation de l'approche montrent bien la préservation de la structure géométrique du visage ainsi que ses détails en comparaison avec un scanner laser 3D. Dans le chapitre suivant, nous traitons le deuxième module de reconnaissance basé sur l'alignement rigide de surfaces tridimensionnelles.

Chapitre 4

Reconnaissance/Authentication de personnes via leurs surfaces faciales

4.1 Introduction

Depuis quelques années, le problème de reconnaissance faciale connaît un gain d'intérêt exponentiel à cause de son caractère peu intrusif par rapport aux autres modalités biométrique. Dans la littérature, plusieurs techniques ont été proposées, néanmoins, les approches sur lesquelles se basent souffrent de sérieuses limitations qui sont dues principalement à l'utilisation d'un modèle 2D du visage (image couleur ou d'intensité). Cette représentation faciale est très sensible aux variations de pose et d'illumination ; sans oublier les variations intrinsèques du visage à savoir les expressions faciales. Comme illustre la figure 4.1, la variabilité des données, dans ce problème de reconnaissance de formes, peuvent être catégorisée en deux catégories : (A) les variations intra-classes et (B) les similarités inter-classes. La première catégorie des problèmes concerne la variabilité au sein d'une même classe d'individu causée par les variations dans les conditions d'éclairage du visage, son orientation, les expressions faciales, le port d'accessoires (lunettes, cache-soleil,...), etc. Ces variations changent considérablement l'apparence de la personne et les caractéristiques utilisées pour la reconnaissance sont, par conséquent, influencées. La seconde catégorie de problèmes présente la similarité d'apparences (ou de caractéristiques) entre deux individus, de même race par exemple. On trouve, également, l'influence des liens familiaux dans les classes des jumeaux ou des jumelles et la ressemblance entre père et fils ou mère et filles.

Les travaux réalisés dans ce domaine ont visé, essentiellement, le développement d'algorithmes de traitement, mathématiquement de plus en plus performants. En effet, la plupart des travaux se sont intéressés aux techniques d'analyse de données, de réduction de dimensionnalité des données ainsi que leurs décorrélation et aussi aux techniques d'apprentissage. Les taux de reconnaissance, qui restent toujours peu satisfaisants, ont fait que d'autres formes

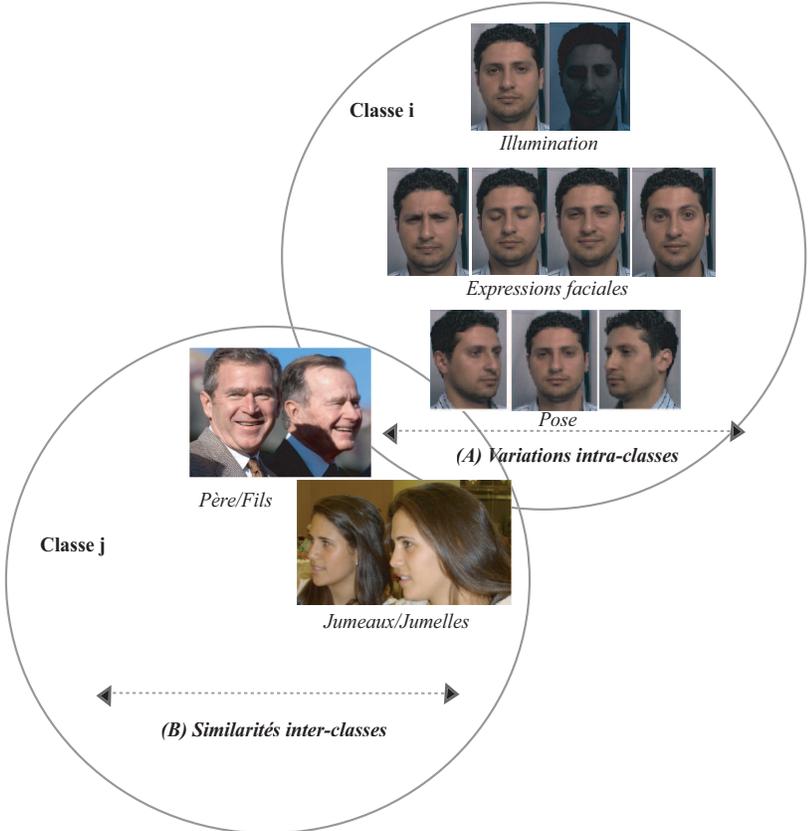
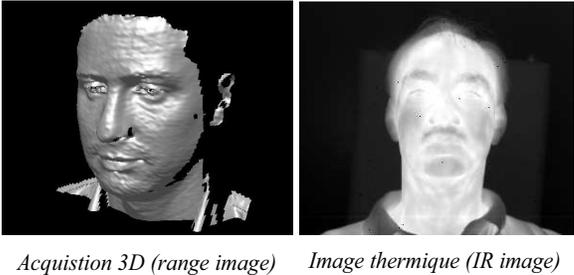


FIG. 4.1 – Problèmes de reconnaissance de visages : (A) variations intra-classes (B) similarité inter-classes.



Acquisition 3D (range image) Image thermique (IR image)

FIG. 4.2 – Nouvelles formes d’acquisition de visage en vue de la reconnaissance.

d'acquisition, moins influencées par ces variations, ont été employé très récemment. Ces nouvelles formes d'acquisition et de représentation du visage se présentent dans les images thermiques prises par des caméras infrarouges (*IR image*), les acquisition 3D (*range image*) et les images couleurs de hautes résolutions (voir figure 4.2). Ces formes de représentation semblent être plus invariantes, chacune d'entre elles, par rapport à quelques variations.

Dans le présent travail, nous proposons d'exploiter une représentation tridimensionnelle du visage. Cette représentation est plus riche par rapport à une représentation texturale et peut apporter des éléments de réponses à ces challenges classiques. En effet, un algorithme basé sur la géométrie (forme) 3D du visage est insensible aux variations d'éclairages du fait que ce dernier fait abstraction de l'information de texture. En outre, par introduction de cette troisième dimension, la reconnaissance peut se faire sous n'importe quelle angle de vue, d'où l'invariance au problème de changement de posture. Cette richesse d'information et d'invariance des représentations 3D par rapport à celles 2D nous ont conduit à proposer un paradigme de reconnaissance/authentification en se basant sur cette nouvelle modalité, à savoir la face 3D.

Dans ce chapitre, nous décrivons d'abord une vue d'ensemble du paradigme que nous proposons avec un contexte d'application englobant à la fois les problèmes d'identification et d'authentification de personnes. Ensuite, nous développons une première approche de matching de surfaces faciales 3D moyennant l'algorithme *ICP* (*Iterative Closest Point*) utilisée pour le calcul de score de similarité entre les visages. Puis, nous présentons notre nouvelle variante *R-ICP* et notre nouvelle métrique orientée région conçu en vue de réduire les effets des changements d'expressions faciale. Cette version améliorée est basée sur une étude anatomique du visage en vue de distinguer les régions du visages de fortes déformations par rapport à celles plus statiques. Ici, un schéma de segmentation unifié est élaboré en se basant sur la précédente étude et une procédure de segmentation automatique de la galerie des visages 3D complet est, également, donnée. Finalement, nous introduisons une approche de reconnaissance faciale 3D par le biais de calcul de cartes géodésiques effectué sur les surfaces faciales après leurs matching.

4.2 Vue d'ensemble de l'approche proposée

Dans cette partie de notre travail, un nouveau schéma de reconnaissance est proposé pour plus d'efficacité et d'invariance à certaines variations décrites dans la section précédente. D'abord, une reconstruction complète des visages à reconnaître ou à authentifier est nécessaire en vue d'effectuer le recalage des surfaces faciales quelque soit l'angle de prise de vue de l'image requête. Cette étape fait partie de l'étape off-line du paradigme proposé. Dans la partie on-line, de reconnaissance ou de vérification d'identité, nous cherchons à recalculer le modèle requête (modèle partiel ou 2 1/2D) à tous les modèles complets de la galerie, s'il s'agit d'un scénario d'identification. Dans le cas d'un scénario d'authentification la comparaison est faite par rapport au modèle complet 3D correspondant. Ce recalage permet une estimation précise de la transformation rigide entre les deux modèles. En plus, il permet d'apparier les deux nuages de points 3D et permet par la suite de calculer la distance qui

les sépare. C'est sur cette distance que se base notre métrique de similarité. En effet, plus la distance est élevée moins similaires les visages sont. Cette métrique a subi des améliorations suite à une étude anatomique du visage couplée avec une étude empirique qui nous ont permis de classer les régions du visage selon leurs degrés de déformations et par conséquent leurs sensibilités aux expressions faciales.

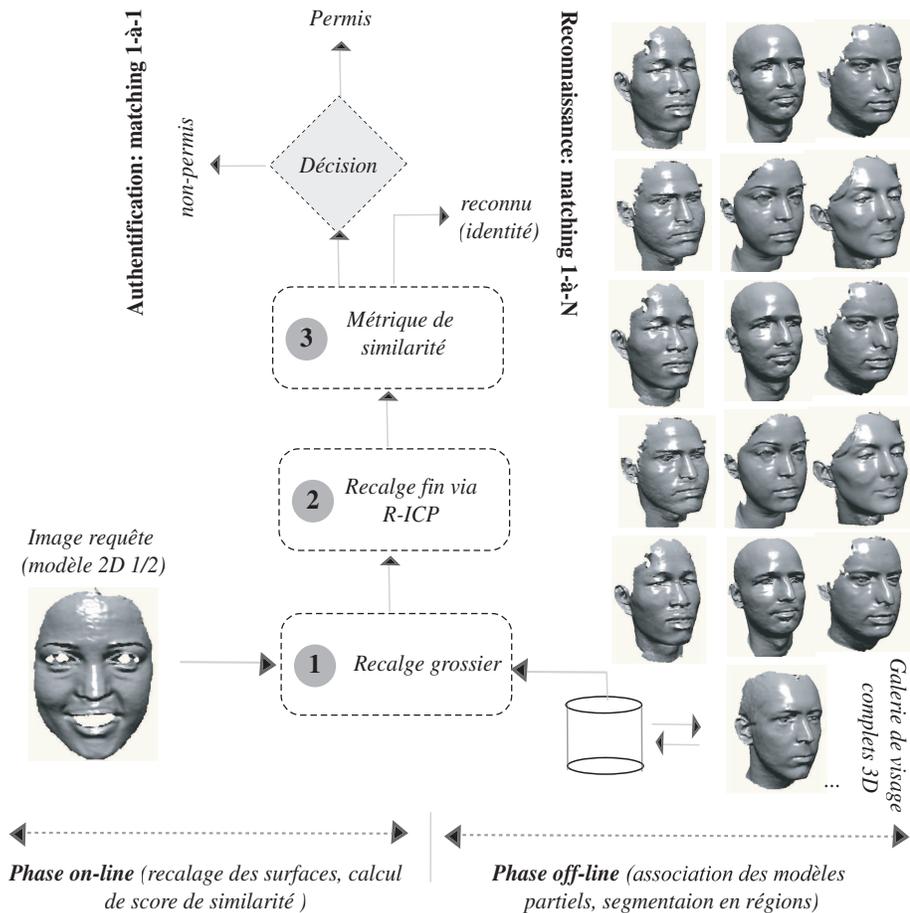


FIG. 4.3 – Vue d'ensemble de l'approche de reconnaissance/authentification proposée.

Plus concrètement, et comme illustré par la figure 4.3, la partie off-line de notre approche consiste, dans une première étape, en la reconstruction des modèles 3D complets qui seront incorporés dans la galerie. Un modèle 3D complet (*Full 3D face*) est obtenu moyennant une procédure de recalage et de fusion de quelques modèles partiels du même visage prises dans différents angles de vue. En pratique, un ensemble de trois modèles partiels suffisent pour restituer la forme faciale complète (région située entre les oreilles, la chevelure et le cou). A

un maillage 3D on associe aussi une carte de texture formée elle aussi à partir des trois textures d'origines fusionnées. Ces modèles 3D sont, ensuite, segmentés automatiquement selon un schéma de segmentation unifié. A Chacune des régions est attribuée un label présenté par une couleur unique et spécifique. Dans la partie on-line de reconnaissance/authentification, une image requête est présentée au système. Elle est recalée aux visages 3D de la base de données en vue de déterminer la transformation rigide entre les deux surfaces et de calculer la déviation spatiale entre elles. Le recalage est fait en deux étapes : un premier recalage grossier suivi d'un recalage plus fin moyennant notre variante de l'algorithme *ICP*, appelé *R-ICP*.

4.3 Appariement de surfaces 3D via *ICP*

Nous souhaitons, dans cette phase, recalcr de manière rigide deux nuages de points correspondants, respectivement, à un modèle requête et un modèle de la galerie de visages 3D. Le recalage consiste à trouver la transformation rigide, entre les deux nuages de point, en vue de garantir un bon matching 2 1/2D vs. 3D. Plusieurs travaux de recherche se sont intéressés à ce sujet pour son application dans, essentiellement, deux domaines clés : l'association de modèles partiels d'un même objet en vue d'obtenir un modèle numérique complet et la reconnaissance de formes 3D (répondre à une requête, naviguer dans une base structurée, etc.). A l'issu de cet intérêt, plusieurs algorithmes ont été proposés pour résoudre ce problème de recalage d'entités tridimensionnelles (nuages de points, surfaces triangulées, surfaces implicites, surfaces paramétriques) depuis le milieu des années 80 [AF86]. Néanmoins, l'immense progrès a eu lieu au début des années 90 avec les publications [BM92] et [CM92], qui décrivent deux techniques itératives et très similaires. En particulier l'algorithme du plus proche voisin itéré *ICP* (*Iterative Closest Point*) [BM92] qui a connu depuis un succès considérable dans ce domaine, où il est utilisé dans la plupart des travaux. Il a surtout connu un grand nombre de variantes. Nous n'en présentons ici que les principes généraux de cet algorithmes qui seront utiles pour comprendre la suite de cette thèse. Le lecteur pourra se reporter à [RL01] pour un état de l'art plus complet, et doublé d'une bonne comparaison expérimentale.

Cette famille d'approches dérivée de l'algorithme *ICP* est basée sur la minimisation, d'une manière itérative, d'une mesure de similarité au sens des moindres carrés entre des paires de points mis en correspondance et appartenant aux deux entités à recalcr. L'algorithme *ICP* (*Iterative Closest Point*) ou encore l'algorithme du plus proche voisin itéré emploie une procédure itérative qui utilise le concept du point le plus proche pour passer d'une itération à la suivante. Il consiste en l'optimisation alternée des appariements et de la transformation rigide. Cette transformation est le plus souvent un déplacement (rotation et translation), mais elle peut être plus complexe, notamment dans le cas d'objets déformables. Nous nous intéressons dans nos travaux uniquement à l'estimation des transformations rigides vu la complexité de modéliser les déformations 3D de la forme d'un visage. Cette question sera posée plus loin dans ce manuscrit. Notre critère de minimisation comporte, par conséquent, deux paramètres de déplacement t et de rotation R .

4.3.1 Présentation de l'algorithme *ICP*

Soit deux ensembles de données P et X les deux nuages de points à recalculer. prenons par exemple le premier ensemble (P) présentant le modèle de la galerie alors que le deuxième ensemble (X) présente un modèle requête. Deux sous-ensembles de ces entités sont mis en correspondances respectivement p_i et y_i via un critère de similarité que nous définirons plus tard. Il s'agit donc de trouver la transformation rigide, noté (R, t) , qui minimise le critère au sens des moindres carrés suivant (4.3.1) :

$$e(R, t) = \frac{1}{N} \sum_{i=0}^N \|(Rp_i + t) - y_i\|^2 = \frac{1}{N} \sum_{i=0}^N \|(Rp_i + t) - \Psi(p_i)\|^2 \quad (4.3.1)$$

Où la fonction Ψ associe chaque point p_i à un point de l'ensemble X et N correspond au nombre des paires de points. A une itération k de la procédure itérative de calcul de la transformation globale est mise à jour d'une façon incrémentale comme suit $R = R_k R$ et $t = t + t_k$.

Finalement, le critère à minimiser devient (4.3.1) :

$$e(R_k, t_k) = \frac{1}{N} \sum_{i=0}^N \|(R_k(Rp_i + t) + t_k - \Psi(p_i))\|^2 \quad (4.3.2)$$

Le fonction de correspondance Ψ est définie, alors, par (4.3.1) :

$$\Psi(p_i) = x \mid \min_{x \in X} d((Rp_i + t), x) \quad (4.3.3)$$

La quasi-totalité des variantes de l'algorithme *ICP* diffère par la manière dont l'appariement de points est effectué ou par la technique de minimisation utilisée. Le principe itératif donné plus haut est toujours le même. En effet, Besel [BM92] utilise la distance euclidienne entre les points pour désigner les correspondances entre les deux nuages à recalculer. Chen [CM92] calcule la distance point-à-surface moyennant l'intersection de la normale à la surface passant par un point de la première surface et la seconde surface. D'autres mesures de similarités ont été proposées reposant sur l'indice couleur [GRB94][JK97] ou l'orientation des normales aux deux surfaces [FAB97][BMR⁺99] ou, parfois, combinés à une des deux premières mesures en vue de garantir un appariement correct et plus pertinent.

L'algorithme *ICP*, dans la peau de ses variantes, est composé principalement de deux étapes. Une première étape de mise en correspondance ou d'appariement qui consiste à associer pour tout point de la première entité géométrique le point de la deuxième entité qui lui est le plus proche. Cette étape nécessite donc la définition d'une procédure pour trouver le point le plus proche à un point dans une entité géométrique. La deuxième étape est l'optimisation, au sens des moindres carrés, de la transformation rigide qui recalage au mieux les deux entités géométriques temporairement appariées. Cette transformation est

appliquée à l'entité à recaler. Les deux étapes sont itérées jusqu'à ce que elles deviennent suffisamment proches. Pour simplifier, revenons sur la formulation mathématique donnée plus haut : $X = y_i$ sur une première entité géométrique 3D dont le type appartient à ceux que nous avons cités plus haut appelé $S = p_i$. L'algorithme peut être alors décrit ainsi :

1. On initialise l'algorithme pour $k = 0$ et on pose $P_0 = P$, $R_0 = I$ et $t = (0, 0, 0)$.
2. A l'itération k , nous cherchons pour tout point y_i^k de X^k , le point de P qui lui est le plus proche selon un critère de distance euclidienne (4.3.4) :

$$d(p, x) = \| p - x \|^2 \quad (4.3.4)$$

Nous obtenons, ainsi, des paires de points (y_i^k, p_i^k) appariés dans les deux nuages P et X exprimé par la fonction de correspondance Ψ (4.3.5) :

$$y_i^k = \Psi(p_i^k) = x \mid \min_{x \in X} d((p_i^k), x) \quad (4.3.5)$$

3. Nous calculons à partir de (y_i^k, p_i^k) la meilleure transformation rigide (R_k, t_k) qui recale les deux ensembles appariés est l'estimons à l'aide de l'une des méthodes d'optimisation de transformation rigide comme suit (4.3.6) :

$$e^k = \min_{R_k, t_k} e(R_k, t_k) \quad (4.3.6)$$

4. L'ensemble des points X^{k+1} est alors défini comme étant le transformé de l'ensemble X^k par la transformation rigide estimée à l'itération k . L'application des paramètres issus de (3.), c'est-à-dire (R_k, t_k) , sur l'ensemble X_k donne naissance à l'ensemble X^{k+1} par (4.3.7).

$$X_i^{k+1} = R_k X_i^k + t_k \quad (4.3.7)$$

5. Si le recalage effectué en (4.) satisfait un critère (dans notre cas le critère le l'erreur absolue), ce procédé est stoppé, sinon on réitère le même procédé en incrémentant k et en revenant en (2.). Si arrêt nous obtenons la transformation rigide finale suivante (5) :

$$\text{si } e_k < \tau \text{ alors } R = R_k \text{ et } t = t_k.$$

Une des techniques pratiques les plus utiliser pour l'estimation des paramètres de transformation rigide est la décomposition en valeurs singulière (*SVD*). Elle est appliqué à chaque itération de l'algorithme *ICP* et l'optimisation de la rotation R est découplée de celle de la translation t de la manière suivante :

- *Recherche de la translation* : La translation optimale est caractérisée par une dérivée nulle du critère à minimiser (4.3.8) :

$$\frac{\partial e}{\partial t} = 0 = 2 \cdot \sum_{i=0}^N (R_k p_i + t_k - y_i) \quad (4.3.8)$$

Ce qui donne (4.3.9) :

$$t = \frac{1}{N} \left(\sum_{i=0}^N y_i - R \sum_{i=0}^N p_i \right) \quad (4.3.9)$$

En posant $\bar{y} = \frac{1}{N} \sum_{i=0}^N y_i$ et $\bar{p} = \frac{1}{N} \sum_{i=0}^N p_i$, ces termes correspondent respectivement aux barycentres des deux ensembles de points X et P , respectivement, on obtient alors l'équation (4.3.10) :

$$t = \bar{y} - R\bar{p} \quad (4.3.10)$$

Une fois calculés, nous pouvons passer les points de la scène et du modèle probe en coordonnées barycentriques en posant $y' = y - \bar{y}$ et $p' = p - \bar{p}$, le critère se simplifie en (4.3.11) et permet de calculer la rotation :

$$e(R, t) = \sum_{i=0}^N \|Rp'_i - y'_i\|^2 \quad (4.3.11)$$

– *Recherche de la rotation* : Le critère simplifié précédent se réécrit alors :

$$e(R, t) = \sum_{i=0}^N \|Rp'_i - y'_i\|^2 \quad (4.3.12)$$

$$= \sum_{i=0}^N \|Rp'_i\|^2 + \|y'_i\|^2 + 2(Rp'_i y'_i) \quad (4.3.13)$$

$$= \sum_{i=0}^N \|p'_i\|^2 + \|y'_i\|^2 - 2Tr \left(\sum_{i=0}^N Rp'_i (y'_i)^t \right) \quad (4.3.14)$$

Comme $\|p'_i\|^2$ et $\|y'_i\|^2$ sont constants, il s'agit de maximiser $E(R) = Tr(RK^{tt})$ où $K^{tt} = \sum_{i=0}^N p'_i (y'_i)^t$ est la matrice de corrélation. Le maximum est alors atteint pour $R = VU^t$, où $U\Lambda V^t = K'$ est une décomposition en valeur singulière de K' et $\Lambda = \text{diag}(1, \dots, 1, \det(U), \det(V))$ et si $\det(R) = +1$. Si $\det(R) = -1$, la matrice R n'est orthogonale. Cela se produit quand un des éléments de la matrice diagonale Λ est nul, ce qui signifie la présence d'un cas dégénéré.

– *Propriétés de l’algorithme ICP :*

Critère d’arrêt : L’itération de l’algorithme s’arrête lorsque la distance entre les deux entités à recalculer est inférieure à un certain seuil. Il est possible d’introduire un facteur d’échelle pour avoir une condition de convergence sans dimension. Ce facteur représente la taille approximative du recouvrement et il est défini par la racine carrée de la trace de la matrice de covariance de l’ensemble des points appariés appartenant à la première entité géométrique.

Convergence : Besel et al. ont démontré dans [BM92] que l’algorithme *ICP* converge toujours vers un minimum du critère (4.3.1). En effet, l’erreur diminue d’une façon monotone, par conséquent l’algorithme converge forcément vers un minimum local. Ce critère n’est cependant pas convexe, et le résultat du recalage dépendra de la cuvette où nous mènera la transformation initiale. Une initialisation de trop mauvaise qualité nous mènera vers un recalage manifestement faux. Les transformations initiales de bonne qualité ne mèneront pas toutes au même recalage, et cette dépendance introduit donc une première source d’imprécision. Nous reprenons, rigoureusement, cette démonstration de convergence vers un minimum local. Cependant la convergence vers un minimum global ne peut être garantie.

A une itération k , la distance quadratique moyenne, d_k entre les ensembles de points $\{p_k\}$ et $\{y_k\}$ est défini par (4.3.15) :

$$e_k = \frac{1}{N} \sum_{i=0}^N \|(R_k p_i^k + t_k) - y_i^k\|^2 = \frac{1}{N} \sum_{i=0}^N \|p_i^{k+1} - y_i^k\|^2 \quad (4.3.15)$$

Par définition, nous avons $e_k \leq d_k$, car si $e_k > d_k$ signifie qu’une transformation produit une erreur quadratique plus petite que celle donnée par la transformation mise à jours (R_k, t_k) ce qui est absurde. On obtient donc la première inégalité suivante (4.3.16) :

$$d_k = \frac{1}{N} \sum_{i=0}^N \|p_i^{k+1} - y_i^k\|^2 \leq \frac{1}{N} \sum_{i=0}^N \|p_i^k - y_i^k\|^2 = e_k, \forall i = 1..N \quad (4.3.16)$$

De plus, à une itération $k + 1$, deux nouveaux ensembles de points p_{k+1} et y_{k+1} plus proches sont obtenus satisfaisant l’équation (4.3.17) :

$$\|p_i^{k+1} - y_i^{k+1}\|^2 \leq \|p_i^{k+1} - y_i^k\|^2 \forall i = 1..N \quad (4.3.17)$$

Simplement parce que y_i^{k+1} est le point le plus proche de p_i^{k+1} par définition, et par conséquent, nous obtenons l’inégalité suivante (4.3.18) :

$$d_{k+1} \leq e_k \quad (4.3.18)$$

Finalement, vue que l’erreur quadratique moyenne ne peut pas être négative. Nous obtenons la relation finale suivante (4.3.19) :

$$0 \leq e_{k+1} \leq d_{k+1} \leq e_k \leq d_k, \forall k \quad (4.3.19)$$

Ce qui prouve la convergence de l'algorithme *ICP*, d'une façon monotone, à un minimum local du critère (4.3.1).

Robustesse (par rapport à la transformation initiale) : On parle de robustesse pour désigner la capacité de l'algorithme à ne pas se diriger vers des résultats complètement faux. Il faut pour cela démarrer dans la cuvette centrée autour de la bonne transformation. La robustesse est donc conditionnée par la qualité de la transformation initiale, mais aussi, par la qualité des données de départ. Notons au passage que les résultats complètement faux sont en général très éloignés de la bonne transformation (puisqu'ils sont à l'extérieur de la cuvette principale). Il est donc primordial que les deux ensembles de points à recalcr soient suffisamment proches pour assurer une bonne convergence.

Bruit interne ou répétabilité : Si nous supposons maintenant que notre algorithme est parfaitement robuste (si nous avons une bonne méthode d'initialisation). Tous nos recalages convergent vers des transformations proche de la transformation recherchée. Ces différents résultats ne seront cependant pas identiques, et on désignera par bruit interne de l'algorithme leur variabilité. Il ne faut pas donc interpréter ce bruit comme la conséquence de l'initialisation, mais comme la conséquence de l'erreur sur l'estimation de la configuration des appariements, et il est donc intrinsèque à l'algorithme.

Incertitude globale : On désigne aussi par incertitude ou bruit global l'imprécision finale de l'algorithme, qui tient compte de toutes les sources d'erreurs possibles.

Nous avons donné dans cette section de présentation de l'algorithme *ICP* un recul mathématique important et indispensable pour la compréhension du fonctionnement de notre approche de recalage 2 1/2D vs. 3D. Revenons, maintenant, à notre cas d'application qui est le matching d'un modèle appartenant à la galerie de visage 3D complet et d'un modèle requête celui de la personne à identifier ou à authentifier son identité. Le but du recalage étant d'estimer la transformation rigide entre les deux modèles puis de l'appliquer sur le modèle requête, afin de calculer un score de correspondance qui servira pour une décision de reconnaissance ou d'authentification. Outre les paramètres de la transformation rigide entre les deux modèles, *ICP* produit également un ensemble de paires de points appariés entre les nuages de points à comparer. Nous nous servirons de cette sortie pour calculer le score de similarité entre les nuages de points 3D. En effet, nous calculons les déplacements qui séparent les points appariés dans les deux surfaces. Nous produisons un histogramme représentant la densité de ces mesures. C'est sur la comparaison des histogrammes que le score de similarité entre visages est fondée.

Les figures 4.4, 4.5 et 4.6 illustrent avec trois cas différents les résultats de l'application de

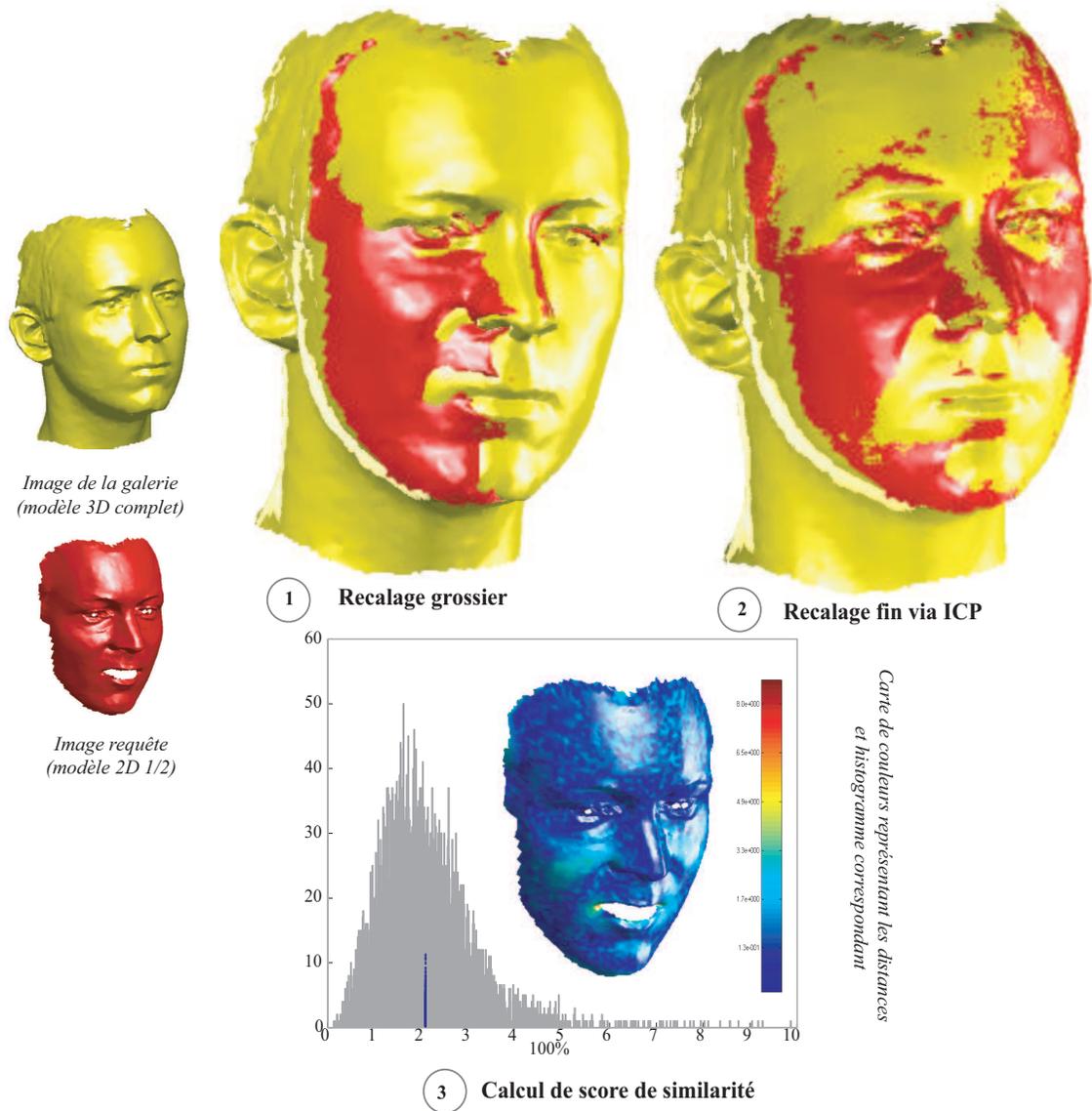


FIG. 4.4 – Exemple1 : recalage et calcul de similarité entre un modèle 3D et un modèle requête de la même personne avec des variations d'expressions faciales.

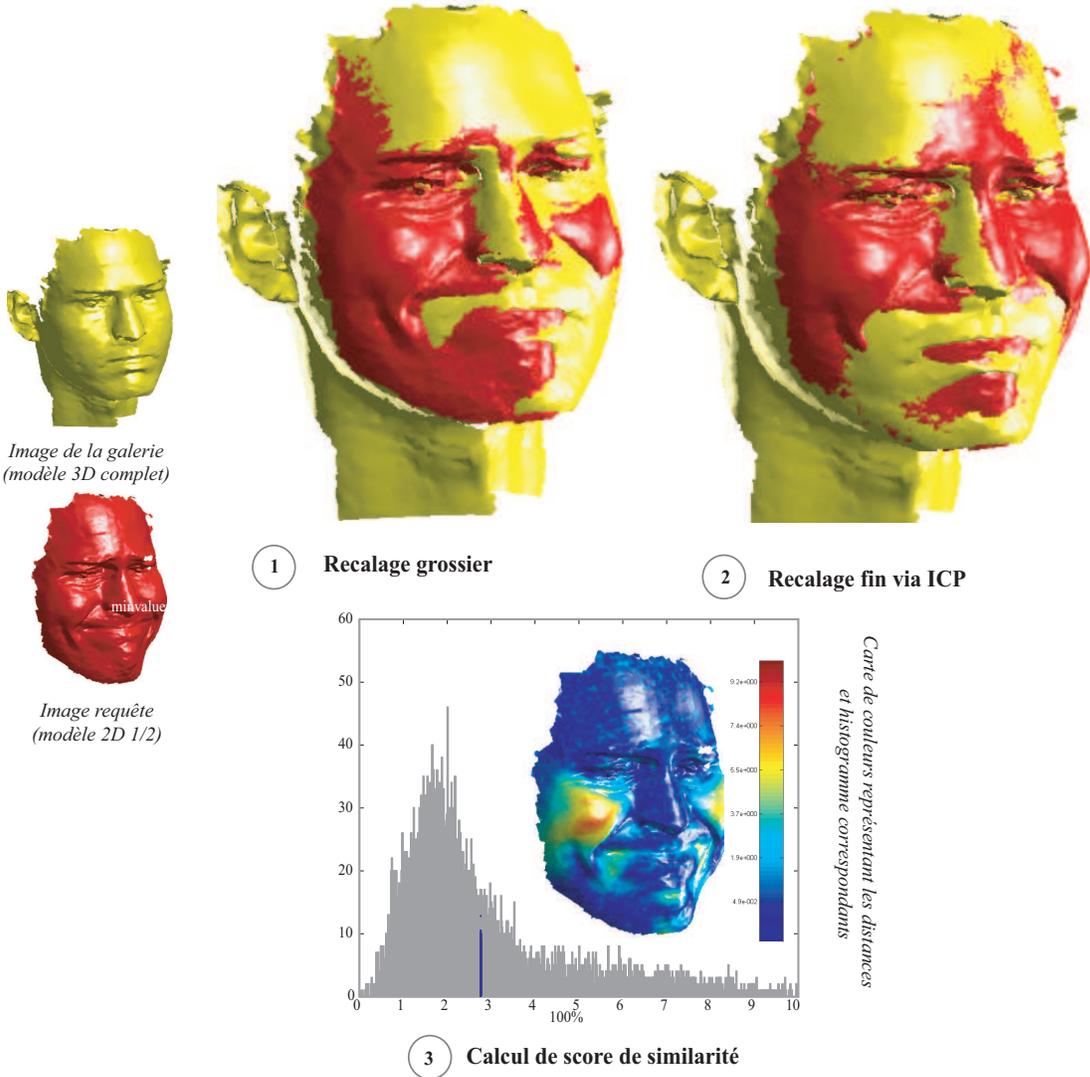


FIG. 4.5 – Exemple2 : recalage et calcul de similarité entre un modèle 3D et un modèle requête de la même personne avec des variations d’expressions faciales plus significatives.

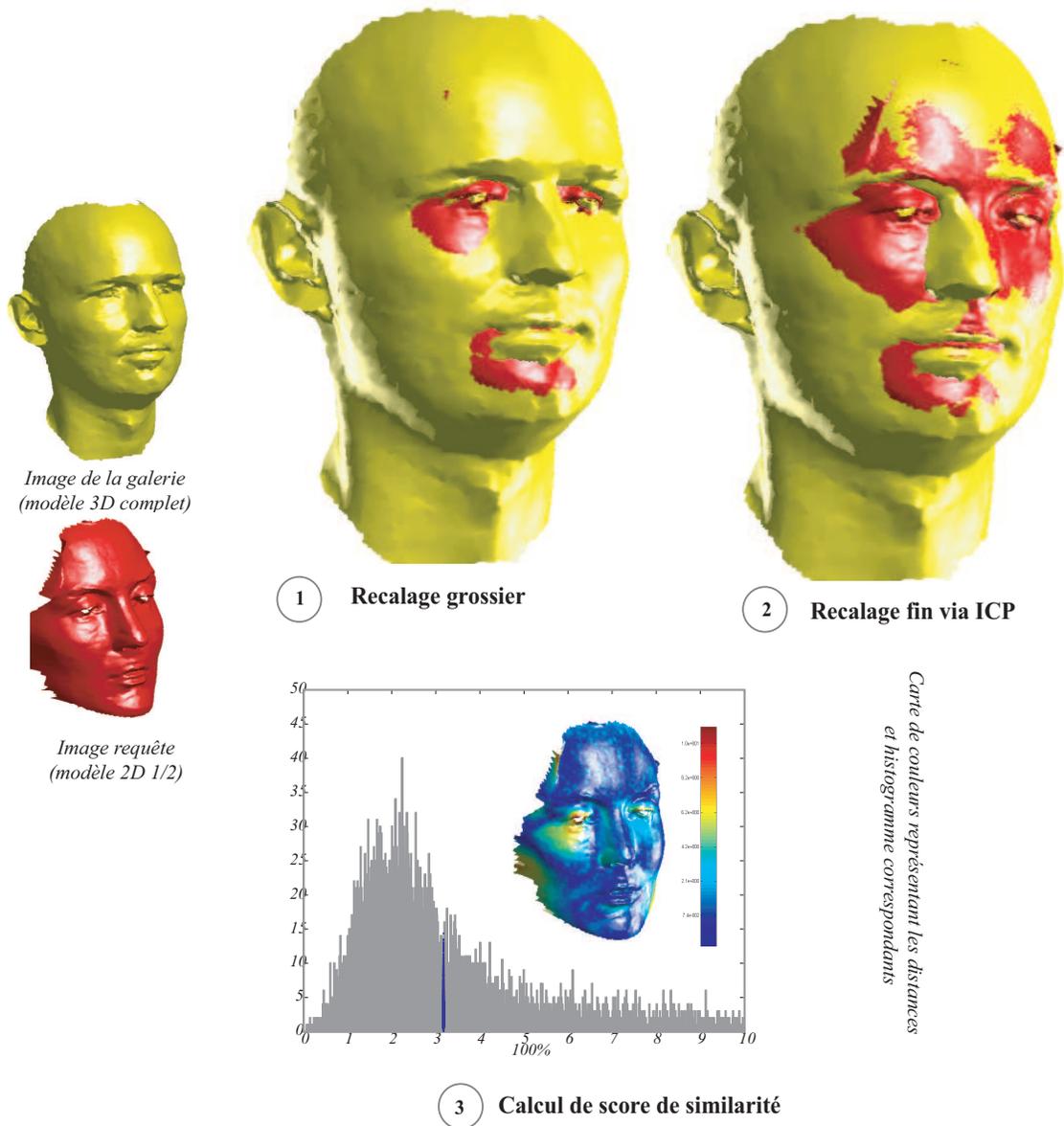


FIG. 4.6 – Exemple3 : recalage et calcul de similarité entre un modèle 3D et un modèle requête de deux personnes différentes.

l'algorithme *ICP* pour le recalage de surface et le calcul des scores de similarité en se basant sur la quasi-totalité des points du visage. A l'issue de cette étape, les deux maillages, celui du modèle complet 3D de la galerie et le modèle requête, sont bien mappés. Les figures affichent aussi les histogrammes issus de chaque comparaison et des cartes de couleurs mappées sur les images requêtes. Les couleurs appartenant à une palette de couleurs HSV (*Hue*, *Saturation*, *Value*) traduisent les valeurs des déplacements entre points appariés des deux modèles. Ces figures présentent différents cas de matching de surfaces : de deux personnes différentes par la figure 4.4, d'une seule personne avec des expressions faciales minimales par la figure 4.5, et enfin la figure 4.6 pour le cas d'un modèle requête avec des déformations plus importantes par rapport à un modèle 3D d'un même individu.

A l'issue de l'observation de ces trois exemples, cette première approche, qui consiste en l'application d'un algorithme de matching rigide de surfaces en vue de calculer leur différence de volumes, semble être sensible aux déformations des visages dues aux expressions faciales. Nos premières expérimentations, présentées dans le chapitre suivant, ont confirmées cette hypothèse très logique. En effet, alors que via ce paradigme de matching 2D vs. 3D, cette approche garantit une certaine invariance aux problèmes de pose et d'illumination, celle-ci est moins précise s'il s'agit d'une déformation faciale considérable. Pour résoudre ce problème, il y a au moins deux solutions qui se présentent : (a) la modification du critère à minimiser moyennant *ICP* par introduction d'un modèle élastique qui tienne en compte les variations des expressions ou (b) l'introduction de modifications sur *ICP* lui-même et l'élaboration d'une nouvelle métrique de similarité qui tient en compte de ces variations. Notre choix était d'éliminer la première solution, vue la complexité de modéliser les expressions d'un visage et par peur de perdre de précision sur la transformation rigide par introduction d'une énorme batterie de paramètres à optimiser. Ce choix est justifié par l'étude anatomique du maillage musculaire du visage. En effet, la structure musculaire peut engendrer un nombre très grand d'expressions via. Nous adoptons ainsi la deuxième solution qui consiste à effectuer des modifications au premier algorithme de matching global basé sur dérivé de *ICP* et de concevoir une nouvelle métrique plus robuste aux changements d'expressions faciales. Les sections ci-après présentent cette nouvelle variante d'*ICP* adaptée pour les visages ainsi que la métrique de comparaison pour faire face aux problèmes d'expressions et d'autres. Mais nous traitons d'abord le problème d'initialisation de l'algorithme *ICP* pour garantir une bonne convergence.

4.3.2 Initialisation de l'algorithme *ICP*

Tous les algorithmes de type *ICP* exigent une transformation initiale comme point de départ, et le résultat final dépend de cette transformation initiale. En fait, elle conditionne le minimum local qui sera trouvé, qui est le minimum de la cuvette dans laquelle se trouve l'initialisation. Comme nous l'avons déjà dit, les minima manifestement faux sont éloignés du recalage recherché, et on atteindra avec une mauvaise initialisation. Avec une bonne initialisation, on tombera dans l'un des nombreux minima situés à proximité de la bonne solution. Nous le répétons ici, cette variabilité des bonnes solutions ne doit pas être interprétée comme une conséquence de l'initialisation, mais plutôt comme le bruit interne de l'algorithme. Nous

présentons dans cette section deux méthodes d’initialisations. La première est celle que nous utilisons, est basée sur des points d’intérêts localisés sur le modèle (un minimum de trois points). La seconde est plus générale, mais très grossière, et en pratique elle utilisable dans le matching de 2 1/2D vs. 2 1/2D. Ce type d’initialisation peut être utilisé uniquement pour le matching de vues frontales.

4.3.2.1 Utilisation de points d’intérêts

Cette méthode propose de calculer la transformation rigide initiale à partir de deux ensembles de points localisés et appariés sur les deux modèles à recaler. La désignation, manuelle ou automatique, d’un certains nombre de points d’intérêts (minimum trois paires) sur les deux nuages de points permet de calculer une bonne transformation (solution) initiale pour démarrer l’algorithme *ICP*. Sur le modèle 3D complet, la sélection de ces points est faite manuellement, dans le cadre de la partie off-line du paradigme proposé. Alors que sur le modèle 2 1/2D, les points sont localisés automatiquement moyennant un certains nombre de descripteurs issus de la géométrie différentielle, à savoir les courbures moyennes, gaussiennes, minimales et maximales. En réalité, le système à résoudre nécessite quatre paires de points correspondants au minimum, puisque c’est un système à douze inconnus. En pratique nous désignons trois paires de points et nous calculons leurs barycentres respectifs comme quatrièmes points correspondants.

Une problématique importante qui se pose à ce niveau du raisonnement, quels points choisir ? Le choix des points à mettre en correspondance, manuellement ou automatiquement, est primordial à cause de son influence direct sur la convergence de l’algorithme *ICP* vers une solution satisfaisante. Ci-après les différents critères de choix à respecter pour la mise au point du protocole :

- Les points d’intérêts doivent être facilement repérable sur les modèles tridimensionnels dans le but de diminuer le risque d’erreur de sélection manuelle ou automatique.
- Les points désignés ne doivent pas être aligné dans le but de calculer correctement la transformation rigide qui les sépare.
- Les points doivent être stables par rapport aux variations d’expressions faciales et doivent appartenir à des régions statiques du visage. Ceci permettra la moindre variabilité entre solutions avec ou sans expressions faciales. Ceci rentre aussi en cohérence avec l’approche de recalage fin que nous proposons. En effet, vu le caractère déformable de notre face, nous proposons de se limiter à quelques régions du visage qualifiées de statiques pour effectuer le matching.

Une fois les paires de points sont localisés sur les modèles 3D et requête, la calcul de la transformation initiale est résolu moyennant une décomposition en valeur singulière (SVD) de la matrice de covariance liant les deux ensembles (de même que pour *ICP*), nous en déduisons ainsi les paramètres R_0 et t_0 qui constitue cette transformation initiale.

4.3.2.2 Utilisation des barycentres

Cette méthode d'initialisation est simple à mettre en œuvre, néanmoins elle reste très grossière. Elle est inspirée de la physique des solides. Il s'agit de mettre en correspondance le barycentre et les axes d'inerties des deux modèles à recaler.

On calcule donc les barycentres des deux nuages de points :

$$\bar{p} = \frac{1}{N} \sum_{i=0}^N p_i \text{ et } \bar{y} = \frac{1}{N} \sum_{i=0}^N y_i$$

La rotation est supposée nulle alors que la translation est donnée par :

$$R_0 = 0 \text{ et } t_0 = \bar{p} - \bar{y}$$

Cette méthode peut être utilisé pour deux types de matching seulement, 2 1/2D vs. 2 1/2D ou alors 3D vs. 3D. Elle est même suffisante pour ces types de matching pour avoir une bonne initialisation. Néanmoins, elle présente des limitations en dehors de ces deux types de matching. En effet, la rotation que peut présenter un modèle 2 1/2D par rapport à un modèle 3D n'est pas négligeable. En plus, le calcul de barycentre dans les deux cas 2 1/2D et 3D ne donne pas le même type de résultats. Dans notre paradigme de reconnaissance nous n'utilisons pas cette méthode d'initialisation mais plutôt la première basée sur les points d'intérêts.

4.3.3 Notre variante de l'algorithme *ICP* : *R-ICP*

Comme illustré par la figure 4.6 l'application de cette procédure itérative de recalage rigide ne traite pas le problème d'expressions faciales qui semble être compliquer en 2D qu'en 3D. En effet, en se basant sur cet algorithme de matching rigide, considéré comme le cœur de notre approche, nous avons essayé de garantir les invariances par rapport aux variations d'éclairage et aux changements de posture. En revanche, l'application d'*ICP* peut présenter des limitations à ce niveau du traitement, puisque qu'il est destinée au recalage d'entités statiques à la différence de la surface faciale qui est une entité déformable et présente une infinité d'états d'expressions faciales grâce à son modèle de muscles complexe. Ces déformations engendrent, essentiellement, des différence d'épaisseurs de plus en plus importantes. En outre, elles ont une influence directe sur l'algorithme itératif de recalage.

Pour remédier à ces problèmes, nous avons pu constater, dans un premier temps, que dans la structure anatomique de notre face, il existe des régions plus statiques que d'autres et des points plus stables que d'autres. D'où l'idée de concevoir une métrique qui traite différemment les différentes régions d'un visage et qui peut avoir d'influences souhaitables sur la mise au point d'une nouvelle variante de l'algorithme *ICP* pour plus d'invariance aux expressions faciales. Il s'agit en plus de mettre en place une nouvelle métrique de similarité qui prenne en compte toutes ces spécificités et ces différences de rigidité entre les différentes zones du visage d'une manière pondérée en vue d'atténuer les effets des expressions du visage.

Avant d'introduire des modifications sur l'algorithme de matching rigide *ICP*, un schéma unifié de segmentation a été élaboré en se basant, d'abord, sur une étude anatomique du maillage musculaire du visage. Puis, des observations et des mesures de déformations ont été, prélevées afin d'en déduire un classement des différents muscles et régions du visage en fonction de leurs degrés de déformation. Cette étude couplée anatomique et empirique a donnée naissance à un schéma de segmentation unifié qui sera appliqué uniquement sur l'ensemble des modèles 3D complets de la galerie. Cette partie du traitement rentre dans la phase offline de notre paradigme de reconnaissance/vérification et nous présentons une procédure automatique pour l'effectuer. Le schéma de segmentation appliqué à chaque modèle 3D servira à définir, d'une part, un ensemble de poids de pondération des scores de similarités élémentaires (issue de chaque région). D'autre part, le fonctionnement de l'algorithme est influencé par cette segmentation vue une pondération binaire qui va être introduite. Cette pondération consiste simplement à rejeter (poids nuls) les points appartenant aux régions mimiques et de considérer que les points les plus statiques pour itérer *ICP*. Le critère à minimiser devient donc (4.3.20) :

$$e_{R-ICP}(R, t) = \frac{1}{N} \sum_{i=0}^N w_i \|(Rp_i + t) - y_i\|^2 \quad (4.3.20)$$

Où w_i présente les pondérations selon l'appartenance du point p_i à une région statique R_s ou non selon la règle suivante (4.3.21) :

$$w_i = \begin{cases} 1 & \text{si } p_i \in R_s \\ 0 & \text{sinon} \end{cases} \quad (4.3.21)$$

Ce nouveau critère à minimiser s'appuie seulement sur les points statiques du visage appartenant à des régions peu déformable. Le matching rigide est, ainsi, plus significatif du fait qu'il est moins sensible aux déformations subies par le visage. Nous considérons les points issus des zones déformable comme du bruit qui peut empêcher l'algorithme de converger vers une solution proche de la solution correcte. Les sections suivantes détailleront l'élaboration de notre schéma de segmentation unifié. D'abord, nous décrivons brièvement la structure anatomique (surtout musculaire) de notre visage, puis la technique de mesure de l'influence de chaque muscle (ou zone) de notre visage sur la déformation globale de celui-ci. En fin, un schéma unifié de segmentation est donné et appliqué aux visages 3D complets de l'ensemble de la galerie.

La figure 4.7 présente le même exemple illustrée dans 4.6 avec différents poids de pondération des régions mimiques du visage. la première colonne de cette figure présente des poids égaux (=1), l'histogramme de déviations spatiales présente de grandes valeurs ce qui diminue la similarité entre les deux modèles d'un même individu et ambigu, par conséquent, la décision de reconnaissance. La deuxième et la troisième colonne de la même figure présentent l'atténuation de l'effet des parties mimiques du visage par attribution d'une pondération moins importante que celle attribuée à la région statique du visage. Il est clair que les silhouettes des distributions affichées présentes moins de grandes déviations devant

une concentration vers l'origine du repère. Les effets de cette nouvelle métrique orientée région sont également illustrée pas les différentes cartes de couleurs traduisant les déviation spatiale entre les modèles.

4.4 Segmentation du visage 3D en régions

Dans cette section, nous allons essayer de répondre à la question suivante : Quel schéma de segmentation adopté? Cette question a beaucoup d'importance et d'influence sur les résultats et les taux de reconnaissance attendus. Vu cette importance, notre segmentation est basée sur deux études complémentaires. La première étude s'intéresse à la structure musculaire de notre visage afin de donner une compréhension sur les états de déformation possibles. Elle fournira aussi des indices sur les muscles et les parties du visages qui contribuent plus à la déformation globale de celui-ci pour synthétiser ses émotions et ses expressions. La deuxième étude consiste à trier ces muscles et ces parties élémentaires du visage en fonction de la déformation introduite par chacun d'entre eux. Ce classement définira par la suite notre schéma de segmentation générique.

4.4.1 Étude anatomique du visage

Le but de cette étude est de fournir une compréhension sur la structure anatomique du visage afin de sculpter nos premières observations et d'en déduire un schéma de segmentation unifié et efficace. C'est essentiellement la structure musculaire qui est étudiée dans cette partie et son influence sur la déformation de la forme globale du visage. D'après de nombreuses études faites dans le domaine de l'anatomie du visage et de la tête en général, notre visage comporte trois étages et une batterie de muscles à chacun d'entre eux :

- **Étage inférieur** : qui comprend la mandibule, le menton et la bouche, représente la vie instinctive et active et donne des indications sur la manière d'utiliser l'énergie et d'agir (puissance, résistance, persévérance, précision, rythme, pragmatisme, etc.) C'est la zone instinctive. Les grandes déformations globales de la forme 3D sont dues essentiellement à cette partie qui a plus d'un degré de liberté.
- **Étage médian** : qui comprend les pommettes, les joues et le nez, représente la vie affective et relationnelle et donne des indications sur la sociabilité, l'empathie, l'autorité, l'influçabilité, la susceptibilité, l'émotivité, etc. C'est la zone affective. C'est une partie figée en comparaison avec la première. Néanmoins, elle comporte un tissu musculaire important ce qui explique son influence assez considérable sur la déformation du visage 3D.
- **Étage supérieur** : qui comprend le front, les tempes et les yeux, représente la vie

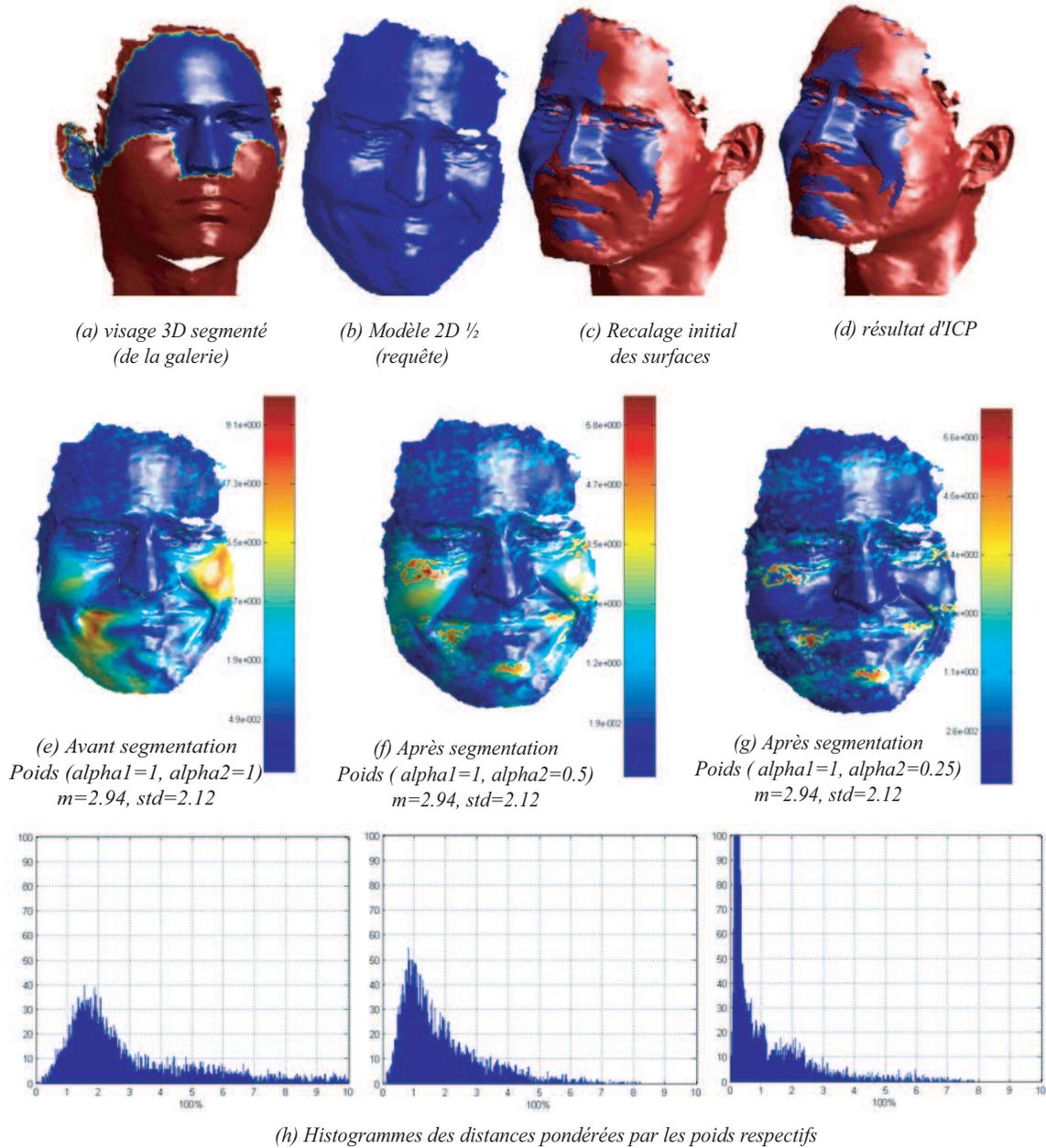


FIG. 4.7 – Compensation des expressions faciale par l'approche région.

cérébrale et donne des indications sur la forme d'intelligence comme la logique, l'intuition, la créativité, le sens de l'observation, de l'analyse, etc. et non le niveau d'intelligence. C'est la vie cérébrale. Cette partie du visage est aussi figée en comparaison avec la première. Les déformations de cet étage sont, notamment, beaucoup moins importantes que celles causées par l'étage inférieur et l'étage médian.

Outre ce fractionnement du visage en étages, chacun d'entre eux dispose d'un tissu musculaire différent et complexe afin de synthétiser toutes les émotions et les expressions possibles. Environ 44 muscles sont impliqués dans les déformations importantes du visage. Ces muscles trouvent leurs origines au niveau des os du crâne et se fixent sur les tissus mous de la peau du visage tels que paupières, nez, joues et lèvres. La figure 4.8¹ illustre ces structures osseuse et musculaire en donnant le label de chacune des parties du crâne et des muscles correspondant.

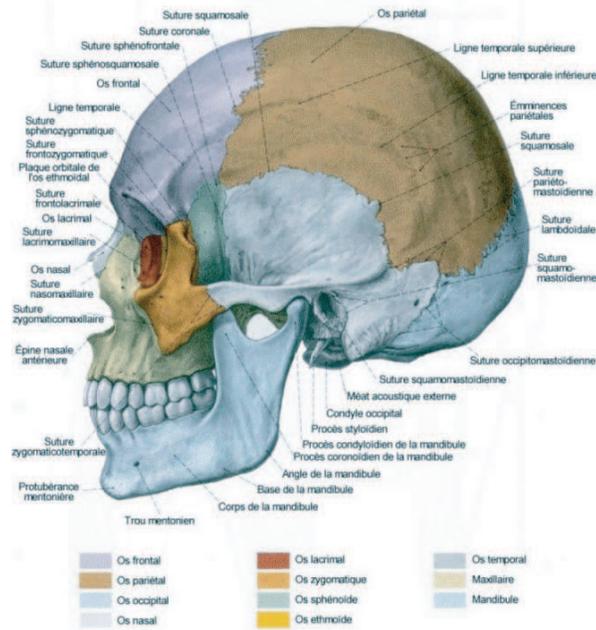
Plusieurs études ont été élaborées en vue de définir des modèles mathématiques en rapport avec ce sujet et pour différents types d'applications. Ces modèles sont utilisés dans le domaine de l'informatique graphique où l'animation faciale 3D présente une des préoccupations de la communauté scientifique. Certains travaux ont essayé, premièrement, d'analyser la structure du visage en vue de fournir une compréhension sur ses différentes parties et surtout son maillage musculaire. Ils ont ainsi essayé de reproduire des animations réelles sur des modèles virtuels via des formules mathématiques agissant sur le maillage de la forme 3D. Ces expressions (déformations) faciales peuvent être approximées moyennant une combinaison linéaire de déformations élémentaires issue de chacune des différents muscles. La forme du visage peut être représentée par (4.4.1) :

$$\vec{s} = \vec{s}_0 + \sum_{i=1}^{Nm} c_i \vec{e}_i \quad (4.4.1)$$

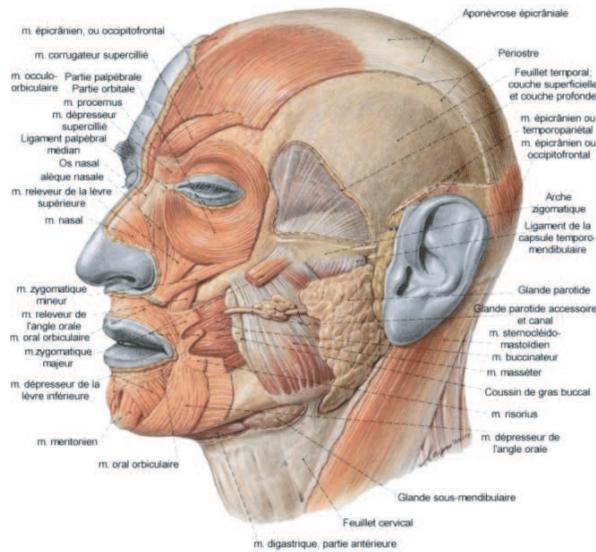
Où \vec{s}_0 représente la forme neutre du visage, \vec{e}_i sont ses déformations élémentaires (ou unité d'action) au niveau de chaque muscle. L'ensemble c_i présente les paramètres de pondérations. Ces paramètres caractérisent la contribution de chaque muscle à la déformation globale du visage au niveau de son rayon d'action.

Après avoir donné cette étude anatomique du visage qui fournit une compréhension sur sa structure géométrique, nous nous intéressons maintenant à la segmentation du visage en régions. Le but étant de limiter l'influence, de certaines régions dans le visage, sur la dégradation des résultats de notre approche de reconnaissance faciale. Nous proposons dans cette étape du travail de mener une étude permettant la localisation de différentes régions et de leurs influences sur la déformation globale de la forme 3D du visage. C'est l'objectif de la section suivante qui présente une étude empirique complémentaire à la précédente et qui cherche à mesurer l'influence, de chaque muscle (les principaux muscles), sur la déformation de la forme tridimensionnelle globale du visage.

¹<http://www.cegep-baie-comeau.qc.ca/bio>



(A) Etude anatomique du visage: Crâne et son contenu (Os, encéphale, méninges)



(B) Etude anatomique du visage: Muscles de la tête (de l'expression faciale, de la mastication)

FIG. 4.8 – Etude de l’anatomie du visage : composition osseuse et maillage musculaire.

4.4.2 Influences des muscles et des régions sur la forme faciale 3D

L'étude précédente nous a permis de comprendre l'architecture en étages et la constitution musculaire que comporte le visage. Nous avons pu, également, identifier quelques muscles élémentaires clés qui donnent naissance aux importantes expressions faciales. En effet, outre la partie inférieure du visage, nous disposons d'un ensemble de 11 muscles particuliers qui sont : les ventres frontaux gauche et droite, le muscle mentonien, les orbiculaires des yeux, la partie palpébrale de l'oeil, le muscle risorius, le corrugateur, le muscle orbiculaire des lèvres, le muscle abaisseur de la bouche et les Zygomatiques. Seules ces muscles sont considérés dans notre étude, la figure 4.9 présente les déformations causées par ces muscles unitaires sur un visage d'un individu. Nous essayons de les classer selon des mesures prises en comparaison avec un modèle neutre du visage. Nous introduisons dans cette section une technique de mesure des déformations causées par chacune des unités musculaires dans leurs rayons d'action.

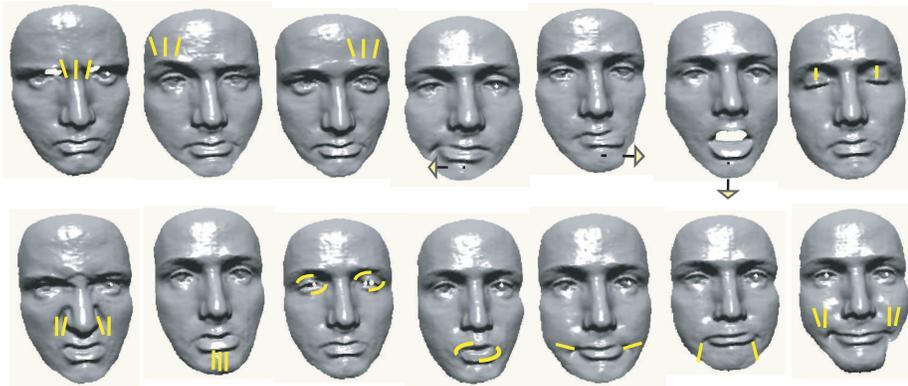


FIG. 4.9 – Ensemble des muscles élémentaires considérés pour l'étude empirique.

Il s'agit donc de considérer un modèle du visage avec une action musculaire unitaire et à le comparer à un état neutre. En pratique, nous avons demandé, à un ensemble d'individus (cinq personnes) d'effectuer un ensemble d'actions musculaires unitaire en essayant de faire bouger un des muscles unitaire énumérés au part avant. Ensuite, nous avons recalé les modèles obtenus par numérisation 3D par rapport à un modèle neutre, en vue de mesurer la déformation conséquent. La figure 4.10 illustre cette procédure de recalage puis de mesure d'erreur entre les deux surfaces 3D. Une carte de couleur caractérise aussi les zones de déformations ainsi que les degrés de celles-ci.

Le recalage et le matching des surfaces 3D par rapport au modèle neutre sont faits moyennant l'algorithme *ICP (Iterative Closest Point)* décrit plus haut. Après l'appariement point à point, une distribution des distances entre les deux surfaces est produite pour les mesures de déformations. Le but est de profiter de ces mesures pour classer les muscles en fonction de leurs contributions à une expression faciale donnée. Suite à ce classement,

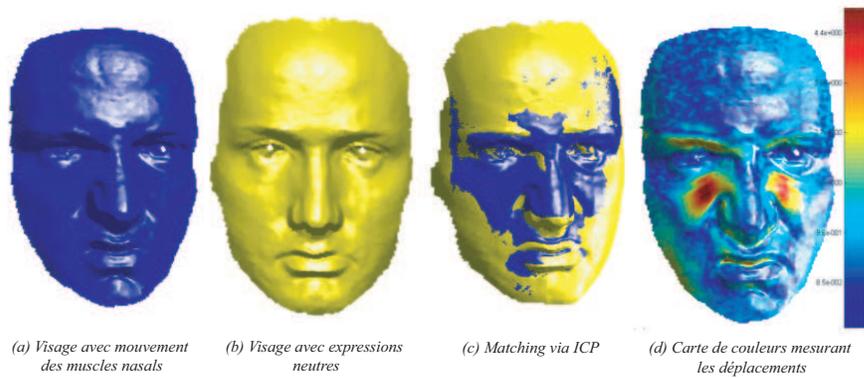


FIG. 4.10 – Technique de mesure de l'influence de chaque muscle ou région du visage sur la déformation de celui-ci par rapport à un état neutre.

nous avons pu élaborer un schéma de segmentation en régions générique pour tous les modèles complets 3D. La figure 4.11 montre un tri croissant des muscles en fonction de leurs influences aux changements de la forme 3D du visage d'un individu. Il est clair que l'étage inférieur est la zone qui contribue plus à ces changements. En effet, c'est les degrés de liberté de cette partie qui est à l'origine de la déformation la plus importante de la forme 3D du visage. Les grands déplacements que peut faire cette partie donne naissance à une différence de volumes énorme calculée moyennant l'algorithme de matching rigide *ICP*. Juste après, on trouve les zygomatiques (petits et grands) et les muscles releveurs de la lèvre supérieure placés au dessous de joues qui sont responsable à l'émotion de joie et de dégoût, respectivement. Moins influents, on trouve les muscles épicroâniens (les deux ventres frontaux), les muscles occulo-orbitaire, c'est à dire les muscles couvrant le fronts et dans les orbites des yeux, respectivement.

Cette étude de la structure musculaire du visage ainsi que le tri des muscles élémentaires selon les déformations causées par chacun d'entre eux, nous permet de proposer, dans un premier temps un fractionnement du visage en deux zones. C'est ce que nous essayons d'élaborer dans cette étape.

4.4.3 Schéma de segmentation adapté

La figure 4.17 illustre par des modèles colorés le schéma de segmentation suggérer pour améliorer les performances du système initial à base de matching rigide des surfaces faciales. Par l'introduction d'une métrique orientée région, le système, sera capable de prendre en compte différents labels de régions. Ce schéma est appliqué à tous les modèles 3D de visages qui composent l'ensemble de la galerie. Les mêmes couleurs sont conservées pour toutes les régions correspondantes à des labels par rapport à une échelle statique/mimique que nous avons défini suite à ces études anatomique et empiriques.

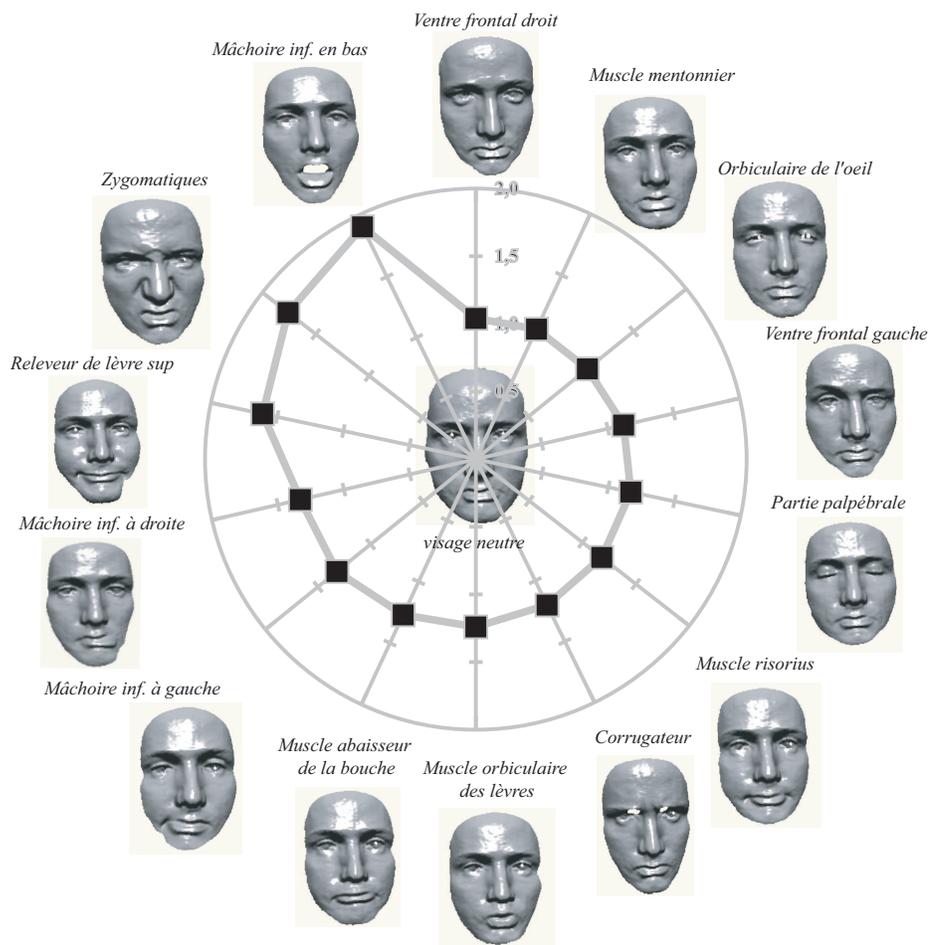


FIG. 4.11 – Ensemble des muscles et parties élémentaires considérés pour l'étude empirique.

La nouvelle métrique agira donc en fonction du label (ou couleur) du point dans le maillage 3D. C'est la somme pondérée des scores élémentaires fournis par chaque région qui définira le degrés de correspondance entre deux visages. Cette phase de segmentation concerne uniquement les modèles 3D de la galerie, en effet, suite à l'application de l'algorithme de recalage *ICP*, un appariement point à point est résultant. Une paire de points correspondants portera le même label. Cette étape de segmentation du visage et d'attribution de label fait partie de la phase off-line où un modèle 3D est d'abord formé puis segmenté.

Schéma de segmentation sur le visage de référence

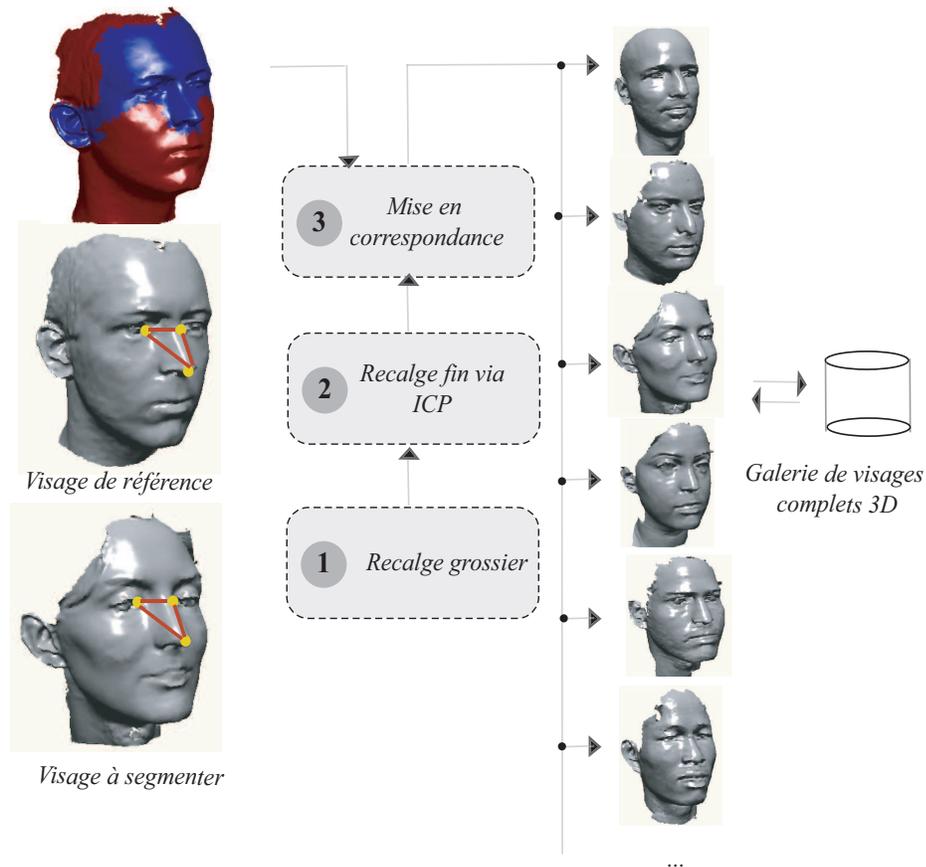


FIG. 4.12 – Procédure de segmentation automatique des visages 3D complets de la galerie.

Nous proposons dans notre approche une procédure de segmentation automatique de la base des visages 3D complets. Cette procédure consiste à recalcr un modèle de référence pré-segmenté par rapport à tout les visages appartenant à la galerie via une étape d'alignement.

Cette étape est faite moyennant l'algorithme *ICP*, qui outre l'estimation des transformations rigides entre les modèles 3D, permet de mettre en correspondance l'ensemble des points des deux modèles. Nous nous servons de cet appariement pour labeller le modèle 3D à segmenter. Un deuxième modèle 3D segmenté est alors obtenu et peut être incorporé à la base de visage 3D segmentés. La figure 4.17 illustre les étapes d'obtention du modèle segmenté moyennant l'alignement par rapport à un visage de référence. Cette procédure est donc entièrement automatique introduite dans la phase off-line de notre paradigme de reconnaissance/authentification.

4.5 Calcul de carte géodésique en vue de la reconnaissance

Une des approches les plus récentes et qui traite, plus particulièrement, le problème de changement d'expressions faciales en reconnaissance de visages 3D, est celle de calcul géodésique sur les surfaces faciales 3D. Dans [BBGK04], les auteurs proposent cette solution en se basant sur l'hypothèse suivante : le visage est une surface isométrique. Ce type de surfaces préserve la métrique des surfaces par rapport aux déformations subies. Cette hypothèse est entièrement empirique déduite d'un ensemble d'observations faites sur le visage avec différentes expressions. Le calcul de géodésie, dans cette approche, vise la production d'une signature invariante aux transformations non-rigides subies par le visage. Dans le présent travail, nous ne reproduirons pas la même approche décrite dans [BBGK04] mais nous nous intéressons à l'hypothèse de base à savoir : la propriété de préservation de longueur des surfaces du calcul géodésique et son application à la reconnaissance faciale 3D.

Dans la suite de cette section, nous donnons, d'abord, une brève description de la théorie sur laquelle se base cette étude. Puis, nous nous intéressons à son application dans le domaine triangulé. Finalement, nous décrivons l'approche adoptée pour inclure cette mesure dans le problème de reconnaissance de visage 3D. Notre objectif par cette implémentation est la comparaison de cette approche à celles basées sur le recalage de surfaces (*ICP*) et le recalage orientée region (*R-ICP*). Une section, dans le chapitre suivant, est consacrée à cette comparaison des trois approches. Elle présente les différents résultats obtenus, en terme reconnaissance et en authentification, sur le corpus *ECL-IV*².

4.5.1 Calcul Géodésique

Le calcul numérique des géodésiques est d'abord passé par la théorie des graphes. En 1959, Edsger Dijkstra propose un algorithme particulièrement élémentaire dans sa conception et redoutablement efficace, quoique gourmand en calculs, pour retrouver le plus court chemin dans un graphe [Dij59]. Depuis, de nombreux algorithmes ont été proposés dans ce contexte, motivés par les applications de ce type de problème. Enfin, plus récemment, Sethian ([Set96]), s'inspirant de l'algorithme de Dijkstra et de résultats sur le calcul variationnel, a proposé un algorithme rapide pour calculer des distances dans un espace discrétisé par une grille orthogonale, il s'agit de l'algorithme de cheminement rapide (*Fast Marching Algorithm*). Sethian et Kimmel proposent par la suite dans ([KS98]) d'étendre la théorie de

cheminement rapide des grilles orthogonales vers les maillages 3D.

4.5.1.1 Définition d'une géodésique

Une ligne géodésique est le plus court chemin pour relier deux points en fonction des contraintes du milieu. Le terme *géodésique* vient de géodésie (du grec *gê* "terre" et *daiein* "partager, diviser")². C'est la science de la mesure de la taille et de la forme de la Terre. Néanmoins, les mathématiques fournissent une définition plus précise. Une géodésique est la généralisation de la notion de ligne droite aux espaces courbes. La définition de la géodésique dépend du type d'espace courbe. Dans le cas où cet espace dispose d'une métrique, alors la géodésique se définit effectivement comme, localement, le chemin le plus court entre deux points de l'espace.

4.5.1.2 Théorie de Sethian

L'idée de Sethian était de décrire de manière analytique les équations qui relient le temps et la position, et de discrétiser ces équations avant de les insérer dans un algorithme de propagation de type Dijkstra. Considérons l'équation du mouvement qui relie la variation du temps dt à la variation de distance dx :

$$dt = \frac{1}{v} dx. \quad (4.5.1)$$

où v est la vitesse de déplacement. Cette équation montre que la distance est proportionnelle au temps. Ainsi, pour une vitesse de déplacement donnée, minimiser le temps de parcours revient à minimiser la distance à parcourir. En considérant x le temps d'arrivée en un point, T une fonction de la coordonnée locale ($t = T(x)$) et en partant d'un point de départ x_0 , l'équation précédente se réécrit :

$$v \frac{dT}{dx} = 1. \quad (4.5.2)$$

En dimensions multiples, le gradient de la fonction T est aussi inversement proportionnel à la vitesse de propagation :

$$\|\nabla T\| = \frac{1}{v}. \quad (4.5.3)$$

Cette équation, connue sous le nom d'équation iconale (du grec *eikon* qui signifie petite image), est notamment utilisée dans le domaine de l'optique géométrique où elle sert à décrire la trajectoire de la lumière. L'équation iconale, dans le cas des géodésiques, peut aussi être retrouvée dans un cadre plus général grâce à des résultats de calcul variationnel.

²<http://www.techno-science.net>

En effet, si l'on note x_0 le point de départ, la fonction T représente la longueur du chemin minimum qui relie x_0 à tout point x :

$$T(x) = \min_{\gamma \in \Gamma_{x_0 x}} \int F(\gamma(s)) ds, \quad (4.5.4)$$

où $F = \frac{1}{v}$. La fonction T réalisant le minimum de cette intégrale satisfait l'équation iconale (4.5.3). La fonction T est donc solution d'une équation aux dérivées partielles (EDP), à laquelle il faut ajouter des conditions aux limites. Dans notre cas, la fonction est nulle à l'origine, c'est-à-dire au point x_0 de départ de la propagation :

$$\begin{cases} \|\nabla T\| = F \\ T(x_0) = 0. \end{cases} \quad (4.5.5)$$

D'autre part, on peut démontrer que les lignes géodésiques sont perpendiculaires aux lignes de front de T , et parallèles à son gradient. Ainsi, le calcul de la fonction T en tout point de l'espace permet de reconstruire toutes les géodésiques entre le point de départ x_0 et tout point x , en remontant les lignes de gradient de T de x vers x_0 .

4.5.1.3 Algorithme de Cheminement Rapide

Le principe de l'algorithme de Sethian [Set96] est de construire la fonction T de sa plus petite à sa plus grande valeur, selon un principe analogue à celui de Dijkstra. La procédure de mise à jour de la valeur de T en un point à partir de son voisinage se fait en résolvant l'équation (4.5.5). A chaque étape de l'algorithme, le noeud de la grille qui possède la plus petite valeur de T est *gelé* (T y prend une valeur définitive), de la même manière que, dans l'algorithme de Dijkstra, on considérait un point comme étant marqué lorsque la fonction de coût y prenait une valeur définitive. En outre, lors de la mise à jour de la fonction de coût, seul le voisinage immédiat des points préalablement marqués, formant ce qu'on appelle la bande étroite, est considéré (voir figure 4.13).

L'algorithme suivant présente de manière schématique l'algorithme de cheminement rapide.

Définitions : Soit PA (Points Acceptés) l'ensemble des nœuds dont la valeur de T est gelée. Soit BE l'ensemble des nœuds de la Bande Étroite, et soit PE (Points Éloignés) l'ensemble des nœuds non encore explorés. On notera $\#E$ le nombre de nœuds appartenant à un ensemble E .

1. Initialisation :

$$X_0 \rightarrow PA, T(X_0) = 0;$$

$$\text{pour tout } X \neq X_0 : X_0 \rightarrow PA, T(X) = \infty;$$

mise à jour de T au voisinage de X_0 ;

2. Tant que $\#BE \neq 0$:

soit X_m le sommet de BE ayant la plus petite valeur de T ;

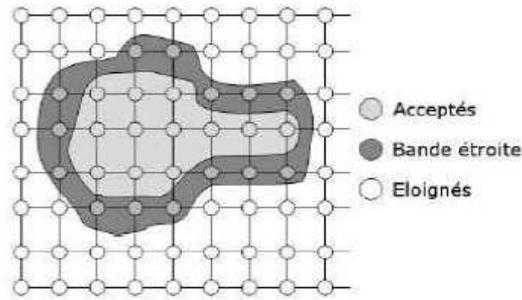


FIG. 4.13 – Illustration de l'algorithme de cheminement rapide.

mise à jour de T au voisinage de X_m ;
 $X_m \rightarrow PA$

La procédure de mise à jour consiste à résoudre l'équation iconale discrétisée (4.5.5). L'algorithme de cheminement rapide devient extrêmement efficace lorsque les points de la bande étroite sont stockés dans une structure de données de type *minheap* (ce qui pourrait être traduit par : tas de minima). Il s'agit de maintenir un arbre binaire équilibré dans lequel chaque nœud représente une valeur, qui est inférieure à celles des nœuds fils. Cela permet une détermination aisée du minimum (extraction de la racine de l'arbre).

- Lors de la destruction d'une valeur, le voisinage voit ses valeurs mises à jour et recalculées, aussi haut que cela est nécessaire pour équilibrer l'arbre.
- Lors de l'insertion d'une valeur, cette valeur est insérée à la fin de l'arbre, et les calculs sont effectués comme pour une mise à jour de valeurs.

L'utilisation de cette structure de données modifie la complexité de l'algorithme de cheminement rapide en la faisant passer de $\log(n^2)$ à $n\log(n)$, où n est le nombre de nœuds de la grille.

4.5.1.4 Cheminement Rapide sur une triangulation aiguë

Considérons une triangulation aiguë autour d'un sommet centre comme le montre la figure ???. Plusieurs triangles peuvent partager ce sommet centre. Il s'agit donc de calculer la valeur de T sur chaque triangle contenant le sommet en question. Puisque plusieurs triangles peuvent fournir une valeur acceptable de T , la valeur à garder est celle minimale sur toutes les valeurs obtenues [KS98]. Nous allons donc décrire la procédure de mise à jour de T au sein d'un triangle non obtus donné.

Soit un triangle ABC où C est le sommet à mettre à jour (voir figure 4.15). Soit $u =$

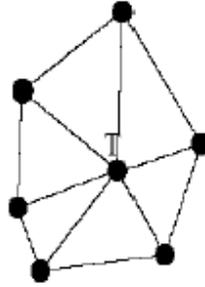


FIG. 4.14 – Triangulation aiguë autour d'un point centre de la grille.

$T(B) - T(A)$, le problème consiste donc à chercher $t = EC$ tel que :

$$\frac{t - u}{h} = F. \quad (4.5.6)$$

Ainsi, la distance géodésique en C sera $T(C) = t + T(A)$. Notons $a = BC$ et $b = AC$; nous avons donc $t/b = DF/AD = u/AD$, ainsi, $CD = b - AD = b - bu/t = b(t - u)/u$. La loi des cosinus donne : $BD^2 = a^2 + CD^2 - 2aCD \cos \theta$, et la loi des sinus donne : $\sin \Phi = \frac{CD}{BD} \sin \theta$. Utilisons l'angle droit du triangle CBG , nous avons :

$$h = a \sin \Phi = a \frac{CD}{BD} \sin \theta = \frac{aCD \sin \theta}{\sqrt{a^2 + CD^2 - 2aCD \cos \theta}}. \quad (4.5.7)$$

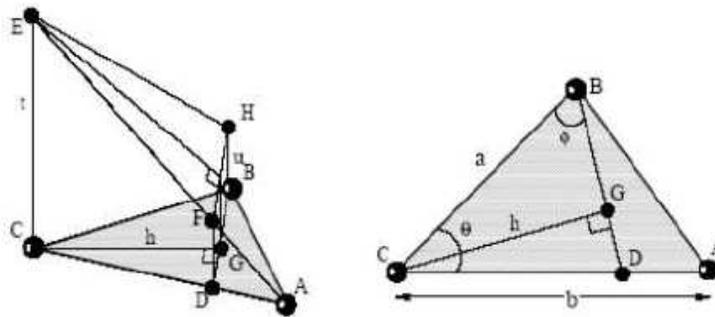


FIG. 4.15 – A gauche : Une vue perspective du triangle supportant les valeurs de $T()$, à droite : Trigonométrie sur le plan défini par le triangle ABC.

Enfin, nous aboutissons à l'équation quadratique de t suivante :

$$(a^2 + b^2 - 2ab \cos \theta)t^2 + 2bu(a \cos \theta - b)t + b^2(u^2 - F^2 a^2 \sin^2 \theta) = 0. \quad (4.5.8)$$

Deux solutions se présentent. Nous choisissons la solution t qui doit satisfaire $u < t$ et doit être mise à jour de l'intérieur du triangle, précisément :

$$a \cos \theta < b(t - u)/u < \frac{a}{\cos \theta}. \quad (4.5.9)$$

Ainsi, la procédure de mise à jour est :

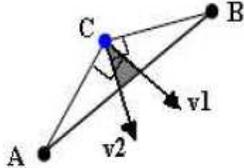
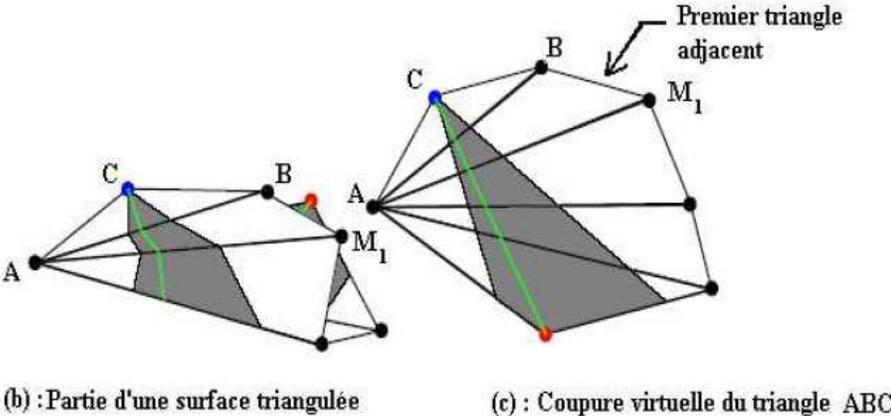
If $u < t$ **and** $a \cos \theta < b(t - u)/u < \frac{a}{\cos \theta}$ **then** $T(C) = \min\{T(C), t + T(A)\}$;
else $T(C) = \min\{T(C), bF + T(A), cF + T(B)\}$;

D'après Sethian et Kimmel [KS98], cette équation constitue une approximation à différences finies de l'équation iconale. Elle est monotone par construction, consistante et converge vers la solution de viscosité et donc elle peut être utilisée pour étendre la méthode de cheminement rapide vers les domaines de triangulation aiguë.

4.5.1.5 Extension vers une triangulation quelconque

Notons que le principe de cheminement rapide décrit ci-dessus exige que la triangulation soit aiguë. En d'autres termes, lorsqu'un front heurte un côté du triangle, il doit fournir les valeurs des deux premiers sommets avant de calculer celle du troisième ce qui garantit la monotonie du cheminement du front et de cette façon nous exigeons que la mise à jour vienne de l'intérieur du triangle [KS98]. Puisque un maillage 3D contient des triangles aigus et d'autres obtus, Sethian et Kimmel proposent de diviser d'une façon bien spécifique les triangles obtus en deux triangles aigus moyennant une arête virtuelle pour ramener le problème vers le cas d'une triangulation aiguë (voir figure 4.16).

Le principe consiste à considérer une zone intérieure à (CA, CB) formée par deux vecteurs v_1 et v_2 issus de C et dont l'angle commun est aigu. Ces deux vecteurs constituent durant tout cet algorithme le repère local à utiliser. Il s'agit d'exprimer CA et CB sur ce repère. Ensuite, on réalise un travail itératif consistant à chercher sur cette zone le premier sommet rencontré. Ce sommet là formera la nouvelle arête divisant le triangle ACB en deux triangles aigus. Cette zone s'appelle la zone de coupure. Pour trouver ce premier sommet du maillage sur la zone de coupure considérée, il suffit de prendre le triangle ABM_1 adjacent à ACB en AB . Nous vérifions si M_1 se trouve à l'intérieur de la zone de coupure en calculant ses coordonnées sur la base (v_1, v_2) : Si M_1 est à l'intérieur de la zone en question, l'algorithme s'arrête et CM_1 est l'arête virtuelle qui va diviser l'angle obtus en deux angles aigus. Sinon, nous vérifions sur quelle arête la zone aiguë intercepte ce triangle. Si elle intercepte la première arête AM_1 , alors nous refaisons le travail sur le triangle adjacent au triangle ABM_1 en AM_1 et si elle intercepte la deuxième arête BM_1 , alors nous refaisons le travail sur le triangle adjacent au triangle ABM_1 en BM_1 . L'algorithme est donc itératif, il évolue jusqu'à ce qu'un sommet du maillage soit rencontré dans la zone de coupure. Dès que toutes les distances sont calculées sur notre maillage, nous serons aptes à calculer n'importe quel



(a) : Angle obtus en C et zone de coupure

FIG. 4.16 – Cas d'une triangulation quelconque.

chemin géodésique menant d'un point donné vers la source en utilisant une descente de gradient par étapes (The Steepest Gradient Descent) [DC00].

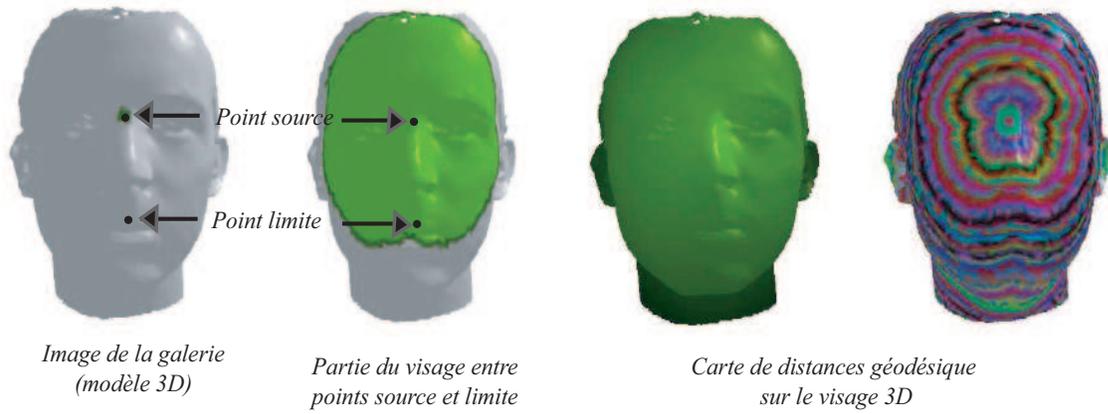
4.5.2 Application de l'algorithme du cheminement rapide dans le cas des surfaces faciales 3D

Comme décrit dans l'introduction de cette section, nous souhaitons intégrer ces mesures géodésiques, dans la procédure de reconnaissance faciales 3D en profitant de l'hypothèse décrite dans [BBGK04]. Elle consiste en la préservation de ce type de distances entre deux points quelconques d'une surface avec différentes expressions faciales. Cette caractéristique permet de faire face donc à ce type de variabilité dont souffrent les approches de reconnaissances de visages 2D et 3D. Pour se faire, nous commençons d'abord, par recalculer les modèles de visages : 2 1/2D de l'image requête et 3D du visage de la galerie. Puis, nous appliquons l'algorithme de cheminement rapide sur les parties communes des deux surfaces en vue de calculer les deux cartes de distances géodésiques. Une carte est formée par les distances mesurées entre deux points du maillage : le point source et le point limite. Finalement, un score de similarité est donné par comparaison des deux vecteurs de distances géodésiques calculées entre les ensembles de points appariés.

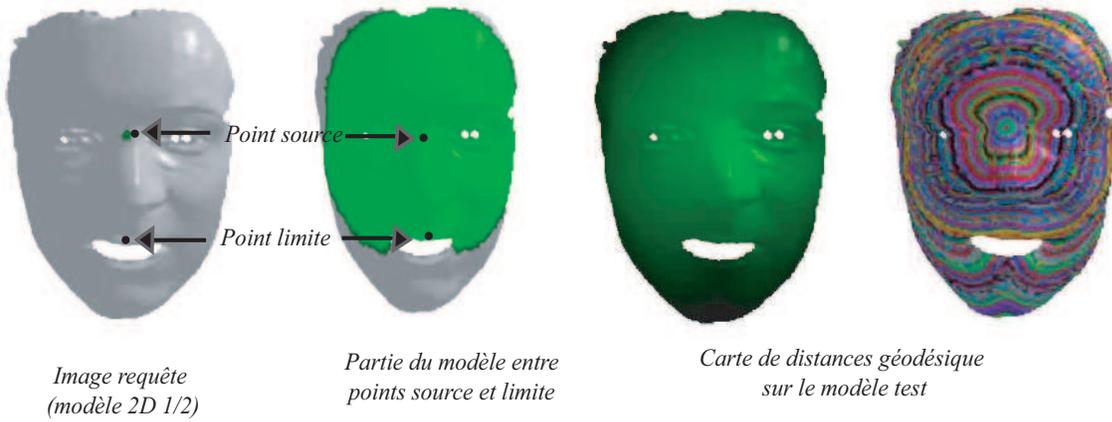
Plus concrètement, une fois le recalage 3D via l'algorithme *ICP* est fait, nous désignons un premier point source du cheminement rapide, sur le modèle 3D. Un second point doit être, également, désigné et considéré comme le point limite et dont nous allons bientôt expliquer l'utilité.

A l'issue de l'application de l'algorithme *ICP*, nous disposons d'une liste de points correspondants et aussi du point source que nous noterons S_1 du maillage 3D, du point limite L_1 sur le maillage 3D et de leurs correspondants le point source S_2 et le point limite L_2 sur le maillage 2 1/2D. Il s'agit donc de calculer les cartes de distances géodésiques sur les modèles 3D et 2 1/2D en appliquant l'extension du principe de cheminement rapide de Sethian et Kimmel [?] sur les domaines triangulés. L'ensemble de ces distances constituent la carte de distances géodésiques qui n'est autre qu'un vecteur V_1 dont chaque ligne i contient la distance géodésique séparant S_1 du point d'indice i du maillage 2 1/2D. Nous calculons enfin le vecteur V tel que $V = \|V_2 - V_1\|$ dont l'écart type constitue la mesure de similarité que nous utiliserons pour la reconnaissance.

Dans le calcul, nous n'avons pris en compte que les points des deux maillages qui se trouvent au dessus de la bouche. En d'autres termes, les vecteurs V_1 et V_2 ne contiennent que les distances géodésiques des points se trouvant au dessus de la lèvre supérieure. Nous calculons toutes les distances géodésiques sur la surface 2 1/2D et nous ne gardons que les distances dont la valeur est inférieure à la distance séparant S_1 de L_1 . De même, nous calculons toutes les distances géodésiques sur la surface 3D et nous ne gardons que les distances dont la valeur est inférieure à la distance séparant S_2 de L_2 . Nous justifions ce choix par le fait que dans le cas où la bouche est ouverte sur le modèle 2 1/2D, et qu'il présente un trou par conséquent, toutes les distances géodésiques correspondants aux points se trouvant au dessous de la lèvre inférieure sont très différentes aux distances géodésiques



(A) Calcul de la carte de distances géodésique sur un modèle 3D de la galerie



(B) Calcul de la carte de distances géodésique sur un modèle 2D 1/2 de test

FIG. 4.17 – Calculs des cartes de distances géodésiques sur les modèles 3D et 2.5D après recalage des surfaces moyennant ICP.

de leurs correspondants dans le visage 3D qui ne présente pas un trou au niveau de la bouche(voir figure ??). La figure ?? récapitule le principe de calcul géodésique que nous adoptons dans notre approche.

Sur cette figure, les points marqués sur les modèles présentent les sources des chemine-ments, et les limites de ceux-ci. Dans les deux cas, le point limite constitue le point le plus éloigné de la source qui peut figurer sur la carte géodésique. Les points du maillage qui sont colorés en vert sont les seuls points du maillage qui vont intervenir au niveau de la carte géodésique.

4.6 Conclusion

Nous avons présenté dans ce chapitre une approche de reconnaissance/vérification de visage par exploitation d'une nouvelle représentation du visage : la forme faciale 3D. Cette approche introduit un schéma de matching 2 1/2D vs. 3D en vue remédier au problème de changement de posture. Elle est basée sur un algorithme de matching de surfaces, *ICP*, qui a été amélioré pour prendre en compte les variations du visage dues aux expressions. En effet, une nouvelle variante de l'algorithme, appelé *R-ICP*, est donnée et une nouvelle métrique orientée région est présentée. Nous avons décrit aussi une approche de reconnaissance par calcul de cartes de distances géodésiques effectué à la suite du recalage 2 1/2D vs. 3D. Dans le chapitre 4, intitulé expérimentation, évaluations et discussions, nous présentons d'abord notre nouvelle base de visages 3D (*ECL-IV²*) et nous expérimentons les algorithmes développés. Les deux problèmes de reconnaissance et d'authentification sont traités par des courbes d'évaluations standards. Une comparaison de ses approches par rapport à l'approche de calcul géodésique effectué sur les surfaces faciale achèvera nos évaluations.



Chapitre 5

Expérimentations, évaluations et discussions

5.1 Introduction

Ce chapitre caractérise les performances des différents algorithmes de reconnaissance faciale développés dans le chapitre précédent. Ces algorithmes sont basés sur une nouvelle modalité biométrique, à savoir *la face 3D*, qui intéresse de plus en plus de chercheurs dans la communauté scientifique. Utilisée seule ou couplée avec la modalité 2D de texture, elle connaît un développement intensif ces quelques dernières années. En effet, la modalité 2D de visage a montré son insuffisance pour atteindre des taux de précision élevés tant en reconnaissance qu'en authentification. C'est, essentiellement, sa sensibilité aux changements de conditions d'éclairages, aux changements de postures et aux expressions faciales qui constitue ses limitations.

Dans nos travaux, nous introduisons cette modalité, seule, dans un paradigme de matching dédiée, 2_{1/2}D vs. 3D, en vue de palier les problèmes classiques de la reconnaissance de visages. Trois approches ont été développées basées toutes sur ce schéma de matching qui présente plus d'invariances aux problèmes de changements de pose et d'illumination par rapport aux approches classiques 2D. En vue d'évaluer ces différentes approches, nous faisons appel à des techniques de test très connues dans le domaine de la biométrie. En matière d'identification ou de vérification, ces techniques permettent d'obtenir des mesures de performances et même des comparaisons objectives entre différents systèmes ou algorithmes au sein d'une même ou différentes technologies. Ces techniques d'expérimentation et d'évaluation ont été amplement décrites dans le rapport publié par *A. J. Mansfield et J. L. Wayman* en 2001¹ et qui commence par donner des éléments explicatifs et des concepts fondamentaux liés à un système biométrique générique et qui s'achève par présenter les différents types d'évaluations ainsi que les critères de performances (taux, courbes, etc.) présents dans la littérature pour garantir une meilleure évaluation. Outre ce rapport, nous

¹www.cesg.gov.uk/site/ast/biometrics/media/BestPractice.pdf

nous basons, également, sur des campagnes d'évaluations qui ont eu lieu dans différents programmes à savoir [PGM⁺03], [PFS⁺03], [GSBC⁺03], etc. Cette étude doublé en théorie et en pratique nous a conduit à concevoir, sur des bases solides, un nouveau protocole d'évaluation de la modalité 3D du visage.

Dans les sections suivantes nous décrivons, d'abord, notre protocole de test de cette modalité utilisée au sein du paradigme proposé et nous revenons sur quelques définitions fondamentales concernant les techniques et les mesures d'évaluation. Ce protocole est parmi les premiers mis en place pour cette nouvelle modalité (*3D face*) qui est intégré au projet technovision² IV² dont le *LIRIS* est un des partenaires académiques. En suite, nous présentons les résultats obtenus des algorithmes développés dans cette thèse, via ce protocole. Finalement, nous présentons une étude comparative de ces techniques et des perspectives d'améliorations.

5.2 Protocole d'expérimentation et d'évaluation

Étant indispensable pour la compréhension et la définition de notre protocole de test, nous rappelons, dans cette section, quelques concepts sur les méthodologies d'évaluation des modalités biométrique. En outre, nous situons par rapport aux protocoles existants en effectuant une comparaison entre notre protocole et le protocole le plus connu, celui élaboré récemment par *NIST*, nommé *FRGC* pour (*Face Recognition grand Challenge*) pour cette modalité. Comme la plupart des protocoles d'évaluation, notre protocole s'articule autour de deux phases : **(a) l'acquisition des données** et **(b) la production des courbes et des taux d'évaluation** à partir de mesures de similarité effectuées en off-line sur ces données. Dans cette partie, nous décrivons, d'une part, la nature des données de test ainsi que les motivations derrière ces choix. D'autres part nous présentons les taux et les courbes d'évaluation suivis d'interprétations et de discussions.

5.2.1 Nouvelle base de données *ECL-IV2*

Malgré leurs développements récents, les bases de données publiques traitant cette nouvelle modalité [PFS⁺03][GSBC⁺03] sont peu nombreuses et peu satisfaisantes. Ceci nous a conduit à développer une nouvelle base de données plus riche en variations et en même temps qui répond aux besoins du paradigme que nous proposons pour la reconnaissance et l'authentification. En effet, comme décrit dans le chapitre précédent, pour garantir une reconnaissance des visages sous n'importe quelle angle de vue, la disponibilité d'un modèle 3D complet des visages dans l'ensemble des modèles de la galerie est indispensable. C'est, donc, la première spécificité de notre base de visages. La deuxième caractéristique de notre corpus est sa richesse par des images de test avec différentes variations touchant notamment aux expressions faciales, à l'illumination et à l'angle de pose du sujet. L'ensemble des données de cette base peut être, également, utiles pour d'autres types d'algorithmes via la

²<http://lsc.univ-evry.fr/techno/iv2/PageWeb-IV2.html>

variété des données qu'il propose (images 2D, modèles 2_{1/2}D et modèle 3D complets). Ceci est due à la disponibilité des deux canaux de forme 3D et de textures 2D qui peuvent être couplée ou découplées selon l'algorithme de traitement. Les algorithmes concernés peuvent être de différents natures selon les données de tests sur lesquelles ils sont basées et les types de matching considérés (voir figure 5.1) :

- **Approches de matching 2D vs. 2D** : il s'agit de comparer une image 2D à une image 2D (frontales ou avec changement d'angle d'orientation du sujet). La mise en correspondance entre les visages concerne les images de textures.
- **Approches de matching 2D vs. 3D ou 3D vs 2D** : il s'agit de comparer une image 2D à une image 3D ou inversement. Le modèle 3D est utilisé comme information intermédiaire pour effectuer la mise en correspondances entre deux images de textures.
- **Approches de matching 2_{1/2}D vs. 3D** : il s'agit comparer une forme partielle 2_{1/2}D à une forme plus complète 3D. La mise en correspondance concerne une information partielle et une information complète 3D.
- **Approches de matching 3D vs. 3D** : il s'agit de comparer une image 3D à une image 3D. La mise en correspondance concerne deux informations 3D complètes des visages.
- **Approches multi-modale 2D+3D** : consiste à fusionner les deux modalités forme et texture. La fusion peut être effectuée à plusieurs niveaux (scores, données, décisions, etc.).

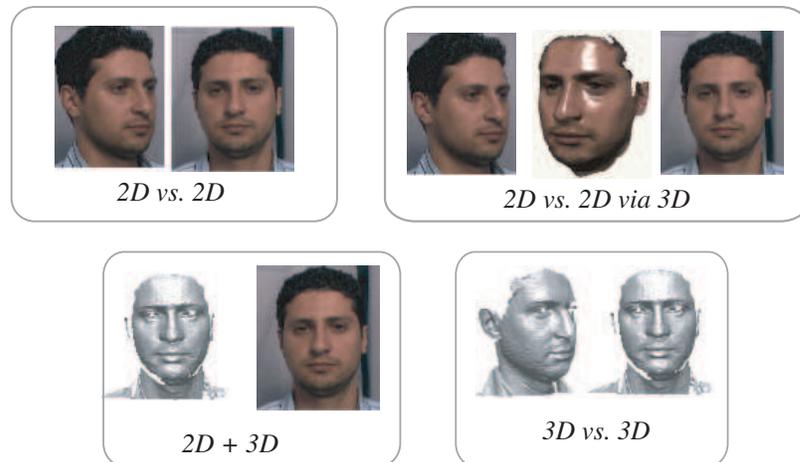


FIG. 5.1 – Variété des données de la base *ECL-IV²* et algorithmes de matching concernés.

Des réponses à ces besoins ont été traitées dans le cadre du projet *IV²*, où plusieurs schémas d'évaluations sont proposées dans un cadre plus général que le notre et visant

l'évaluation de plusieurs types d'algorithmes utilisant les deux modalités 2D et 3D. Concernant nos algorithmes, qui s'intéressent plus particulièrement aux comparaisons de type 2 1/2D vs. 3D, ils nécessitent la formation d'un modèle complet du visage. Ce dernier est obtenu moyennant une fusion de plusieurs modèles partiels qui sont d'abord acquis puis recalés et enfin fusionnés, comme illustre la figure 5.2. Ainsi, le visage 3D est prêt pour être inclus dans la base de reconnaissance sachant qu'il présente des expressions faciales neutres (avec les yeux ouverts). C'est donc de ce types d'images que se compose le premier ensemble de notre protocole d'évaluation. Cet ensemble est appelé dans la suite *la galerie*.

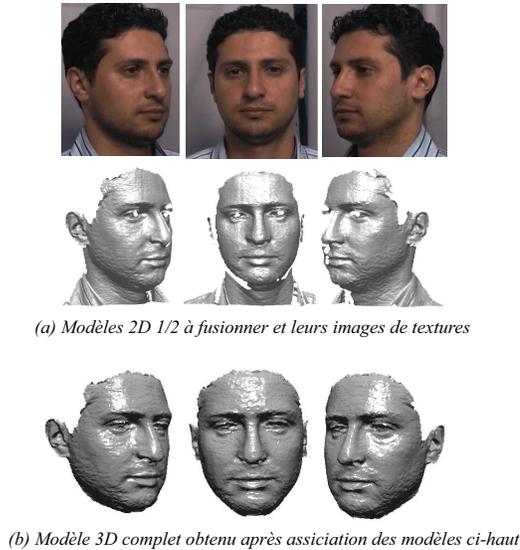


FIG. 5.2 – Association de modèles partiels en vue d'obtenir un modèle 3D complet.

Le deuxième ensemble que considère de notre protocole d'évaluation est formé de modèles de test (*probes*) au format 3D incomplet (c'est à dire 2 1/2D). C'est des clichés pris par le biais d'un scanner laser 3D de type *Minolta VI300*. Cet ensemble, *de test*, comporte pour chaque sujet 8 modèles partiels pris avec des variations de conditions d'éclairage, d'expressions faciales et de pose comme illustre la figure 5.3. Il comporte un sous-ensemble de 5 expressions faciales représentant la surprise, le dégoût, le bonheur, une expression neutre et une dernière avec les yeux fermés, en suite, trois poses différentes, une de face et deux de profils gauche et droite trois quart ($40-60^\circ$), enfin, deux conditions d'éclairage une contrôlée (par l'emploi d'une source de lumière continue pour l'éclairage) et une deuxième non-contrôlée toutes les deux d'intérieure. Ces 8 modèles par sujet seront utilisés exclusivement pour la phase de test, c'est à dire lorsqu'on simule l'accès d'un client ou d'un imposteur en mettant en correspondance l'image de test et l'image de référence relative à la personne prétendu être. Ce deuxième ensemble est complètement disjoint du premier.

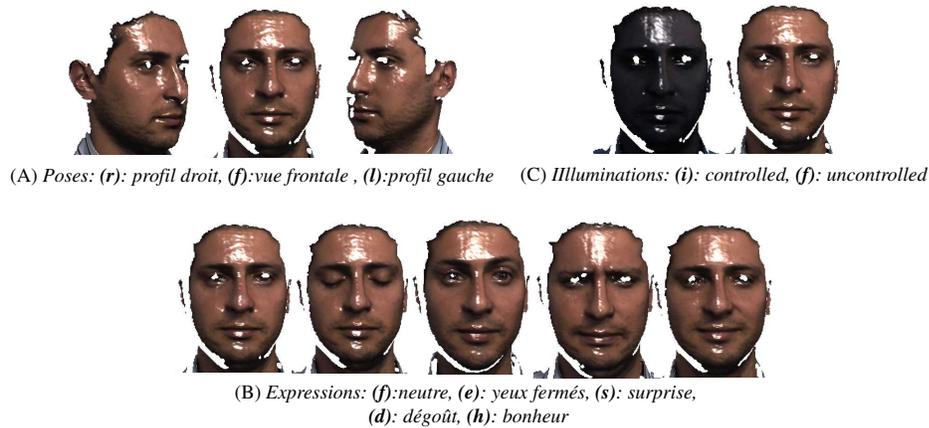
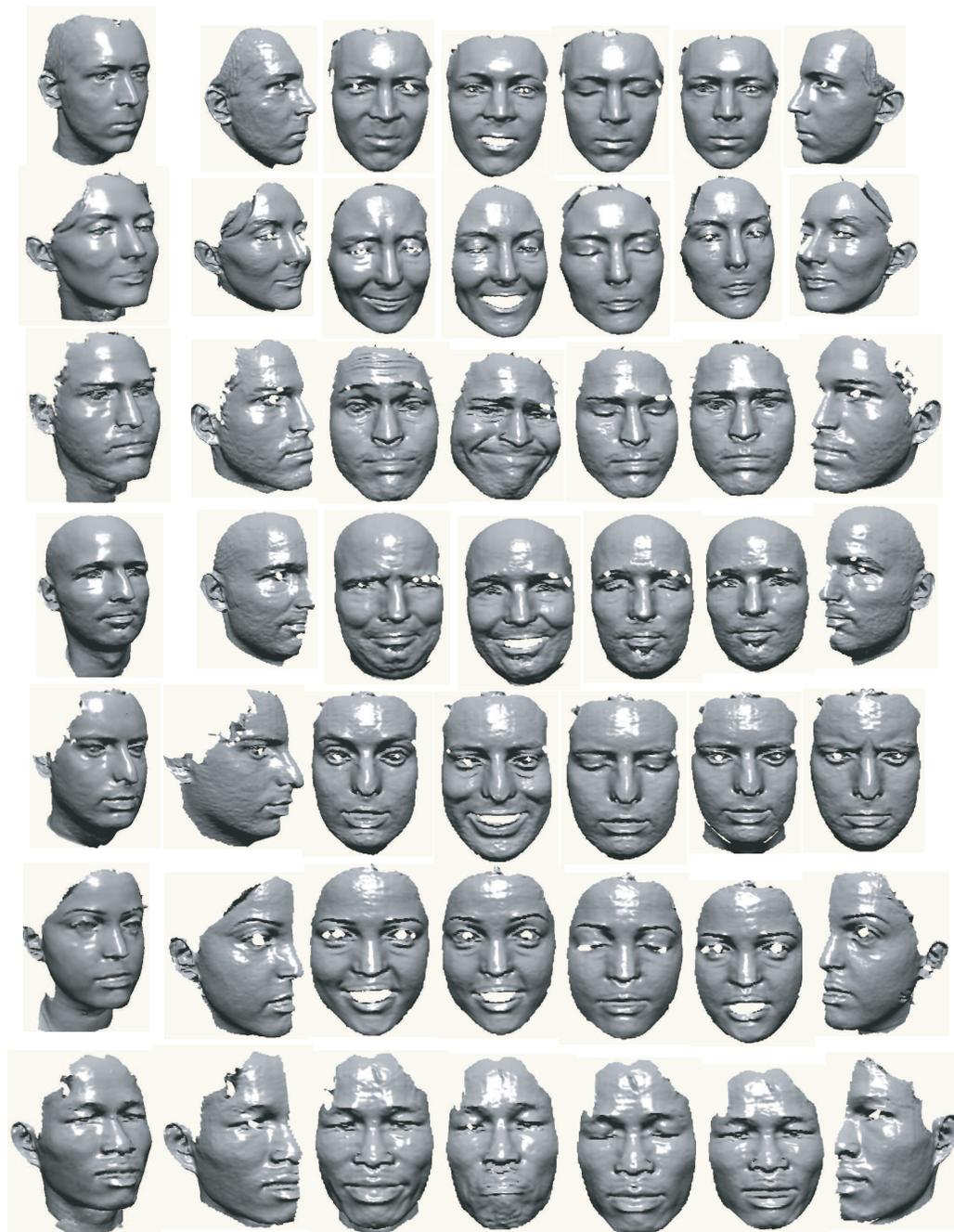


FIG. 5.3 – Différentes variations dans les modèles de test : (A) poses, (B) expressions et (C) illumination.

La figure 5.5 présente un extrait de notre base de visage *ECL-IV²*, la première colonne, dans cette figure, illustre les modèles 3D après fusion. Le reste des colonnes présente des modèles partiels servant comme images requêtes. Ces dernières présentent différentes variations comme mentionné plus haut. Toutes les données de la base sont au format *VRML* (*Virtual Reality Modeling Language*), c'est des modèles qui comportent autour de 7000 points 3D (*vertices*) maillées entre eux pour interpoler la forme 3D du visage. La résolution choisie est suffisante pour conserver les détails du visage avec un minimum de points 3D, en adoptant une technique de remaillage adaptatif. Cette dernière est basée sur le calcul de la valeur de courbure en chaque point du maillage et de fournir un remaillage adaptatif selon les variations des régions de la forme 3D. En réalité, c'est un moyen très efficace pour garder le maximum de détails du visage par régénération d'une bonne densité de points dans les régions de hautes variations (grandes valeurs positives et négatives de courbures) et une faible densité dans les régions homogènes. Une opération de segmentation de la partie faciale, considérée comme la partie informative, précède cette opération.

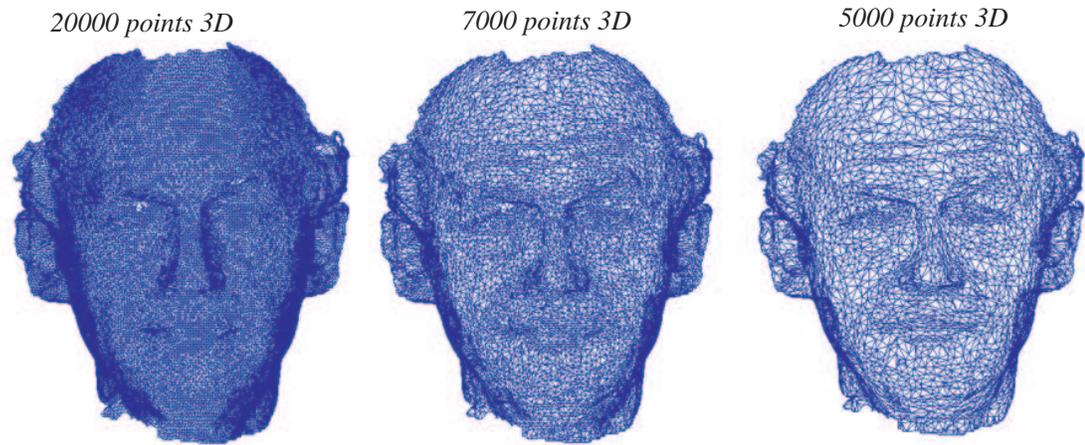
A ces ensembles de données d'évaluation, nous associons un ensemble de fichiers de méta-données à des fins d'organisation. Son but est d'inclure les identités des individus, les sessions correspondantes, les conditions d'acquisition (éclairage et pose) et les variations intrinsèques du visage en terme d'expressions faciales. Ces fichiers au format XML (*eXtensible Markup Language*) contiennent aussi les coordonnées 3D des points d'intérêts sélectionnés manuellement sur la totalité de l'ensemble des modèles 2 1/2D et 3D. L'ensemble de ces fichiers de données et de méta-données sont organisés et gérés par des structures XML. Ces structures définissent l'emplacement des fichiers et servent comme entrée pour l'environnement d'évaluation mis en place pour le calcul des matrices de similarités entre probes (images de tests) et galerie (image de la base). Cette base de données a évolué durant l'année scolaire 2005/2006 de 50 individus ce qui correspond à 50 modèles 3D et 400 images



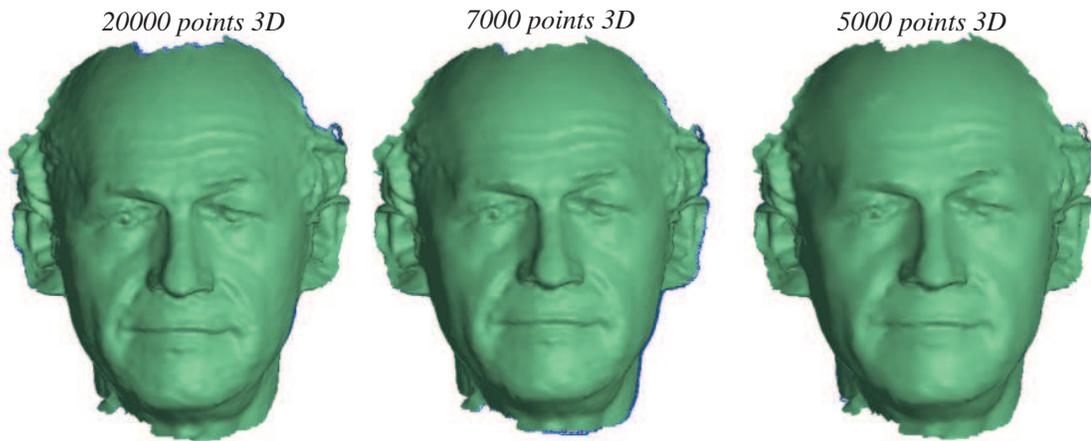
Modèles complets 3D

Modèles partiels 2D 1/2 avec variations

FIG. 5.4 – Extrait de notre base d'évaluation *ECL-IV*² : Modèles complets 3D de la BD, modèles partiel de test.



(a) Maillages 3D obtenus par maillage adaptatif



(b) Formes 3D correspondantes

FIG. 5.5 – Différentes résolutions du visage 3D obtenues par remaillage adaptatif et formes 3D correspondantes.

de test, collectés pendant le workshop résidentiel *Biosecure* 2005, à 120 individu et 1000 images de test pendant la période d'installation de la cabine d'acquisition à l'*ECL*. Ce chiffre doit doubler dans quelques mois avec le déplacement du matériel à l'*université d'evry* puis à l'*INT* dans le cadre du projet technovision national *IV*². Les expérimentations que nous présentons dans ce rapport sont relatives au premier ensemble de données acquis à *Biosecure* 2005 et qui contient 50 modèles 3D complets et 400 images de tests.

5.2.2 Comparaison avec la base *FRGC*

Avec le développement des approches de reconnaissance faciales 3D, plusieurs bases de données ont été développées pour la recherche et pour l'évaluation. Dans ce paragraphe, nous énumérons quelques unes et nous achevons par une comparaison de notre base *ECL-IV*² avec celle de *FRGC*.

- **La base 3D_RMA** : cette base a été créée au sein du SIC³ (Bruxelle-Belgique). Un prototype de vision active (*structured light-based*) est utilisé. Elle contient les images de *120 individus* sous formes de points 3D non-maillé.
- **La base GavabDB** : cette base contient *427 images 3D* correspondantes à *61 sujets* (45 hommes et 16 femmes) avec 7 différentes images par personne. Des variations systématiques de poses et d'expressions faciales sont incluses dans cette base⁴.
- **La base de York University** :⁵ cette base contient des images qui correspondent à *97 individus*, elle contient *10 captures par personnes* avec différentes poses. 2 vues de cet ensemble présente deux petites expressions faciales et une vue avec une occlusion.
- **La base XM2VTS** :⁶ cette base multi-modale développée au sein du projet européen ACTS, contient des images fixes, des séquences vidéo et des images faciales 3D de *295 personnes*. Un capteur stéréo actif a été utilisé pour l'acquisition et les données sont converties en *VRML* incluant les deux canaux de formes et de texture.
- **La base de FRGC** : Elle contient des acquisitions frontales 3D de *466 individus*, pour chaque individu il y a entre 1 et 22 images de profondeur pour différentes sessions. En total, cette base renferme *4007 scans* 3D frontales avec des conditions d'éclairage contrôlées.

Le tableau suivant donne quelques points de comparaison entre la base de données développée au sein du projet *FRGC* et la base *ECL-IV*². Il est à noter que la base *FRGC* est plus riche en données et ce parce que les acquisitions ont débuté pendant l'année scolaire 2002-2003, alors que pour la base *ECL-IV*² les premières acquisitions ont commencé en août 2005. EN outre, la base *ECL-IV*² comporte dans sa première version une seule session,

³<http://www.sic.rma.ac.be/>

⁴<http://gavab.escet.urjc.es/articulos/GavabDB.pdf>

⁵<http://www-users.cs.york.ac.uk/~tomh/3DFaceDatabase.html>

⁶<http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/>

néanmoins elle englobe huit différentes acquisitions. A la différence de la base *FRGC*, la base *ECL-IV²* comporte des variations significatives d'expressions faciales et deux conditions d'éclairage contrôlée/incontrôlée. Ces variations systématiques permettent d'élaborer des sous expérimentations cohérentes caractérisant chaque type de variations.

Comparaison des données 3D	<i>FRGCv2</i>	<i>ECL-IV² V1</i>
<i>3D complet</i>	▼ non disponible	▲ disponible
<i>Multi-vues (poses)</i>	▼ (1 pose frontale)	▲ (3 poses)
<i>Illumination</i>	▼ contrôlée	▲ contrôlée/uncontrôlée
<i>Expressions</i>	▼ 2 expressions	▲ 5 expressions
<i>2D et 3D</i>	▲ disponibles	▲ disponibles
<i>Taille</i>	▲ 4000 scans	▼ 500 scans
<i>Materiel</i>	VI910 (640×480)	VI300 (400×400)

5.2.3 Scénarios des expérimentations

Comme mentionné dans la section précédente, le but étant l'évaluation de cette nouvelle modalité biométrique au sein du paradigme proposé sur une base de donnée de taille significative. Il s'agit donc d'une évaluation de type technologique (*technology evaluation*) qui vise l'évaluation d'un ou plusieurs algorithmes sur une base de données unifiée de la même population et avec les mêmes conditions environnementales d'acquisition. Quant au calcul de scores de similarités, il s'effectue en off-line en produisant une ou plusieurs matrices de similarités selon le test effectué. Deux scénarios d'utilisation possibles sont proposés dans ce protocole et ont été testés : (a) l'identification, (b) l'authentification. Rappelons que l'identification est le procédé d'identifier une personne dans une base de données de plusieurs personnes, tant dis que l'authentification représente la vérification d'une identité. Ces deux scénarios correspondent, respectivement, à deux questions différentes qui sont :

- **Vérification** : Suis-je celui que je déclare être ? L'utilisateur présente son visage et une identité. L'algorithme de reconnaissance évalue la similarité entre le visage de l'utilisateur et celui de la base de données associé à l'identité fournie. Ce type de reconnaissance est aussi appelé "*one-to-one*" : le visage n'est comparé qu'à un seul visage (celui associé à l'identité annoncée).
- **Identification** : Qui suis-je ? L'utilisateur présente son visage. L'algorithme cherche dans sa base de données un visage semblable, et renvoie l'identité associée. Ce type de reconnaissance est aussi appelé "*one-to-many*" : le visage est comparé à tous les visages de la base (galerie).

Les expérimentations que présente notre protocole se déroulent de deux façons différentes :

- **Expérimentation globale** : Cette expérimentation concerne toutes les images de la base de test, abstraction faite du type de variation que présente le modèle de test.

Elle fournit, par conséquent, des taux et des courbes caractéristiques d'ordre global. L'avantage étant la taille significative (400) des images de test.

- **Expérimentations élémentaires** : Ces expérimentations concerne, chacune d'entre elles, un type de variation particulier. 8 expérimentation élémentaires constituent l'ensemble, incluant des changements de posture, d'illuminations et de conditions d'éclairage. L'objectif de ces expérimentation et de voir le comportement de chaque algorithme par rapport au type de variation considérée. Le corpus de test pour chaque expérimentation inclut 50 images de test, présentant toute la même émotion faciale ou le même type d'éclairage employé ou encore le même angle de vue.

5.2.4 Critères de performances

Pour chaque scénario d'utilisation, il existe critères qui caractérisent les performances des algorithmes développés. La production de ces critère nécessite le calcul d'une matrice de similarité commune dans laquelle les images de test indexe les lignes et les images de galerie indexent les colonnes ou inversement. Chaque case de cette matrice correspond à la mesure de similarité (degré de correspondance) entre le modèle de test et le modèle de la galerie. Le protocole d'évaluation fourni un masque de même taille que cette matrice et qui renseigne sur les "match scores" et les "non-matches scores" dont voici la définition : Un *match score* est quand l'image de test et l'image de galerie représentent le même individu alors que les *non-match scores* correspondent au cas contraire. La matrice de similarité contient donc tout les résultats de l'application d'un algorithme donné sur une paire d'ensemble d'images de test et de galerie. Une fois produite, différents taux et courbes caractéristiques peuvent être calculés :

- **False Acceptance Rate ou FAR** : Le taux de fausse acceptation est la proportion d'imposteurs ayant réussi à usurper l'identité d'un client.
- **False Rejection Rate ou FRR** : le taux de faux rejet est le proportion des clients rejetés par le système.

Ces mesures sont intimement liées à la valeur du seuil d'acceptation k . Pour rendre cette dépendance plus explicite, la fausse acceptation et le faux rejet peuvent être écrits sous formes de fonctions $FA(k)$ et $FR(k)$. de par leur définition, il paraît que $FA(k)$ ne peut être qu'une fonction monotone croissante et $FR(k)$ une fonction monotone décroissante. Les critères de performances choisies sont alors les suivants :

- **La courbe DET (Error trade-off curve)** : Cette courbe caractéristique donne pour chaque valeur de FA la valeur de FR qui lui est associée. Elle est obtenu en faisant varier continuellement le seuil d'acceptation k et en traçant l'ensemble des couples $(FA(k), FR(k))$. Cette courbe fournit de façon graphique un aperçu de tous les compromis FAR/FRR possibles et permet de sélectionner un seuil k adéquat selon l'application envisagée. Un algorithme sera d'autant meilleure que sa courbe DET sera proche de l'origine du repère.

- **Error Equal Rate ou ERR** : Ce taux correspond au seuil k' tel que $FA(k') = FR(k') = EER$. ce taux, lui même, résume assez bien les performances que l'on peut attendre d'un algorithme.
- **Successful Rate ou SR** : Ce taux est équivalent à la valeur maximale de $1 - FA(k) - FR(k)$ et qui fournit une idée des meilleures performances que peut attendre le système globalement.
- **Half Total Error Rate ou HTER** : Ce taux est équivalent à la moyenne de $FA(k)$ et $FR(k)$. C'est une valeur qui résume également les performances d'un algorithme en un seuil donné.
- **FRR en un point choisi de FAR** : C'est la valeur du FRR observée lorsque l'on limite le taux de fausse acceptation à 0,1% par exemple. A cette valeur, il donne une idée des performances du système lorsque l'application envisagée requiert un taux d'imposture relativement faible (1/1000).
- **Rank-one Recognition Rate ou RR** : C'est le taux d'identification fournit par un algorithme, en se limitant au premier rang dans la liste triée des personnes.

Pour nos évaluations nous affichons la quasi-totalité de ces taux et de ces courbes couplés avec des interprétations et des discussions sur les performances obtenus. Ces évaluation sont suivies par une comparaison des différents algorithmes pour les deux scénarios d'identification et de reconnaissance.

5.3 Résultats obtenus et discussions

Nous évaluons dans cette partie trois algorithmes de reconnaissance/authentification par visage 3D. Il sont tous basés sur le paradigme de matching 2 1/2D vs. 3D décrit dans le chapitre 3. Il s'agit dans la première approche d'appliquer l'algorithme *ICP* pour le recalage des surfaces faciales. La mesure de similarité est donnée par une métrique calculant l'erreur résiduelle moyenne entre les surfaces recalées. Dans le deuxième algorithme, nous appliquons notre variante de l'algorithme *ICP*, *R-ICP*, basé sur un schéma de segmentation du visage 3D en régions statiques et non-statiques. Une nouvelle métrique est également introduite en calculons une somme pondérée des erreurs résiduelles issues de chaque région du visage selon de schéma précédent. Ici, un banc de poids de pondération ont été testé en vue de voir le comportement des résultats en fonction de cette variation. Enfin, nous comparons ces deux algorithmes par rapport à un troisième basé sur le calcul de cartes de distances géodésique sur les surfaces recalées des visages. Comme décrit plus haut, pour chaque évaluation d'un algorithme, nous considérons deux ensembles : (a) l'ensemble de la galerie (*gallery dataset*) incluant les modèles complets des visages et (b) l'ensemble de test (*probe dataset*) incluant les modèles partiels (2 1/2D). Nous calculons par la suite la matrice de similarité basée sur chacun des algorithmes de matching.

5.3.1 Algorithme basé sur *ICP*

Cette section caractérise l'ensemble des résultats obtenu suite à l'application de l'algorithme *ICP* pour le recalage puis le calcul de la déviation spatiale entre les nuages de points 3D. Une première étape d'alignement grossier entre les modèles test et les modèles de la galerie est effectuée moyennant des points d'intérêts localisées sur les visages. Une étape d'alignement fin suivra la première étape, c'est l'application de l'algorithme *ICP* en vue de minimiser l'erreur de déviation entre les modèles à matcher. Nous retenons comme mesure de dissimilarité la déviation moyenne résultante de calcul de déviation point à point appariées moyennant *ICP*.

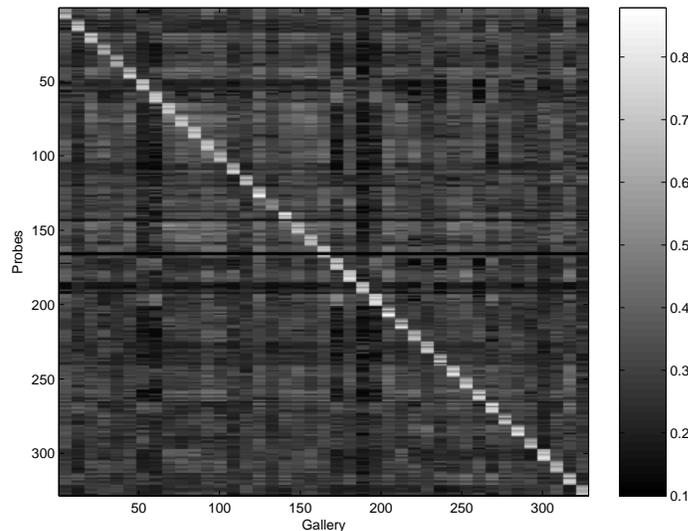


FIG. 5.6 – Matrice de scores de similarité calculées entre les images de la galerie et les images de test.

La figure 5.6 illustre avec une carte en niveaux de gris les mesures de similarité calculé moyennant ce premier algorithme. Les niveau de gris clair correspondent aux mesures de similarité élevé, alors que les niveaux de gris sombre caractérisent les faibles scores. Il est clair que ces valeurs sont situé quasiment en dehors des termes diagonales contrairement aux premières valeurs.

La figure 5.7 présente les taux d'identification RR (*Rank-one recognition rate*) pour une expérimentation globale et les expérimentations élémentaires. Elle affiche un taux de reconnaissance global au tour de $RR_g = 97,25\%$. Pour des images de test avec des changements d'éclairage nous obtenons des bonnes performances ($RR_i = 97,56\%$, $RR_f = 100\%$). Pour des images de test présentant des changements d'angle de posture, les expérimentations donnent un taux de $97,56\%$ ($RR_t = RR_r = 97,56\%$). Pour le reste des expérimentations les taux sont moins bons vu les déformations subies par la surface faciale pendant une expression faciale. Quant'à l'expression de dégoût un taux d'identification de $RR_d = 92,68\%$ est

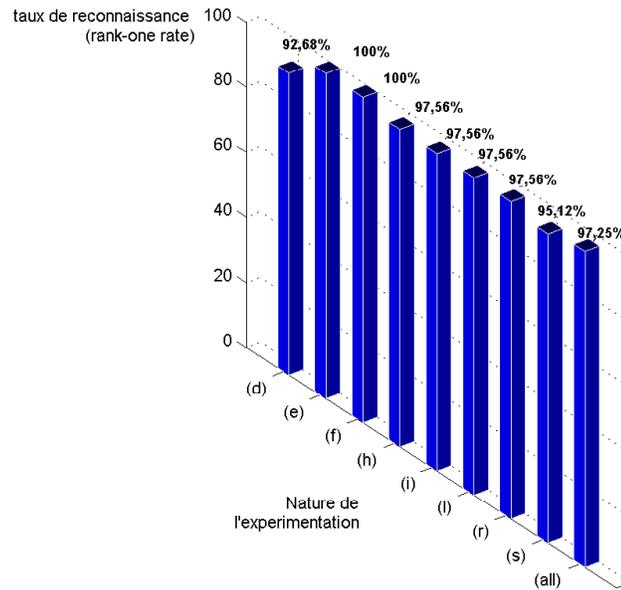
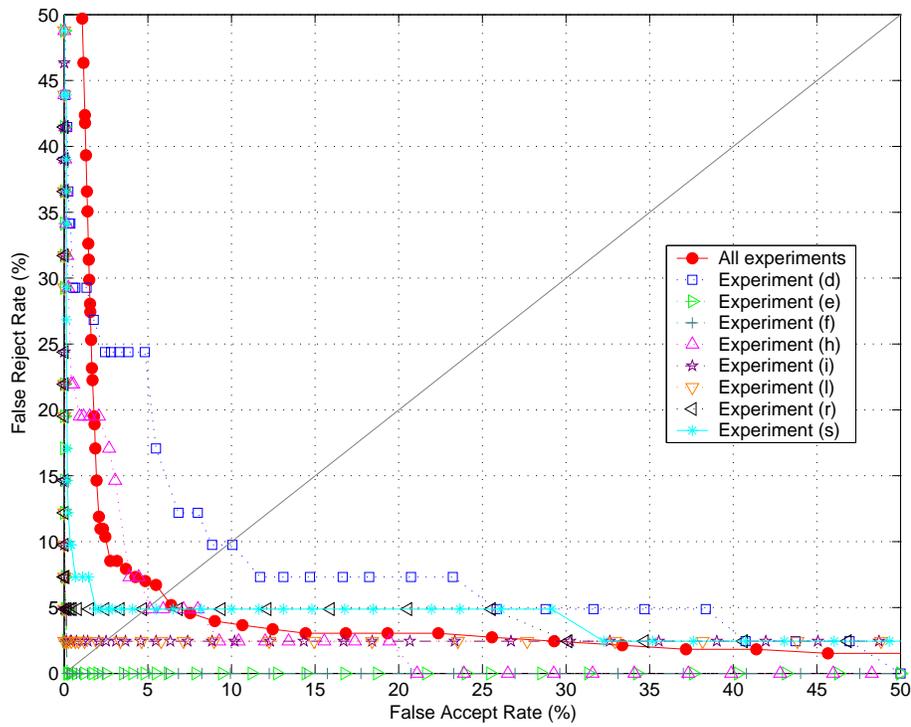


FIG. 5.7 – Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'expérimentation effectuée.

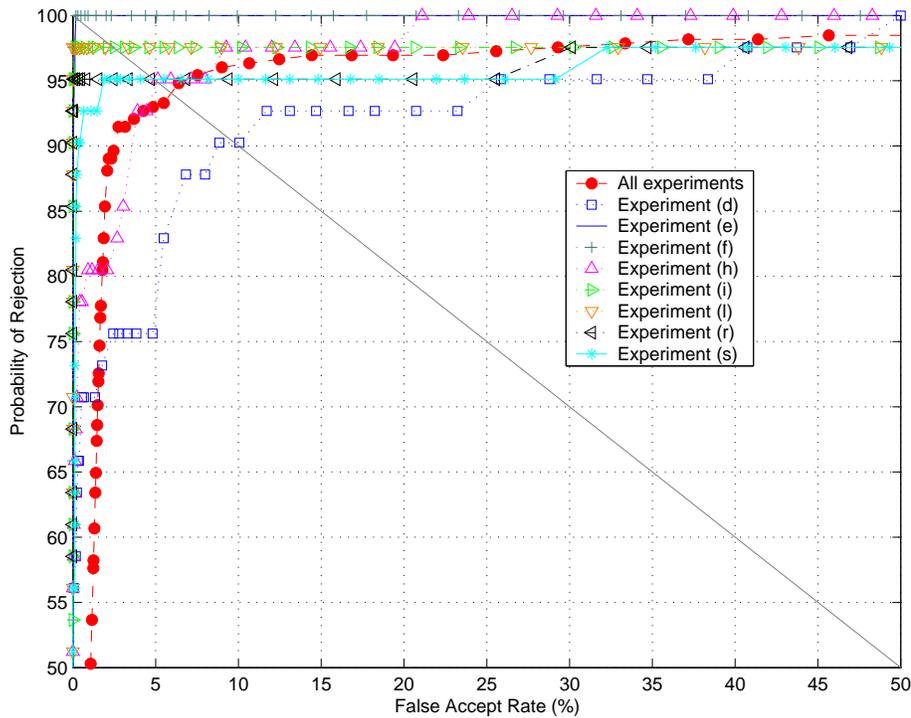
résultant, $RR_s = 95,12\%$ présente le taux pour une expression de surprise, l'expression de bonheur affiche un taux de $RR_h = 97,56\%$ et finalement une expression minimale en 3D (yeux fermé) n'affecte pas l'algorithme et un taux d'identification $RR_e = 100\%$ est affiché.

La figure 5.8 présente les deux courbes caractéristiques *DET* (*Error trade-off*) et *ROC* (*Receiver operating characteristic*) des expérimentations. Ces courbes sont tracées, également, pour chaque type de variation de notre protocole ainsi que pour l'ensemble. Un taux de $ERR_g = 6,5\%$ est obtenu pour l'expérimentation globale (toutes les images probes). Les taux $ERR_{elementaires}$ dépendent du type de variation testé. Par ordre décroissant nous avons obtenu : $ERR_f = ERR_e = 0\%$, $ERR_i = 2,5\%$, $ERR_l = ERR_r = ERR_h = 5\%$ et $ERR_d = 10\%$.

Ces critères de performances, taux d'identification, courbes *DET* et taux de *ERR*, nous mènent aux conclusions suivantes : (a) l'algorithme *ICP* employé au sein du paradigme proposé est quasi-invariant aux variations dues aux changements d'illumination et de pose. Ceci confirme nos hypothèses de départ pour l'utilisation de cette nouvelle modalité. Néanmoins, (b) cette approche est sensible aux changements d'expressions faciales. Nous pouvons en déduire, également, que les performances décroissent en fonction de la gravité de l'expressions faciales présentée comme requête au système. En effet, des petites variations 3D sur la surface faciale ont des effets négligeables sur la mise en correspondance rigide de celle-ci par rapport à une surface dans un état neutre. Les expressions faciales présentent alors, d'après ces expérimentations, le vrai challenge de cette nouvelle modalité biométrique. Ces



(a)



(b)

FIG. 5.8 – (a) Courbes DET (Error trade-off), (b) Courbes ROC (Receiver operating characteristic), pour toutes les images de test et pour chaque test individuellement.

expérimentations nous ont conduit à concevoir une nouvelle approche basée sur l'algorithme *ICP*, *R-ICP* associée à une nouvelle métrique de similarité, et qui prend en compte sa sensibilité par rapport aux expressions faciales.

5.3.2 Algorithme basé sur notre variante *R-ICP*

Dans les évaluations que présente cette section, nous visons notre variante de l'algorithme *ICP* (*R-ICP*) décrite dans le chapitre précédent. Elle modifie l'algorithme *ICP* afin de l'adapter au problèmes d'expressions faciales présentés dans la section précédente. En plus, une nouvelle métrique de similarité a été introduite afin de réduire les effets de ces variations. Dans la partie off-line de préparation de la galerie de reconnaissance, les visages 3D complets sont segmentés en deux régions (statique et non-statique) ayant deux label différents. Cette étape est effectuée via la procédure automatique présenté dans le chapitre précédent. Il s'agit par la suite d'attribuer différents poids de pondérations différents pour ces régions dans le but de mettre l'accent sur la région statique (zone haute) et d'atténuer l'effet de la partie mimique (zone basse). Une batterie de jeux de poids a été testée dont voici les valeurs et le principe :

- $\alpha_1 = \alpha_2 = 1$ (**appelé R-ICP global**) : ici les deux régions possèdent le même poids de pondération qui vaut 1 ce qui correspondant à un algorithme *ICP* classique. Les itérations de minimisation de l'erreur résiduelle sont basées sur les points 3D de la partie statique.
- $\alpha_1 = 1, \alpha_2 = 0$ (**appelé ICP statique**) : ici seule la partie statique du visage fournit le score de similarité entre deux visages. Les points 3D de la partie non-statique sont totalement excluent.
- $\alpha_1 = 1, 0 < \alpha_2 < 1$ (**appelé ICP partielle**) : ici la partie statique fournit son score de similarité élémentaire avec un poids égal à l'unité alors que une variation de α_2 entraîne une atténuation des effets des expressions faciales.

L'emploi de plusieurs valeurs du poids de la partie mimique, α_2 , permet d'afficher plusieurs courbes et plusieurs taux ayant comme objectif de voir le comportement de l'approche en faisant varier ce poids de la partie mimique du visage dans la procédure de recalage et la métrique de comparaison. En outre, elle permet de définir une valeur optimale selon le cas d'application. Les figures 5.9 et 5.10 présentent les courbes *DET* et *ROC*, respectivement, en faisant varier la valeur de α_2 dans l'ensemble $\{1, 0.5, 0.25, 0.125, 0.0625, 0\}$. Elles affichent clairement le comportement de l'approche en fonction de ce paramètre. Ces courbes sont relatives à des expérimentations globales, c'est à dire toutes les images de test ont été considérées dans ces évaluations.

La première interprétation de ces courbes est l'apport de la variation du poids de la partie mimique dans le calcul du score de similarité global par une somme pondérée de scores élémentaires. En effet, la courbe verte qui correspond à $\alpha_2 = 0.5$ présente le meilleure

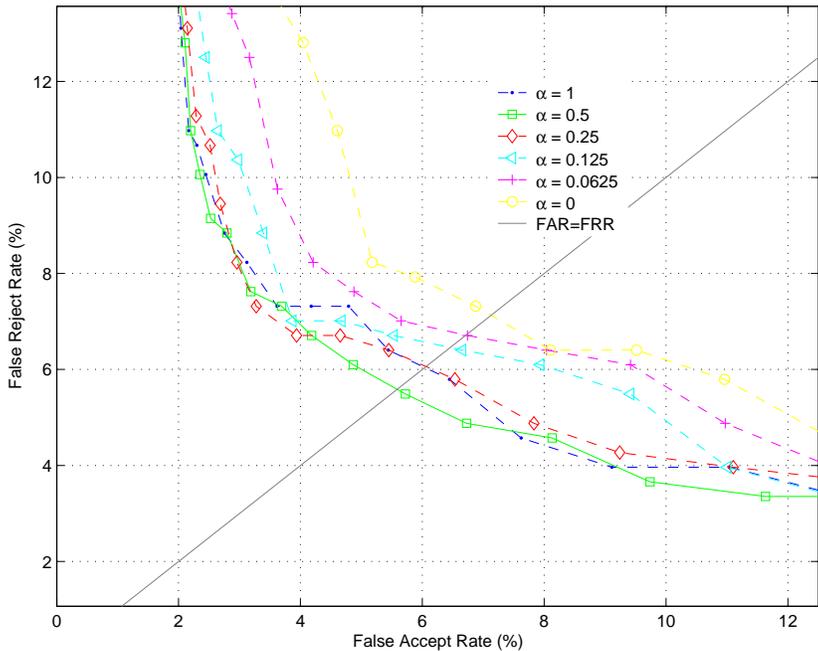
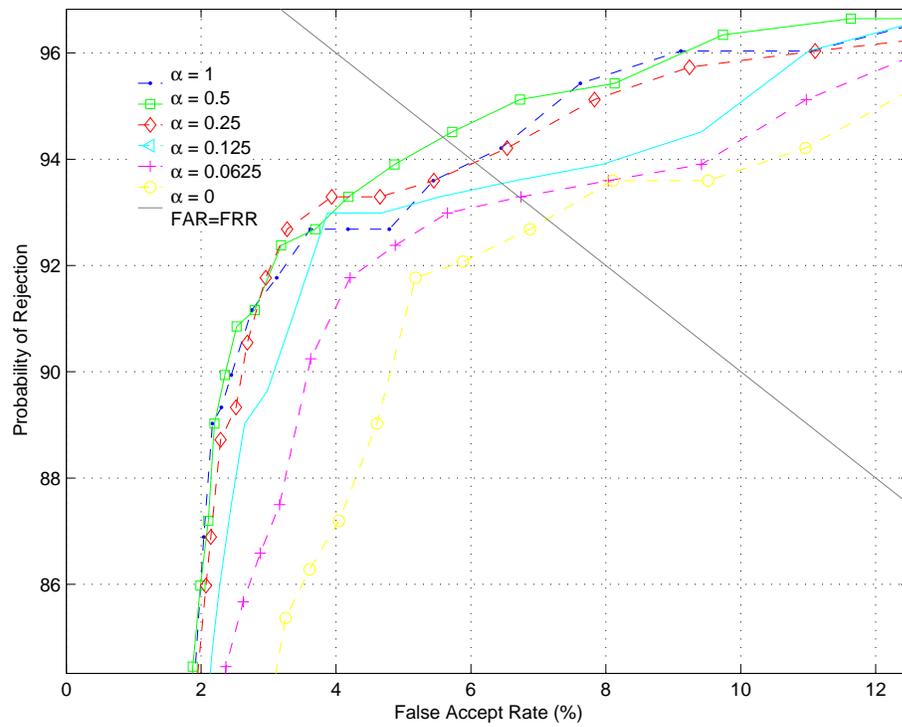


FIG. 5.9 – Courbes DET (Error trade-off) pour toutes les images probes de la BD en variant α .

FIG. 5.10 – Courbes ROC (Receiver operating characteristic) pour tout les probes de la BD en variant α .

résultat sur l'ensemble des courbes du fait qu'elle a tendance de s'approcher de l'origine du repère. Elle présente également la plus petite valeurs du taux $ERR_{0.5} = 5,5\%$ devant un taux de $ERR_1 = 6\%$ pour une valeur de $\alpha_2 = 1$. Outre cette première interprétation, nous pouvons constater que la restriction sur la partie haute du visage $\alpha_2 = 0$ est insuffisante pour avoir de bonnes performances. Ce qui conduit à conclure que la forme 3D de la partie haute du visage est loin d'être, seule, le support de reconnaissance. Dans la suite de ces évaluations et vu que cette nouvelle approche orientée région est destinée pour réduire les effets des expressions faciales, nous présentons dans la figure 5.11 différentes courbes DET pour les différentes variations d'expressions faciales. Pour plus de lisibilité des résultats, un affichage logarithmique est adaptée pour l'axe des FAR , et un affichage normal est affiché pour l'axe des FRR . Les variations considérées dans ces sous-expérimentations sont les expressions de (d) de dégoût, (s) de surprise, (h) de bonheur et (e) des yeux fermés.

Ces courbes confirment les résultats de l'expérimentation globale en faisant varier la valeur α_2 , le poids de la partie non statique du visage. Des interprétations plus avancées peuvent être tirées. En effet, si l'on considère une application de haute sécurité (c'est à dire une application qui requiert un taux d'imposteur relativement faible), cette approche orientée région affichent des résultats plus intéressants. Prenant comme valeur de référence du $FAR = 10^{-1}$, les différentes courbes DET (pour la valeur de $\alpha_2 = 0.5$ affichent un gain maximal au tour de 10% en matière de FRR dans le cas de l'expression (s) et (h) et une valeur minimale de 3% dans le cas de l'expression (d). En conclusion, cette nouvelle approche avec une métrique pondérée est intéressante à introduire dans un contexte de haute sécurité.

Dans la suite des expérimentations, nous considérons la valeur $\alpha_2 = 0.5$ comme valeur optimale. La figure 5.12 présente une carte en niveau de gris de la matrice de similarité calculé pour cette approche. De même que la première matrice calculée pour l'algorithme ICP , les cases sombres correspondent aux petites valeurs de correspondance entre une image probe et une image de la galerie. Les cases claires correspondent aux valeurs de correspondance élevées. Elles sont concentrées sur la diagonale de la matrice calculé.

Dans la figure 5.15, nous listons les taux d'identifications RR (*Rank-one Recognition Rate*) obtenus pour chaque expérimentation élémentaires ainsi que pour l'expérimentation globale. Un taux global de $RR_g = 97.56\%$ est affiché pour l'introduction de toutes les images de test dans cette expérimentation. Un gain de performance, par rapport à l'approche initiale ICP , de 0.31% est obtenu pour le taux d'identification. Ce gain est résultant des gains élémentaires obtenus dans les expérimentations élémentaire d'expressions faciales. En effet, nous passons, par exemple, d'un $RR_d = 92.68\%$ dans le cas d' ICP à un taux $RR_d = 97.56\%$ par la nouvelle variante $R-ICP$.

Dans cette section, nous avons présenté les évaluations relatives à l'algorithme $R-ICP$ associé à une métrique orientée région. Ces évaluations montrent des gains de performances et de précision par rapport au premier algorithme proposé, basé sur ICP . Dans la sections suivante, nous nous intéressons à une approche de calcul géodésique introduite après le matching 2 1/2D vs. 3D via ICP . Cette approche est amplement décrite dans le chapitre 3.

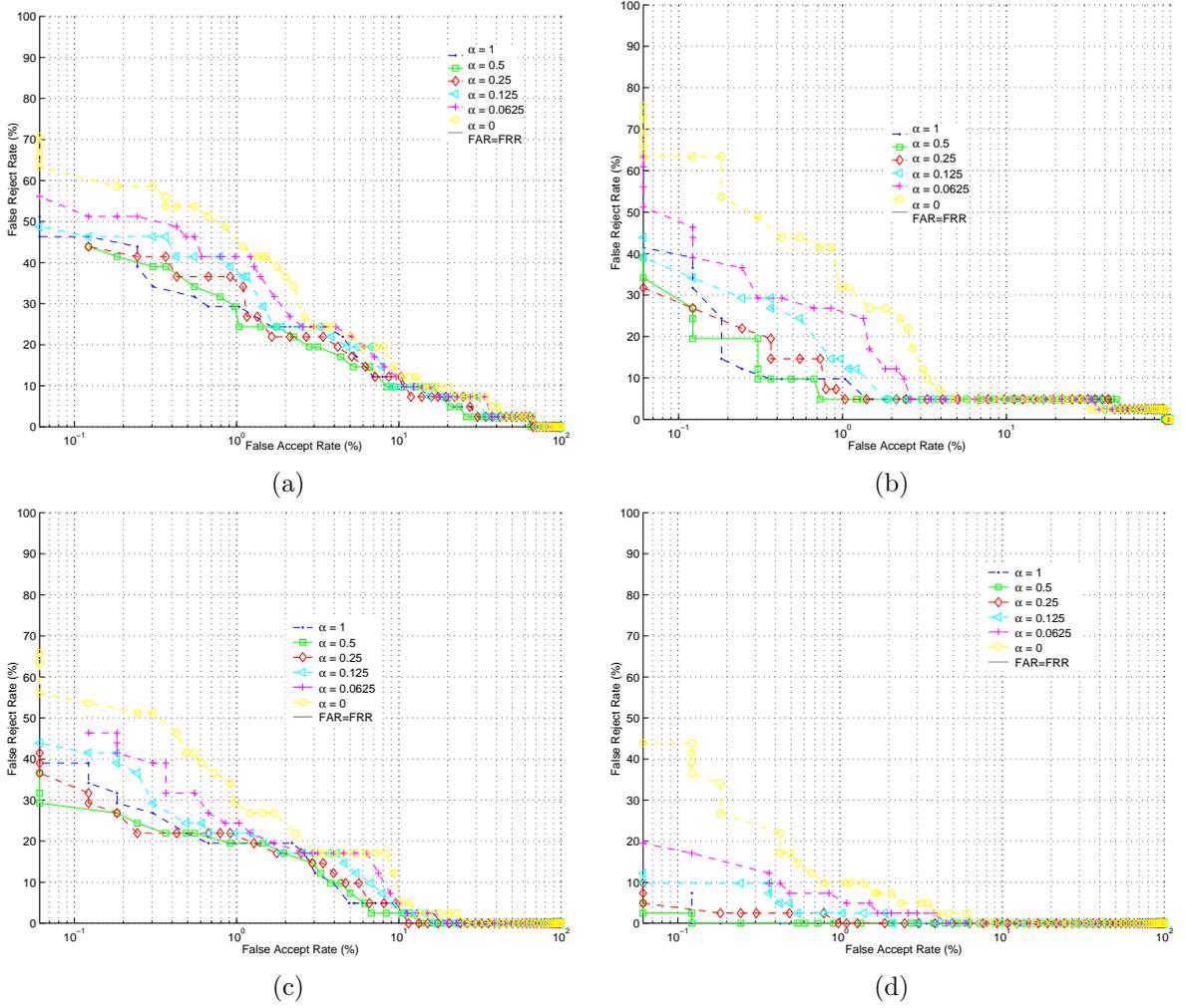


FIG. 5.11 – De gauche à droite et de haut en bas : courbes DET (Error trade-off) pour les images probes avec des expressions faciales (d) : dégoût, (s) : surprise, (h) : heureux et (e) : yeux fermés en variant α .

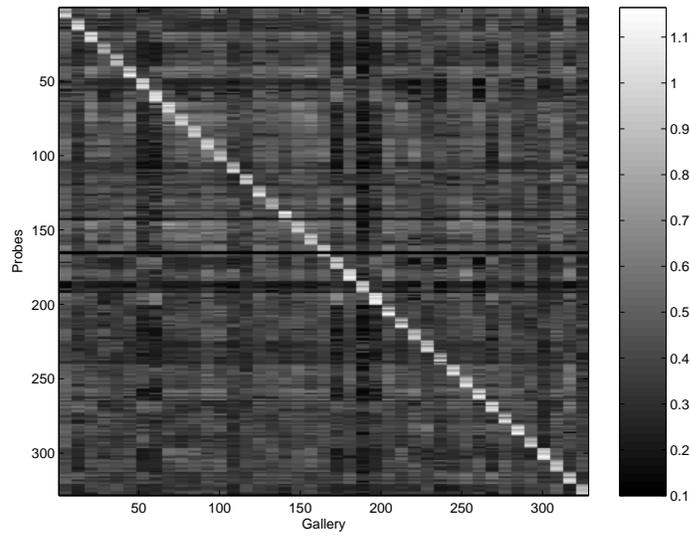


FIG. 5.12 – Matrice de scores de similarité calculées entre les images de la galerie et les images de test.

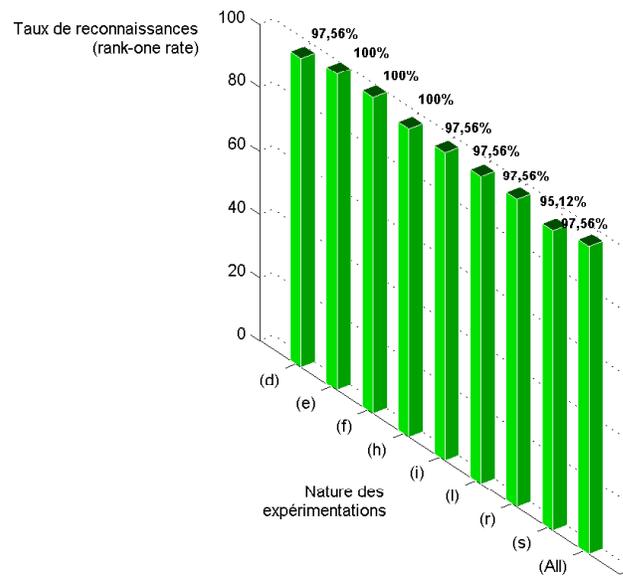


FIG. 5.13 – Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'expérimentation effectuée.

5.3.3 Algorithme basé sur le calcul de cartes géodésiques

Un troisième algorithme a été mis en place et testé dans cette thèse. Cet algorithme part de l'hypothèse suivante : la surface faciale appartient à l'ensemble de surfaces dites *isométriques*. Une surface isométrique est une surface qui préserve les distances tout au long de la surface quelque soit la déformation subie par celle-ci. Dans ce contexte, la distance appliquée dans le cas des surfaces est la distance géodésique qui consiste à relier deux points de la surface en minimisant la distance entre eux tout en restant sur la surface. Par analogie, cette distance peut être comparée à la distance euclidienne dans le contexte d'un plan 2D. Cette hypothèse est le résultat d'une étude empirique de [BBK05] exploitée par la suite pour mettre en place un nouvel algorithme de reconnaissance invariant aux expressions faciales. Les auteurs affichent un résultat très intéressant par conclusion que leur algorithme, basé sur cette hypothèse, sait distinguer entre les jumeaux même en présence d'expressions.

Vu cette caractéristique très importante de la surface faciale, émanée de [BBK05], nous avons choisi de le considérer comme troisième approche de notre étude comparative. Nous avons ainsi mis en place une approche qui exploite le résultat du recalage effectué moyennant *ICP* pour calculer des cartes de distances géodésiques entre les deux nuages de points mises en correspondance. Pour simplifier plus le problème et rendre plus robuste cette approche par rapport à une faiblesse remarquable celle quand la bouche est ouverte, nous désignons sur le modèle 3D de la galerie un point source (point entre les centres des yeux) et un point limite (point en dessous de la base du nez) pour le calcul de la carte de distance. L'objectif de ce choix est de remédier à cette limitation, en effet, quand la bouche est ouverte, l'hypothèse n'est plus vérifiée comme illustre la figure . le chemin minimal empreinte par l'algorithme de cheminement entre les deux points n'est pas le même.

À l'image des deux expérimentations précédentes, la matrice de similarité est produite suite à cette étape de matching de modèles probes et de la galerie et de calcul de cartes de distances géodésiques. La figure 5.14 illustre par une carte en niveaux de gris les valeurs des mesures de similarité entre modèles de test et les modèles 3D de référence.

La figure 5.15 affiche les taux d'identification *RR* (*Rank-one Recognition Rate*) relatifs à chaque expérimentation élémentaire ainsi que pour l'expérimentation globale. Un taux de $RR_g = 92,68\%$ est obtenu pour l'expérimentation globale, par l'introduction de l'ensemble de l'image de test qui est traduit le niveau de performance modeste par rapport aux deux techniques présentées dans les deux sections précédentes, *ICP* et *R-ICP*. Des taux quasi-parfaits sont obtenus pour les sous-expérimentations (f), (i), (l) et (r) autour de 97,56%, ces images de test ne présentent pas de changement d'expressions faciales. Des taux faibles est inattendus sont à l'origine des expressions faciales à l'image de $RR_d = 80,48\%$, $RR_d = 92,68\%$, $RR_h = 90,24\%$ et $RR_s = 92,68\%$.

Les courbes *DET* et *ROC* concernant l'évaluation de cet algorithme confirment les mêmes conclusions. En effet, des taux *ERR* élevés ont été obtenus, comme illustre 5.16.

Malgré que nous avons évité la condition pour laquelle l'hypothèse de départ, celle que le visage est une surface *isométrique*, ne soit vérifiée, les résultats obtenus contredit cette hypothèse. Notre solution, relative à cette condition, consiste à restreindre le calcul de carte

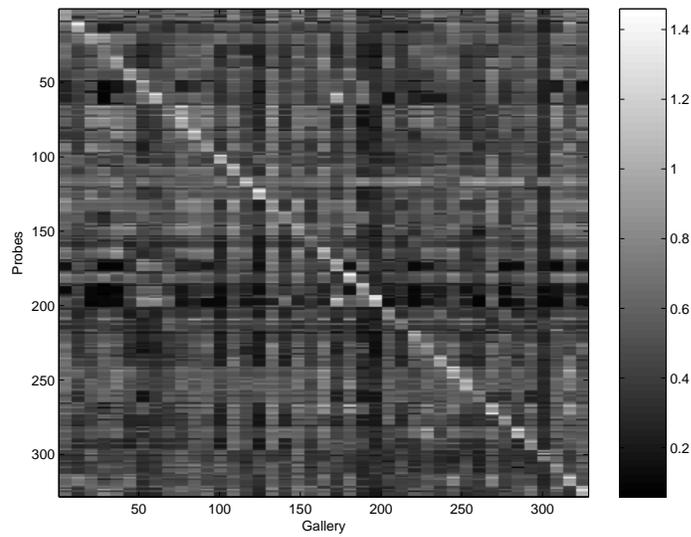


FIG. 5.14 – Matrice de scores de similarité calculées entre les images de la galerie et les images de test.

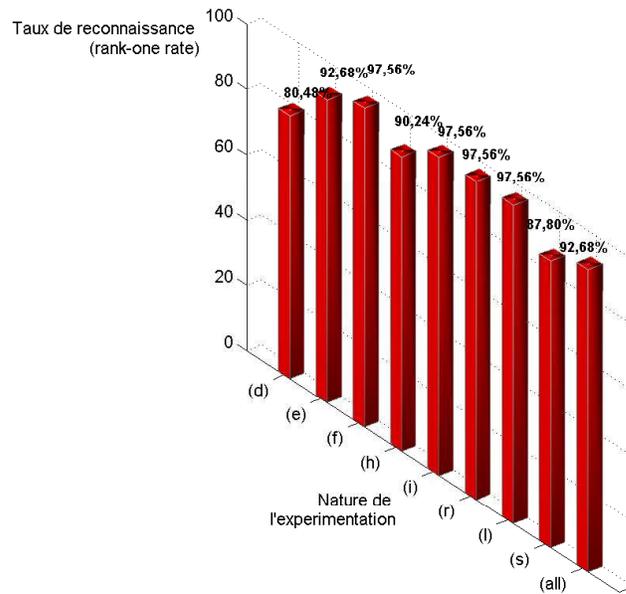
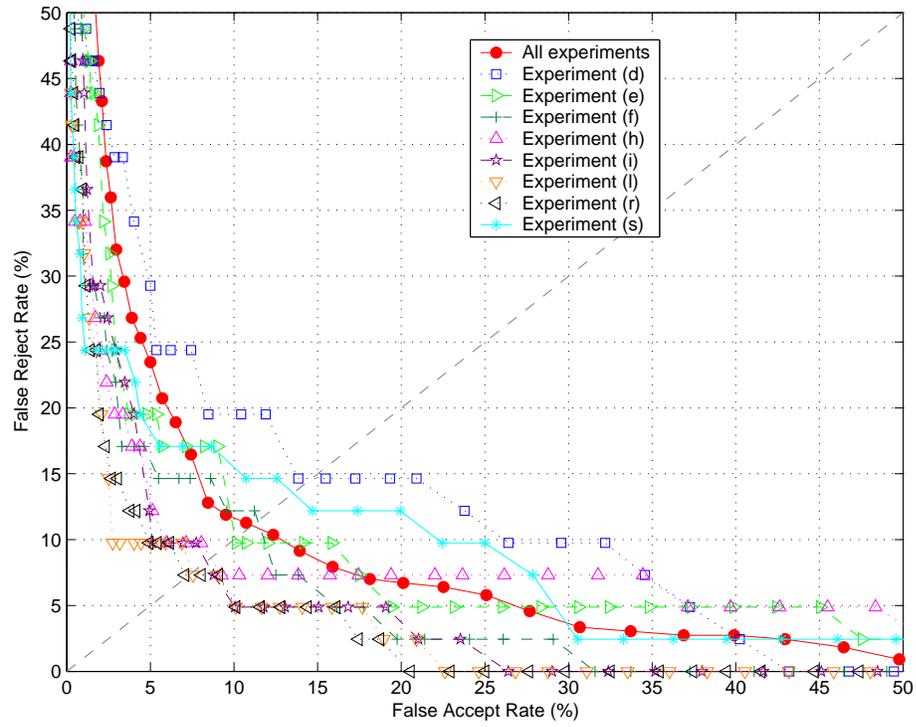
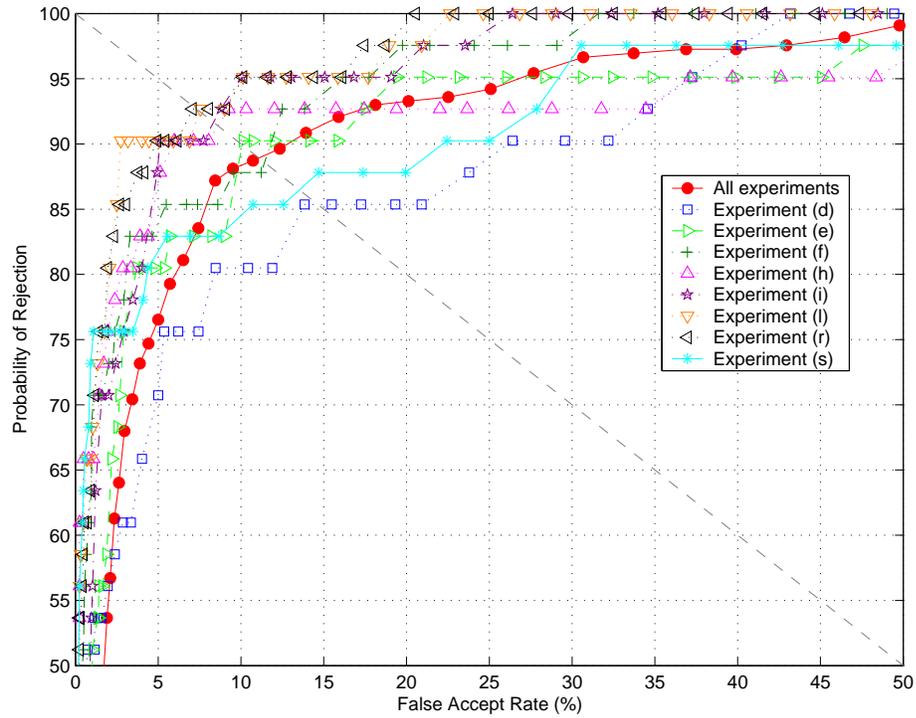


FIG. 5.15 – Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'expérimentation effectuée.



(a)



(b)

FIG. 5.16 – (a) Courbes DET (Error trade-off), (b) Courbes ROC (Receiver operating characteristic), pour toutes les images de test et pour chaque test individuellement.

de distance géodésique entre deux points source et limite afin d'éviter le problème de bouche ouverte dans les images de test. Nous essayons dans ce paragraphe de trouver des explication à ces taux inattendus obtenus par cette approche. La première explication que nous abordons est relative à la nature des maillages des images de test, ces maillages présentent des trous dues à l'acquisition. La présence de ces trous, affectent le calcul de distances géodésiques, par définitions de celles-ci. Pour remédier à ce problème, une interpolation peut être introduite en vue de boucher les trous et par conséquent avoir des distances plus correctes. La deuxième explications que nous apportons est en rapport avec la résolution des maillages utilisés. En effet, les modèles utilisés sont de résolution autour de 7000 points et seul un nombre autour de 5000 points 3D est utilisé pour le calcul de cartes géodésiques. Cette résolution peut être insuffisante pour effectuer correctement le calcul. Une augmentation de la résolution du maillage peut avoir de bon effets sur le calcul de ces cartes de distances.

5.3.4 Comparaison des algorithmes *ICP*, *R-ICP* et cartes géodésiques

Dans cette sections nous souhaitons comparer les trois algorithmes développés dans le chapitre précédent et expérimenté dans les sections précédentes. Sur le même corpus d'images, et avec le même protocole d'évaluation décrits au part avant, nous souhaitons avoir une étude comparative entre ces trois techniques qui se basent toutes sur un paradigme de matching 2 1/2D vs. 3D. La première approche développée est basée sur l'algorithme du plus proche voisin itéré (*ICP*). C'est une technique de matching rigide entre les surfaces faciales et de production de la différence d'épaisseur entre elles sur laquelle est fondée notre métrique de similarité. La deuxième approche est une amélioration de la première, elle est basée sur une nouvelle variante du même algorithme *ICP*. Néanmoins, elle réduit les effets des expressions faciales par l'introduction d'une nouvelle métrique orientée régions. La troisième approche est basée sur la comparaison de cartes de distances géodésiques mesurées sur les visages à comparer. Cette technique est basée sur une hypothèse introduite dans [BBK05], à savoir la propriété *isométrique* de la surface faciale.

Dans cette étude comparative, nous proposons de tracer les courbes présentant les proportions de fausses acceptations et de faux rejets en fonction du seuil d'acceptation k : $FA(k)$ et $FR(k)$. Ces courbes pourrons nous renseigner sur la quantité de l'erreur présenté par chaque classificateur. C'est la zone d'intersection des deux aires engendrées par ces courbes. Les figures 5.17 (a), 5.17 (c) et 5.17 (e) donnent ces courbes caractéristiques des différents classificateurs, respectivement *ICP*, *R-ICP* et *Geodesics*. Nous traçons également pour chaque algorithme les distributions de clients et d'imposteurs en fonction des scores de similarité calculé par chaque algorithmes. Les figures 5.17 (b), 5.17 (d) et 5.17 (f) illustrent ces distributions et renseigne également sur les qualités des algorithmes.

En comparant les figures 5.17 (a), 5.17 (c) et 5.17 (e), et les figures 5.17 (b), 5.17 (d) et 5.17 (f), nous pouvons constater que l'algorithme *R-ICP* engendre moins d'erreur par rapport aux deux autres. C'est les parties communes entre les taux FAR/FRR , d'une part, et les distribution *clients/imposteurs*, présentes, dans les deux types de courbes qui nous renseigne sur cette erreur. Nous pouvons constater également qu'une proportion des vrais

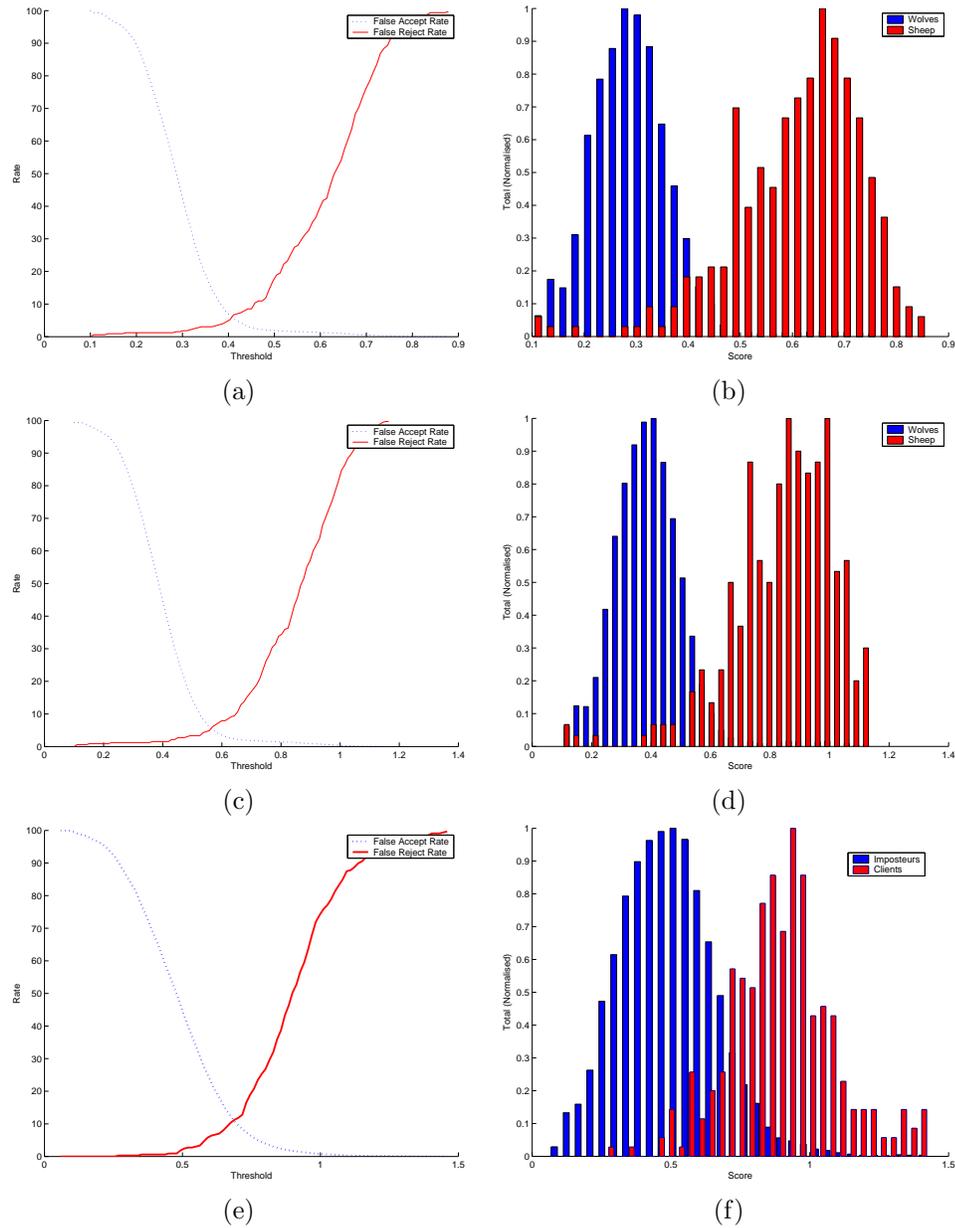


FIG. 5.17 – (a) Courbes FAR/FRR pour l’algorithme ICP, (b) Distributions des *clients/imposteurs* pour l’algorithme ICP, (c) Courbes FAR/FRR pour l’algorithme R-ICP, (d) Distributions des clients/imposteurs pour l’algorithme R-ICP, (e) Courbes FAR/FRR pour l’algorithme Geodesics, (f) Distributions des clients/imposteurs pour l’algorithme Geodesics.

clients présente un score de similarité faible avec leurs modèles 3D correspondants. Ceci est du à l'éloignement par rapport à la solution recherché par recalage rigide via *ICP* ou sa variante *R-ICP*.

Nous donnons, dans la figure 5.18, les taux d'identifications *RR* (*Rank-one recognition Rate*) pour chaque algorithme et pour chaque expérimentation élémentaire et globale. En matière de performance, l'algorithme R-ICP associée à la nouvelle métrique que nous proposons donne les meilleurs résultats de $RR_g = 97,56\%$ devant un $RR_g = 97,25\%$ donnés par *ICP* et $RR_g = 92,68\%$ obtenus via le calcul de cartes géodésiques (*Geodesics*).

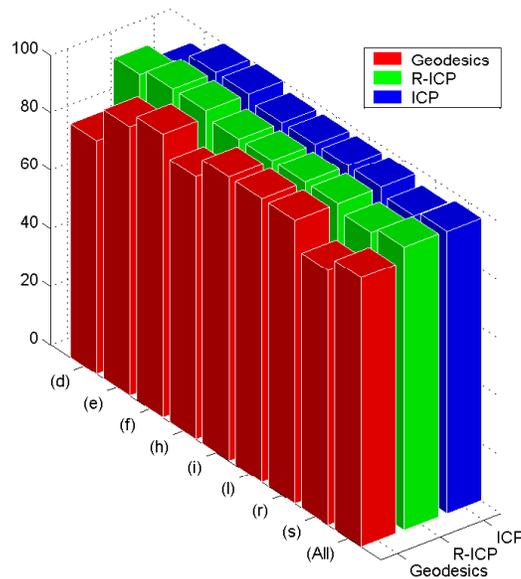
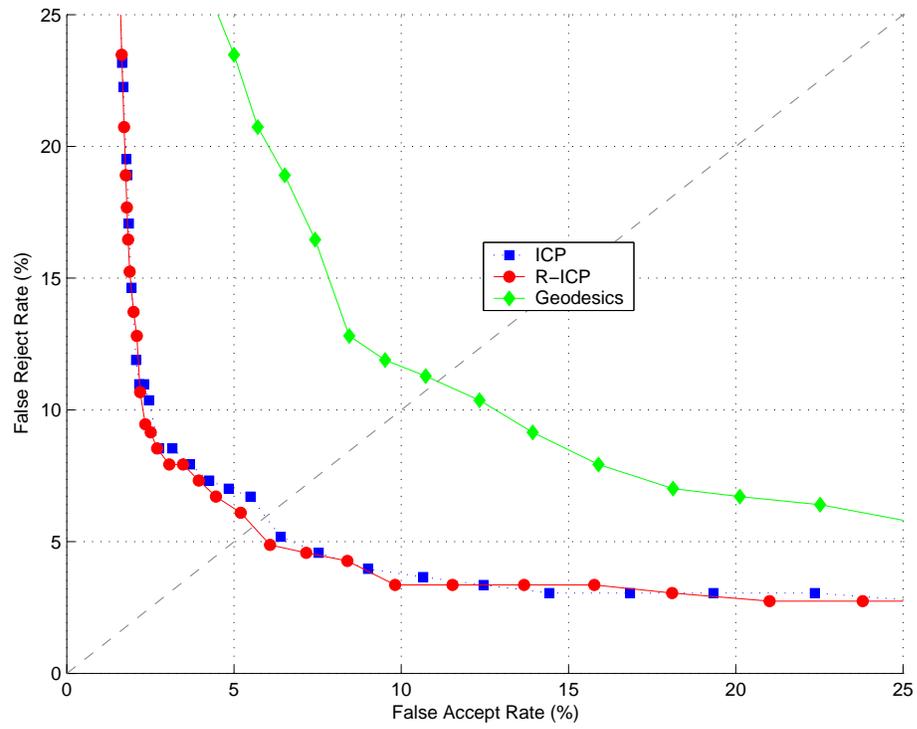


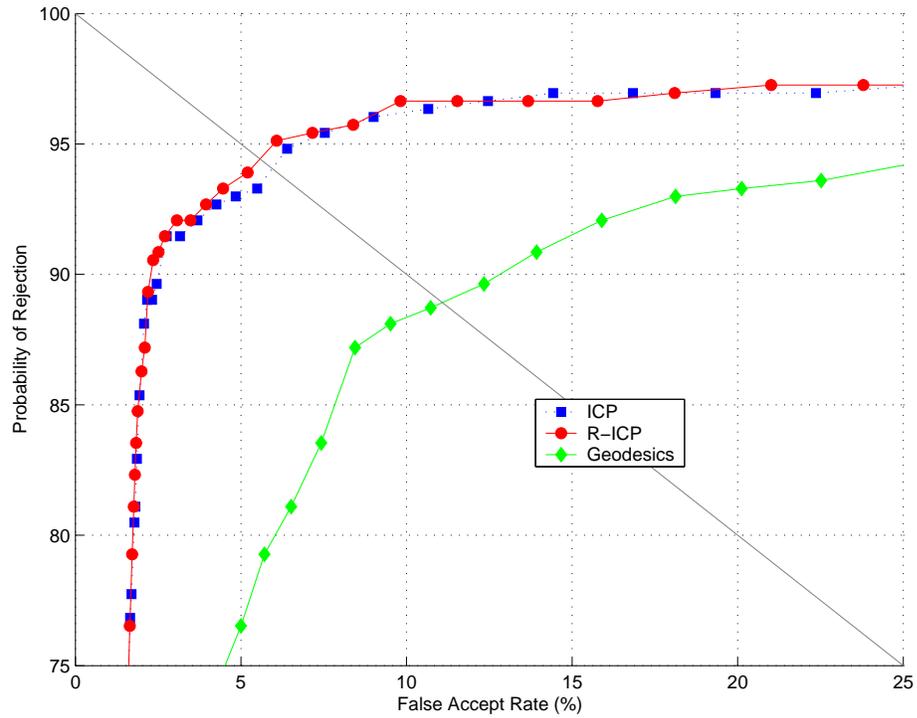
FIG. 5.18 – Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'expérimentation effectuée pour les trois algorithmes ICP, R-ICP et calcul de cartes géodésiques.

La figure 5.19 présente les courbes *DET* pour chaque algorithme tracées dans le même repère. Ces courbes montrent clairement, dans cette étude comparative, l'apport de l'approche région sur l'algorithme *ICP* et l'insuffisance de l'approche géodésique pour son application à nos données de traitement actuelles. En effet, en matière de taux d'erreur pour les quelles les fausses acceptations sont égales aux faux rejets, *EER*, *R-ICP* présente la valeur la plus faible de $EER = 5,5\%$. Plus généralement, la courbe *DET* correspondante à l'algorithme *R-ICP*, présente dans la quasi-totalité des cas un taux succès *TS* plus élevé et un taux *HTER* plus faible.

Dans cette section nous avons présenté une étude comparative concernant trois algorithmes développés durant cette thèse. Les trois algorithmes sont quasi-invariants aux



(a)



(b)

FIG. 5.19 – Courbes DET relatives aux trois algorithmes *ICP*, *R-ICP* et *Geodesics* pour une expérimentation globale.

problèmes de changements de pose ou d'éclairage. Cependant, ils présentent des résultats différents autour du problèmes de changement d'expressions faciales. Premièrement, un premier résultat inattendu de l'introduction de l'approche géodésique a été affiché. En effet, l'hypothèse de départ sur laquelle est fondée cette approche n'est pas vérifiée du moins par rapport à nos données 3D de traitement. Deuxièmement, une approche de recalage rigide entre surfaces faciales et de calcul de la déviation spatiale entre elles a été implémentée et testée. Deux métriques de similarité ont été testés, la première considère le visage comme une entité homogène alors que la deuxième le dissocie en deux zones statiques et non-statiques. Cette segmentation du visage en régions est la sources des amélioration du premier algorithme.

5.4 Conclusion

Nous avons présenté dans ce chapitre une approche de reconnaissance/authentification de personne via une nouvelle modalité biométrique, c'est la forme 3D du visage. Les résultats obtenus sur un corpus de test assez significatif montre l'efficacité de cette modalité pour distinguer entre les individus malgré les variations intervenu sur les images de test tout en respectant le caractère peu intrusif de l'acquisition. Cette caractéristique très importante de cette modalité (visage 2D ou 3D) qui fait que son acceptation par les utilisateurs est immédiate par rapport à d'autres modalités comme l'empreinte digitale, ou le scan de l'iris qui demandent plus de coopérations des utilisateurs.

La solution présentée de reconnaissance/authentification est basée uniquement sur la forme 3D du visage. Une abstraction complète a été faite sur l'apparence à cause de sa sensibilité dramatique aux conditions d'éclairage. Cette modalité reste toujours très informative, qui donne du réalisme aux visage humains et qui sert comme moyen unique pour reconnaître les uns les autres. Nous projetons dans nos travaux futurs l'introduction de cette modalité avec la forme 3D du visage en vue d'une fusion d'informations qui se complète parfaitement. Cette technique multimodale 2D+3D nous permettra d'augmenter la précision de notre système actuel.

Chapitre 6

Conclusion et perspectives

Deux domaines clés de la vision et de la reconnaissance des formes ont été étudiés dans cette thèse, ceux de la numérisation tridimensionnelle et de la reconnaissance faciale. Des solutions aux problèmes ont été données, tout au long de cette thèse, en vue d'une acquisition plus satisfaisante en terme de coûts et de qualité de modèle produit. En reconnaissance, nous avons visé l'invariance par rapport aux limitations dont souffrent les algorithmes existants. Une nouvelle modalité (*3D face*) a été testée dans un paradigme efficace de matching 2.5D vs. 3D. Les résultats obtenus, en acquisition et en reconnaissance 3D, affichent une validation de la démarche scientifique de notre travail. Pour ces deux parties, nous présentons dans cette section un résumé de nos contributions et quelques perspectives que nous envisageons dans nos travaux futurs :

- **En reconstruction 3D de visages en vue de la reconnaissance** : une nouvelle approche hybride d'acquisition tridimensionnelle de visage a été mise au point. Elle associe les techniques de vision active et celles de modélisation géométrique afin d'obtenir des rendus du visage de bonne qualité et avec une bonne précision. En effet, un premier modèle grossier est obtenu par stéréovision active dense sur une seule direction (direction parallèle aux franges de lumières) et semi-dense sur la deuxième direction complètement orthogonale. Un gain de précision dans le calcul de profondeur est apporté moyennant la localisation sous-pixélique de primitives ce qui a affiché une continuité au niveau des reconstructions obtenues. Un modèle raffiné du modèle grossier est obtenu par l'introduction de courbes d'interpolation dédiées dans la direction où la reconstruction est semi-dense. L'importance d'une telle combinaison réside dans la complémentarité dans les reconstructions. Notre méthode est, par conséquent, moins sensible que les approches qui emploient un codage par projection de plusieurs motifs lumineux sur le visage par rapport aux déformations que peut subir ce dernier pendant l'acquisition. En effet, seulement deux modèles de lumière positif et négatif sont successivement envoyés sur le visage dans un laps de temps très réduit. Cependant, plusieurs améliorations peuvent être apportées au schéma proposé. Elles peuvent être de deux types : des améliorations d'ordre pratique nécessaires pour

pouvoir passer d'un système en phase de développement à un système opérationnel, et des améliorations d'ordre plus conceptuel et théorique. D'abord, Pour éviter la périodicité de la lumière projetée, nous pouvons introduire des traits de couleurs sur les franges éclairées à fin d'enlever complètement l'ambiguïté autour du problème de mise en correspondance. En suite, une contrainte très importante peut être introduite au sein de l'algorithme d'appariement afin d'éviter le traitement des lignes épipolaires de manière indépendante. C'est donc la programmation dynamique qui doit être adaptée dans un contexte 2D au lieu du présent 1D. En ce qui concerne les améliorations d'ordre pratiques, nous souhaitons migrer d'un prototype en cours de développement à un prototype figé et opérationnel. Le but est d'intégrer notre plateforme d'acquisition à faible coût dans les évaluations de reconnaissance faciales 3D en vue de profiter des développements effectués dans ce domaine.

- **En identification/authentification de personnes par leurs surfaces faciales 3D** : Les résultats que nous venons de présenter dans le chapitre précédent montrent la robustesse de cette nouvelle modalité biométrique qui est la forme 3D du visage. Cette robustesse se manifeste dans les précisions apportées en terme de taux de reconnaissance et d'authentification par rapport à l'existant. Le paradigme d'utilisation apporte lui-même des améliorations du fait qu'il permet l'invariance à deux problèmes majeurs de la reconnaissance faciale qui sont les changements d'éclairage et de posture. Premièrement, disposer d'une information complète 3D du visage dans la galerie permet la reconnaissance de la requête quelque soit l'angle de prise de vue. Deuxièmement, le même paradigme permet également de s'affranchir au problème d'illumination du fait qu'il est basé seulement sur la forme 3D du visage. En fin, la métrique orientée région vient pour compléter l'ensemble et remédier au problème d'expressions faciales qui a des effets d'imprécision sur les algorithmes au tant en 2D qu'en 3D. Cette nouvelle métrique concentre ses mesures sur des régions qui varient peu intrinsèquement. Les résultats montrent bien les améliorations apportées par le premier schéma de fractionnement du visage en deux parties de haut et de bas via une étude anatomique couplée à une étude expérimentale. Nous pourrions envisager les perspectives suivantes, dans cette partie de notre travail : D'abord, nous souhaitons améliorer l'approche par calcul de distance géodésique et ce par bouchage des trous sur les modèles 3D par interpolation et l'amélioration des maillages pour un calcul plus significatif de distances géodésiques. Puis, nous souhaitons développer un détecteur de points d'intérêt sur la forme tridimensionnelle du visage plus robuste et plus efficace basé sur la géométrie différentielle des surfaces. En suite, nous envisageons la combinaison entre cette approche avec l'approche région qui peut donner des taux de reconnaissance et d'authentification plus élevée. La fusion des deux approches peut être faite par le biais du schéma de segmentation proposé. En effet, le calcul de cartes géodésiques peut être effectué uniquement sur la région statique du visage présente sur le modèle complet 3D (voir figure 6.1). Enfin, une combinaison avec une approche 2D (utilisant l'information de texture) peut être intéressante. Cette fusion peut être au niveau des scores de similarités issus de chaque modalité.

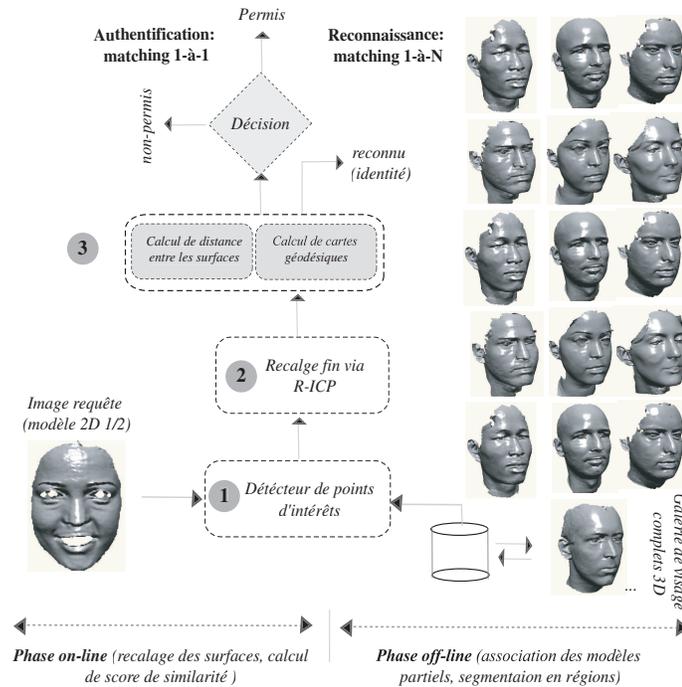


FIG. 6.1 – Vue d’ensemble des travaux futurs en reconnaissance faciale 3D.

- **En évaluation des approches développées** : Les évaluations que nous avons présenté dans le chapitre précédent ont été effectuées sur une base de visage comportant 50 individus et 400 images de test. Nous souhaitons dans nos travaux futurs, élaborer des évaluations plus significatives avec l’augmentation de ces chiffres. En effet, la base de visage 3D *ECL-IV²* sur laquelle nous avons expérimenté nos algorithmes est en cours de développement en gardant le même protocole de test. Nous souhaitons aboutir à un ensemble de 500 individus avec en total 4000 scans 3D dans le cadre du projet *IV²*. La collection de cette base et la mise au point du protocole d’évaluation découlent d’un travail de partenariat entre l’*LIRIS - ECL* et nos partenaires du projet¹.

En conclusion, nous avons développé dans cette thèse une étude de l’existant dans les domaines de reconnaissance faciale et d’acquisition tridimensionnelle. Nous avons essayé par la suite de donner des solutions aux problèmes touchant à ces deux domaines de recherche.

¹GET - INT, GET - ENST, IMEDIA - INRIA, Institut EURECOM, Let It Wave, LSC - Université d’Evry, THALES (TRT) et URATEK : Ce projet a pour but de créer des ressources et les conditions d’une évaluation à l’échelle nationale et internationale de différents systèmes liés à l’information du visage, de l’iris et de la voix, dans des milieux semi-contraints. La base de données biométriques que nous nous proposons de constituer permettra d’évaluer les performances de systèmes d’identification par l’iris, par le visage 2D et 3D et par l’analyse de visages parlants, de systèmes de détection de la position des yeux dans les images 2D et de systèmes de reconstruction 3D du visage

Deux études couplées nous ont permis d'acquérir une compréhension plus approfondie du sujet et de répondre aux besoins de l'une à l'autre. L'ensemble de ces travaux ont fait l'objets de plusieurs publications nationales et internationales : [BAC05a][BAC05c][BAC06b][BAC05b], en reconstruction 3D de visages et [BOAC05][BAC06c][BAC06a] en reconnaissance faciales 3D. D'autres communications sont en cours de révision, elles sont soumises à des revues internationales et des conférences internationales.

Bibliographie

- [AAM03] ANSARI A.N., ABDEL-MOTTALEB M., « 3-D Face Modeling Using Two Views and a Generic Face Model with Application to 3-D Face Recognition. », in *AVSS*, pp. 37–44. 2003.
- [AF86] AYACHE N., FAUGERAS O., « HYPER : a new approach for the recognition and positioning of 2D objects », *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8(1) :44–54. 1986.
- [AJB97] ACHERMANN B., JIANG X., BUNKE H., « Face Recognition using Range Data », in *Proc. International Conference on Virtual Systems and Multimedia*, pp. 129–136, Geneva, Switzerland : IEEE Press. 1997.
- [BA00] BEUMIER C., ACHEROY M., « Automatic 3D Face Authentication », *Image and Vision Computing*, 18(4) :315–321. 2000.
- [BA01a] BEUMIER C., ACHEROY M., « Face verification from 3D and grey level clues. », *Pattern Recognition Letters*, 22(12) :1321–1329. 2001.
- [BA01b] BEUMIER C., ACHEROY M., « Face Verification from 3D and Grey Level Clues », *Pattern Recognition Letters*, 22 :1321–1329. 2001.
- [BAC05a] BENAMOR B., ARDABILIAN M., CHEN L., « 3D Face Modeling Based on Structured-light Assisted Stereo Sensor », in *Proceeding of International Conference on Image Analysis and Processing*, Cagliari, Italy. september 2005.
- [BAC05b] BENAMOR B., ARDABILIAN M., CHEN L., « Acquisition 3D de visages par vision active : une solution de bout en bout », in *Proceeding of CORESA'05*, Renne, France. November 2005.
- [BAC05c] BENAMOR B., ARDABILIAN M., CHEN L., « Efficient and low-cost 2.5D and 3D face photography for recognition », in *Proceeding of IEEE International Conference On Signal-Image Technology and Internet-based Systems*, Yaounde, Cameroun. November 2005.
- [BAC06a] BENAMOR B., ARDABILIAN M., CHEN L., « Enhancing 3D Face Recognition By Mimics Segmentation », Chine. june 2006.
- [BAC06b] BENAMOR B., ARDABILIAN M., CHEN L., « An Improved 3D Human Face Reconstruction Approach Based on Cubic Splines Models », in *Third International Symposium on 3D Data Processing, Visualization and Transmission*, USA. june 2006.

- [BAC06c] BENAMOR B., ARDABILIAN M., CHEN L., « New Experiments on ICP-based 3D face recognition and authentication », in *Proceeding of International Conference on Pattern Recognition*, Hong kong. aout 2006.
- [BB81] BAKER H.H., BINFORD T.O., « Depth from Edge and Intensity Based Stereo. », in *IJCAI*, pp. 631–636. 1981.
- [BBGK04] BRONSTEIN A.M., BRONSTEIN M.M., GORDON E., KIMMEL R., « Fusion of 2D and 3D in Three-Dimensional Face Recognition », in *Proc. IEEE International Conference on Image Processing*, pp. 87–90, Singapore, oral. 2004.
- [BBK05] BRONSTEIN A.M., BRONSTEIN M.M., KIMMEL R., « Three-dimensional face recognition », *International Journal of Computer Vision (IJCV)*, 64(1) :5–30. August 2005.
- [BHK97] BELHUMEUR P.N., HESPANHA J.P., KRIEGMAN D.J., « Eigenfaces vs. Fisherfaces : Recognition Using Class Specific Linear Projection », *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7) :711–720, ISSN 0162-8828. 1997.
- [Bla04] BLAIS F., « Review of 20 Years of Range Sensor Development », *Journal of Electronic Imaging*, 13(1) :231–240. Janvier 2004.
- [BM92] BESL P.J., MCKAY N.D., « A Method for Registration of 3-D Shapes », *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2) :239–256, ISSN 0162-8828. 1992.
- [BMR⁺99] BERNARDINI F., MITTLEMAN J., RUSHMEIER H., SILVA C., TAUBIN G., « The Ball-Pivoting Algorithm for Surface Reconstruction », *IEEE Transactions on Visualization and Computer Graphics*, 5(4) :349–359, ISSN 1077-2626. 1999.
- [BOAC05] BENAMOR B., OUJI K., ARDABILIAN M., CHEN L., « 3D Face recognition by ICP-based shape matching », in *Proceeding of IEEE International Conference on Machine Intelligence*, Tozeur, Tunisie. November 2005.
- [BRV02] BLANZ V., ROMDHANI S., VETTER T., « Face Identification across Different Poses and Illuminations with a 3D Morphable Model. », in *FGR*, pp. 202–207. 2002.
- [BV99] BLANZ V., VETTER T., « A morphable model for the synthesis of 3D faces », in *SIGGRAPH '99 : Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 187–194, New York, NY, USA : ACM Press/Addison-Wesley Publishing Co. 1999.
- [BV03] BLANZ V., VETTER T., « Face Recognition Based on Fitting a 3D Morphable Model. », *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9) :1063–1074. 2003.
- [C.03] C. B., *Authentification d'identité par analyse tridimensionnelle du visage*, Thèse de doctorat, ENST. 2003.
- [CBF03] CHANG K.I., BOWYER K.W., FLYNN P.J., « Face Recognition Using 2D and 3D Facial Data », in *ACM Workshop on Multimodal User Authentication*, pp. 25–32, Santa Barbara, California. 2003.

- [CBF05] CHANG K.I., BOWYER K., FLYNN P.J., « Adaptive Rigid Multi-region Selection for Handling Expression Variation in 3D Face Recognition », in *CVPR '05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, p. 157, Washington, DC, USA : IEEE Computer Society. 2005.
- [CC03] CHOWDHURY A.K.R., CHELLAPPA R., « Face reconstruction from monocular video using uncertainty analysis and a generic model », *Comput. Vis. Image Underst.*, 91(1-2) :188–213, ISSN 1077-3142. 2003.
- [CET98] COOTES T.F., EDWARDS G.J., TAYLOR C.J., « Active Appearance Models », in *ECCV '98 : Proceedings of the 5th European Conference on Computer Vision-Volume II*, pp. 484–498, London, UK : Springer-Verlag. 1998.
- [CFFB06] CHEN X., FALTEMIER T., FLYNN P., BOWYER K., « Human Face Modeling and Recognition Through Multi-View High Resolution Stereopsis », *cvprw*, 0 :50. 2006.
- [CH85] CARRIHILL B., HUMMEL R., « Experiments with the Intensity Ratio Data Sensor », 32(3) :337–358. December 1985.
- [CM92] CHEN Y., MEDIONI G., « Object modelling by registration of multiple range images », *Image Vision Comput.*, 10(3) :145–155, ISSN 0262-8856. 1992.
- [DC00] DESCHAMPS T., COHEN L., « Minimal Paths in 3D Images and Application to Virtual Endoscopy », in *Proceedings of the Sixth European Conference on Computer Vision (ECCV'00)*, Dublin, Ireland. June 2000.
- [Dij59] DIJKSTRA E.W., « A note on two problems in connexion with graphs », in *Numerische Mathematik*, volume 1, pp. 269–271, Mathematisch Centrum, Amsterdam, The Netherlands. 1959.
- [EJCF82] E.L. H., J.B.K. T., C.A. M., F.A. S., « Measuring Curved Surfaces for Robot Vision », 15(12) :42–54. December 1982.
- [FAB97] FELDMAR J., AYACHE N., BETTING F., « 3D-2D projective registration of free-form curves and surfaces », *Comput. Vis. Image Underst.*, 65(3) :403–424, ISSN 1077-3142. 1997.
- [FP02] FORSYTH D.A., PONCE J., *Computer Vision : A Modern Approach*, Prentice Hall Professional Technical Reference. 2002.
- [FT87] FAUGERAS O., TOSCANI G., « Camera Calibration for 3D Computer Vision. », in *International Workshop on Machine Vision and Machine Intelligence*, pp. 240–247. 1987.
- [GDD01] GARCIA E., DUGELAY J.L., DELINGETTE H., « Low cost 3D face acquisition and modeling », in *ITCC 2001, IEEE International Conference on Information Technology : Coding and Computing, April 2-4, 2001, Las Vegas, USA*. Apr 2001.
- [Güh01] GÜHRING J., « Dense 3d surface acquisition by structured light using off-the-shelf components. », in *Videometrics and Optical Methods for 3D Shape Measurement*, pp. 220–231. 2001.

- [GRB94] GODIN G., RIOUX M., BARIBEAU R., « Three-dimensional registration using range and intensity information », in EL-HAKIM S.F., editor, *Proc. SPIE Vol. 2350*, p. 279-290, *Videometrics III*, Sabry F. El-Hakim; Ed., pp. 279–290. October 1994.
- [GSBC⁺03] GARCIA-SALICETTI S., BEUMIER C., CHOLLET G., DORIZZI B., LES JARDINS J.L., LUNTER J., NI Y., PETROVSKA-DELACRÉTAZ D., « BIOMET : A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities. », in *AVBPA*, pp. 845–853. 2003.
- [HBGdM05] HUSKEN M., BRAUCKMANN M., GEHLEN S., DER MALSBURG C.V., « Strategies and Benefits of Fusion of 2D and 3D Face Recognition », in *CVPR '05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, p. 174, Washington, DC, USA : IEEE Computer Society. 2005.
- [Her04] HERNÁNDEZ C., *Stereo and Silhouette Fusion for 3D Object Modeling from Uncalibrated Images Under Circular Motion*, Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications. May 2004.
- [HS04] HERNÁNDEZ C., SCHMITT F., « Silhouette and Stereo Fusion for 3D Object Modeling », *Computer Vision and Image Understanding, special issue on 'Model-based and image-based 3D Scene Representation for Interactive Visualization'*, 96(3) :367–392. December 2004.
- [HSE03] HESHER C., SRIVASTAVA A., ERLEBACHER G., « A novel technique for face recognition using range imaging. », in *ISSPA*, pp. 201–204. 2003.
- [IY96] IP H.H.S., YIN L., « Constructing a 3D individualized head model from two orthogonal views. », *The Visual Computer*, 12(5) :254–266. 1996.
- [JD96] JACKWAY P.T., DERICHE M., « Scale-Space Properties of the Multiscale Morphological Dilation-Erosion », *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(1) :38–51, ISSN 0162-8828. 1996.
- [JK97] JOHNSON A.E., KANG S.B., « Registration and integration of textured 3-D data », in *NRC '97 : Proceedings of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, p. 234, Washington, DC, USA : IEEE Computer Society. 1997.
- [Kel73] KELLY M., « Visual Identification of People by computer », Technical report, Basel and Stuttgart : Birkhauser. 1973.
- [Kos02] KOSTKOVA J., *Stereoscopic Matching : Problems and Solutions*, Thèse de doctorat, Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University. 2002.
- [KS98] KIMMEL R., SETHIAN J., « Computing geodesic paths on manifolds », . 1998.
- [KTP00] KOTROPOULOS C., TEFAS A., PITAS I., « Frontal face authentication using morphological elastic graph matching. », *IEEE Transactions on Image Processing*, 9(4) :555–560. 2000.

- [LJ05] LU X., JAIN A.K., « Integrating Range and Texture Information for 3D Face Recognition. », in *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pp. 156–163. 2005.
- [LM90] LEE J.C., MILIOS E., « Matching Range Images of Human Faces », in *Proc. IEEE International Conference on Computer Vision*, pp. 722–726. 1990.
- [LMT00] LEE W.S., MAGNENAT-THALMANN N., « Fast head modeling for animation. », *Image Vision Comput.*, 18(4) :355–364. 2000.
- [LTM96] LENGAGNE R., TAREL J.P., MONGA O., « From 2D Images to 3D Face Geometry », in *FG '96 : Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96)*, p. 301, Washington, DC, USA : IEEE Computer Society. 1996.
- [MA93] MARUYAMA M., ABE S., « Range Sensing by Projecting Multiple Slits with Random Cuts », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6) :647–651, ISSN 0162-8828. 1993.
- [MFK99] MATSUMOTO Y., FUJIMURA K., KITAMURA T., « Shape-from-Silhouette/Stereo and Its Application to 3-D Digitizer. », in *DGCI*, pp. 177–190. 1999.
- [MGM⁺05] MAURER T., GUIGONIS D., MASLOV I., PESENTI B., TSAREGORODTSEV A., WEST D., MEDIONI G., « Performance of Geometrix ActiveIDTM 3D Face Recognition Engine on the FRGC Data », in *CVPR 05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, p. 154, Washington, DC, USA : IEEE Computer Society. 2005.
- [MOC⁺98] MORANO R., OZTURK C., CONN R., DUBIN S., ZIETZ S., NISSANOV J., « Structured Light Using Pseudorandom Codes », 20(3) :322–327. March 1998.
- [MW03] MADIOUNI G., WAUPOTITSCH R., « Face recognition and modeling in 3D », in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 232–233. 2003.
- [Nic02] NICLOLAT D., « Surface measurement and tracking of human body parts from multi image video sequences », *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(5-6) :360–375. December 2002.
- [Nic03] NICLOLAT D., *Surface measurement and tracking of human body parts from multi station video sequences*, Thèse de doctorat, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland. 2003.
- [OK85] OHTA Y., KANADE T., « Stereo by Two-Level Dynamic Programming. », in *IJCAI*, pp. 1120–1126. 1985.
- [PA82] POSDAMER J., ALTSCHULER M., « Surface Measurement by Space-Encoded Projected Beam System », *CGIP*, 18(1) :1–17. January 1982.
- [PFS⁺03] PHILLIPS P.J., FLYNN P.J., SCRUGGS T., BOWYER K.W., CHANG J., HOFFMAN K., MARQUES J., MIN J., WOREK W., « Overview of the Face

- Recognition Grand Challenge », in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*. june 2003.
- [PGM⁺03] PHILLIPS P., GROTH P., MICHEALS R., BLACKBURN D., TABASSI E., BONE J., « FRVT 2002 : Evaluation Report », Technical report, NIST. March 2003.
- [PJ06] PARK U., JAIN A.K., « 3D Face Reconstruction from Stereo Video », in *CRV '06 : Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, p. 41, Washington, DC, USA : IEEE Computer Society. 2006.
- [PMRR00] PHILLIPS P.J., MOON H., RIZVI S.A., RAUSS P.J., « The FERET Evaluation Methodology for Face-Recognition Algorithms », *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10) :1090–1104, ISSN 0162-8828. 2000.
- [PR04] PAPTODOROU T., RUECKERT D., « Evaluation of Automatic 4D Face Recognition Using Surface and Texture Registration. », in *FGR*, pp. 321–326. 2004.
- [PV03] PLATANIOTIS J.L., VENETSANOPOULOS K., « Face recognition using LDA-based algorithms », *Neural Networks, IEEE Transactions on*, 14 :195– 200. 2003.
- [PW05] PAN G., WU Z., « 3D Face Recognition from Range Data », *International Journal of Image and Graphics*, 5(3) :1–21. 2005.
- [RL01] RUSINKIEWICZ S., LEVOY M., « Efficient variants of the ICP algorithm », in *3DIM01*, pp. 145–152. 2001.
- [Set96] SETHIAN J., « A Fast Marching Level Set Method for Monotonically Advancing Fronts », in *Proc. Nat. Acad. Sci.*, volume 93, pp. 1591–1595. 1996.
- [SI85] SATO K., INOKUCHI S., « Three-dimensional surface measurement by space encoding range imaging », 2(1) :27–39. 1985.
- [SPB04] SALVI J., PAGÈS J., BATLLE J., « Pattern codification strategies in structured light systems. », *Pattern Recognition*, 37(4) :827–849. 2004.
- [SW95] SAKAS G., WALTER S., « Extracting surfaces from fuzzy 3D-ultrasound data », in *SIGGRAPH '95 : Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 465–474, New York, NY, USA : ACM Press. 1995.
- [TI90] TAJIMA J., IWAKAWA M., « 3D data acquisition by rainbow range finder », in *In Proc. of International Conference on Pattern Recognition*, pp. 309–31. 1990.
- [TIC98] TANAKA H.T., IKEDA M., CHIAKI H., « Curvature-based Face Surface Recognition using Spherical Correlation - Principal Directions for Curved Object Recognition », in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 372– 377. 1998.

- [TKP02] TEFAS A., KOTROPOULOS C., PITAS I., « Face verification using elastic graph matching based on morphological signal decomposition », *Signal Process.*, 82(6) :833–851, ISSN 0165-1684. 2002.
- [TP91] TURK M., PENTLAND A., « Eigenfaces for recognition », *Journal of Cognitive Neuroscience*, 3(1) :71–86. 1991.
- [Tro95] TROBINA M., « Error Model of a Coded-Light Range Sensor », Technical report, Communication Technology Laboratory. september 1995.
- [Tsa87] TSAI R.Y., « A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses », pp. 323–344. 1987.
- [TTS03] TSALAKANIDOU F., TZOVARAS D., STRINTZIS M.G., « Use of depth and colour eigenfaces for face recognition », *Pattern Recogn. Lett.*, 24(9-10) :1427–1435, ISSN 0167-8655. 2003.
- [TV98] TRUCCO E., VERRI A., *Introductory Techniques for 3-D Computer Vision*, Upper Saddle River, NJ, USA : Prentice Hall PTR. 1998.
- [VO90] VUYLSTEKE P., OOSTERLINCK A., « Range Image Acquisition with a Single Binary-Encoded Light Pattern », *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(2) :148–164, ISSN 0162-8828. 1990.
- [VTC05] VOGIATZIS G., TORR P.H.S., CIPOLLA R., « Multi-View Stereo via Volumetric Graph-Cuts », *cvpr*, 2 :391–398, ISSN 1063-6919. 2005.
- [WCH92] WENG J., COHEN P., HERNIOU M., « Camera Calibration with Distortion Models and Accuracy Evaluation », *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(10) :965–980, ISSN 0162-8828. 1992.
- [XJF01] X. S., J.Y. E., F.H. D., « Three-dimensional multivision determination of human body segment characteristics. », in *18th Congress of the International Society of Biomechanics*. 2001.
- [Y.99] Y. B.J., *Visual methods for three-dimensional modeling*, Thèse de doctorat, California Institute of Technology Pasadena California. 1999.
- [Ye88] YE Q.Z., *Range Cameras Based on Structured Light*, Thèse de doctorat, Linköping University, Linköping, Sweden. 1988.
- [YY01] YU H., YANG J., « A Direct LDA Algorithm for High-Dimensional Data - with Application to Face Recognition », *Pattern Recognition*, 34 :2067–2070. 2001.
- [ZCPR03] ZHAO W., CHELLAPPA R., PHILLIPS P.J., ROSENFELD A., « Face recognition : A literature survey », *ACM Comput. Surv.*, 35(4) :399–458, ISSN 0360-0300. 2003.
- [ZCS02] ZHANG L., CURLESS B., SEITZ S.M., « Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. », in *3DPVT*, pp. 24–37. 2002.

- [Zha00] ZHANG Z., « A Flexible New Technique for Camera Calibration », *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11) :1330–1334, ISSN 0162-8828. 2000.
- [Zhe94] ZHENG J.Y., « Acquiring 3-D Models from Sequences of Contours », *PAMI*, 16(2) :163 – 178. February 1994.
- [ZYL97] ZHANG J., YAN Y., LADES M., « Face Recognition : Eigenface[S], Elastic Matching, And Neural Nets », *P-IEEE*, 85 :1422–1435. 1997.

Liste des figures

2.1	Reconstruction ultrasonore d'un fœtus (gauche), de la surface 3D (milieu) et la photo du bébé 24h après l'accouchement (droite).	8
2.2	Taxonomie des techniques d'acquisition de formes 3D.	9
2.3	Principe d'acquisition 3D par triangulation active.	11
2.4	Extrait de la base de corps humains 3D <i>CAESAR anthropometric database</i>	11
2.5	Extrait des bases de visages IV2 et FRGC1.0.	13
2.6	Deux exemples de codage temporel et les résultats respectifs de reconstruction.	14
2.7	Exemples de motifs pour une stratégie de codage spatial.	15
2.8	Exemple de motifs de lumières en couleurs envoyé sur l'objet à numériser.	15
2.9	Construction de l'enveloppe visuelle par intersection des cônes définis par les centres optiques des caméras et les silhouettes de l'objet.	17
2.10	Taxonomie des approches de reconnaissance de visages 2D, 2D/3D, 3D et 2D+3D.	20
3.1	Pipeline des étapes de notre approche de reconstruction 3D de visages.	31
3.2	Relation entre un point 3D et sa projection 2D.	34
3.3	Interface de la boîte à outils de calibration utilisée pour la calibration de notre capteur stéréo.	40
3.4	Disposition des caméras par rapport à la mire en mouvement.	41
3.5	Principe de projection de lumière et de détection sous-pixélique de primitives.	42
3.6	Déviations entre intersection par modèle linéaire et modèle de splines cubiques.	43
3.7	Principe de projection de lumière et de détection sous-pixélique de primitives.	45
3.8	Contrainte d'ordre (monotonie) : cas du visage.	48
3.9	Exemple de calcul de matrice des coûts et du chemin optimal : (A) cas d'absence d'occlusion, (B) cas de présence d'occlusion.	49
3.10	Interpolation par Splines cubiques pour densifier la reconstruction.	53
3.11	Dualité : Diagramme de Voronoï/Triangulation de Delaunay.	55
3.12	Formation du modèle 2.5D par interpolation puis par maillage des points	56

3.13	Recalage et fusion de modèles partiels en vue de produire un modèle 3D complet.	57
3.14	Résultats finaux de la reconstruction 3D	60
3.15	Qualité de la reconstruction 3D : parties forme 3D et texture correspondantes	61
3.16	Déviations spatiales entre les données de références (laser) et les données mesurées (notre capteur)	62
3.17	Différentes sections des données de références (laser) et les données mesurées (notre capteur)	63
4.1	Problèmes de reconnaissance de visages : (A) variations intra-classes (B) similarité inter-classes.	66
4.2	Nouvelles formes d'acquisition de visage en vue de la reconnaissance.	66
4.3	Vue d'ensemble de l'approche de reconnaissance/authentification proposée.	68
4.4	Exemple1 : recalage et calcul de similarité entre un modèle 3D et un modèle requête de la même personne avec des variations d'expressions faciales.	75
4.5	Exemple2 : recalage et calcul de similarité entre un modèle 3D et un modèle requête de la même personne avec des variations d'expressions faciales plus significatives.	76
4.6	Exemple3 : recalage et calcul de similarité entre un modèle 3D et un modèle requête de deux personnes différentes.	77
4.7	Compensation des expressions faciale par l'approche région.	83
4.8	Etude de l'anatomie du visage : composition osseuse et maillage musculaire.	85
4.9	Ensemble des muscles élémentaires considérés pour l'étude empirique.	86
4.10	Technique de mesure de l'influence de chaque muscle ou région du visage sur la déformation de celui-ci par rapport à un état neutre.	87
4.11	Ensemble des muscles et parties élémentaires considérés pour l'étude empirique.	88
4.12	Procédure de segmentation automatique des visages 3D complets de la galerie.	89
4.13	Illustration de l'algorithme de cheminement rapide.	93
4.14	Triangulation aiguë autour d'un point centre de la grille.	94
4.15	A gauche : Une vue perspective du triangle supportant les valeurs de $T()$, à droite : Trigonométrie sur le plan défini par le triangle ABC.	94
4.16	Cas d'une triangulation quelconque.	96
4.17	Calculs des cartes de distances géodésiques sur les modèles 3D et 2.5D après recalage des surfaces moyennant ICP.	98
5.1	Variété des données de la base $ECL - IV^2$ et algorithmes de matching concernés.	103
5.2	Association de modèles partiels en vue d'obtenir un modèle 3D complet	104

5.3	Différentes variations dans les modèles de test : (A) poses, (B) expressions et (C) illumination.	105
5.4	Extrait de notre base d'évaluation <i>ECL-IV</i> ² : Modèles complets 3D de la BD, modèles partiel de test.	106
5.5	Différentes résolutions du visage 3D obtenues par remaillage adaptatif et formes 3D correspondantes.	107
5.6	Matrice de scores de similarité calculées entre les images de la galerie et les images de test.	112
5.7	Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'experimentation effectuée.	113
5.8	(a) Courbes DET (Error trade-off), (b) Courbes ROC (Receiver operating characteristic), pour toutes les images de test et pour chaque test individuellement.	114
5.9	Courbes DET (Error trade-off) pour tout les probes de la BD en variant α	116
5.10	Courbes ROC (Receiver operating characteristic) pour tout les probes de la BD en variant α	117
5.11	De gauche à droite et de haut en bas : courbes DET (Error trade-off) pour les images probes avec expressions faciales en variant α	119
5.12	Matrice de scores de similarité calculées entre les images de la galerie et les images de test.	120
5.13	Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'experimentation effectuée.	120
5.14	Matrice de scores de similarité calculées entre les images de la galerie et les images de test.	122
5.15	Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'experimentation effectuée.	122
5.16	(a) Courbes DET (Error trade-off), (b) Courbes ROC (Receiver operating characteristic), pour toutes les images de test et pour chaque test individuellement.	123
5.17	(a) Courbes <i>FAR/FRR</i> pour l'algorithme <i>ICP</i> , (b) Distributions des <i>clients/imposteurs</i> pour l'algorithme <i>ICP</i> , (c) Courbes <i>FAR/FRR</i> pour l'algorithme <i>R-ICP</i> , (d) Distributions des <i>clients/imposteurs</i> pour l'algorithme <i>R-ICP</i> , (e) Courbes <i>FAR/FRR</i> pour l'algorithme <i>Geodesics</i> , (f) Distributions des <i>clients/imposteurs</i> pour l'algorithme <i>Geodesics</i>	125
5.18	Résultats de la reconnaissance : taux de reconnaissance en fonction de la nature de l'experimentation effectuée pour les trois algorithmes <i>ICP</i> , <i>R-ICP</i> et calcul de cartes géodésiques.	126
5.19	Courbes DET relatives aux trois algorithmes <i>ICP</i> , <i>R-ICP</i> et <i>Geodesics</i> pour une expérimentation globale.	127

6.1 Vue d'ensemble des travaux futurs en reconnaissance faciale 3D. 131