Contributions à l'analyse de visages en 3D : approche régions, approche holistique et étude de dégradations

Pierre Lemaire

THESE

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

présentée et soutenue publiquement par

Pierre LEMAIRE

Contributions à l'analyse de visages en 3D : Approche régions, approche holistique et étude de dégradations

---

Ecole Doctorale InfoMaths

Directeur de thèse : Liming CHEN
Co-directeur de thèse : Mohamed DAOUDI

JURY

Prof. Alice CAPLIER	Université INP	Rapporteur
Prof. Youssef CHAHIR	Université de Caen	Rapporteur
Prof. Samir AKKOUCHE	Université Claude Bernard	Examinateur
Dr. Kevin BAILLY	Université Pierre et Marie Curie	Examinateur
Prof. Liming CHEN	Ecole Centrale de Lyon	Directeur de thèse
Prof. Mohamed DAOUDI	Telecom Lille1	Co-directeur de thèse

Remerciements

L'aboutissement de ce travail doit beaucoup aux nombreuses personnes qui m'ont soutenu, porté, et sur lesquelles j'ai pu me reposer tout au long de ces quatre ans et demi de thèse.

Je tiens en premier lieu à remercier mes encadrants, Liming Chen et Mohamed Daoudi, pour leur support logistique, matériel et scientifique. Celà n'a pas toujours été simple de me soutenir, et je remercie particulièrement Liming pour sa persévérance.

Je tiens également à adresser mes remerciements à Mohsen Ardabilian, qui a su se rendre présent et altruiste dans les moments les plus opportuns.

Je souhaite remercier les membres de mon jury de thèse, Alice Caplier, Youssef Chahir, Kevin Bailly et Samir Akkouche pour le temps qu'ils m'ont accordé et leurs remarques constructives.

Je veux remercier Boulbaba Ben Amor et Dimitris Samaras pour avoir su me guider dans mon travail.

Je salue l'ensemble des membres du projet FAR-3D pour le travail réalisé... et je souhaite aussi m'excuser auprès de Joseph pour les appels intempestifs.

Je souhaite remercier Jean-Pierre Bertoglio, Johannes Kellendonk, Attila Baskurt, Mohand-Said Hacid pour m'avoir écouté, puis accordé leur confiance au moment où celà comptait le plus.

J'aimerais dire ma gratitude à l'ensemble du département Math-Info de l'École Centrale de Lyon, et particulièrement Christian, Emmanuel et Charles-Edmond pour leur support scientifique, mais aussi pédagogique lorsque j'enseignais.

Je tiens aussi tout particulièrement à remercier Isabelle et Colette pour leur indispensable travail de l'ombre, leur très grande disponibilité, leur gentillesse et la psychologie dont elles ont souvent su faire preuve au quotidien.

Je souhaite aussi remercier Przemek, Hui-Bin, Zhao, Wael, Karima et Alex pour les bons moments partagés au travail, mais aussi en dehors. Alex, remets-toi bien.

Je veux aussi saluer Aliaksandr, avec qui j'ai toujours adoré discuter, mais aussi faire les poubelles...

Je veux remercier Tom pour m'avoir donné l'idée de faire de la recherche, et pour m'avoir mis le pied à l'étrier lorsque je n'étais que stagiaire à Meylan.

Je veux dire merci à Nicolas et Julien pour m'avoir permis de m'évader dans le vacarme. Pour m'avoir aidé à construire quelque chose en dehors de ma thèse aussi.

J'ai une petite pensée pour Ludo, Nico, Adyl, Bastien et Thomas, qui m'ont suivi en filigrane pendant toutes ces années.

Je veux dire merci à ma famille, et surtout à ma sœur et mes parents pour les (trop) nombreuses relectures.

Enfin, je ne veux surtout pas oublier Sophie, qui a réussi à me supporter au quotidien pendant ces quatre ans et demi. Je n'aurais pas démarré, et encore moins achevé, ce projet sans toi. Mille mercis.

Table des matières

Résumé

Historiquement et socialement, le visage est chez l'humain une modalité de prédilection pour déterminer l'identité et l'état émotionnel d'une personne. Il est naturellement exploité en vision par ordinateur pour les problèmes de reconnaissance de personnes et d'émotions. Les algorithmes d'analyse faciale automatique doivent relever de nombreux défis : ils doivent être robustes aux conditions d'acquisition ainsi qu'aux expressions du visage, à l'identité, au vieillissement ou aux occultations selon le scénario. La modalité 3D a ainsi été récemment investiguée. Elle a l'avantage de permettre aux algorithmes d'être, en principe, robustes aux conditions d'éclairage ainsi qu'à la pose.

Cette thèse est consacrée à l'analyse de visages en 3D, et plus précisément la reconnaissance faciale ainsi que la reconnaissance d'expressions faciales en 3D sans texture.

Nous avons dans un premier temps axé notre travail sur l'apport que pouvait constituer une approche régions aux problèmes d'analyse faciale en 3D. L'idée générale est que le visage, pour réaliser les expressions faciales, est déformé localement par l'activation de muscles ou de groupes musculaires. Il est alors concevable de décomposer le visage en régions mimiques et statiques, et d'en tirer ainsi profit en analyse faciale. Nous avons proposé une paramétrisation spécifique, basée sur les distances géodésiques, pour rendre la localisation des régions mimiques et statiques le plus robustes possible aux expressions. Nous avons également proposé une approche régions pour la reconnaissance d'expressions du visage, qui permet de compenser les erreurs liées à la localisation automatique de points d'intérêt. Les deux approches proposées dans ce chapitre ont été évaluées sur des bases standards de l'état de l'art.

Nous avons également souhaité aborder le problème de l'analyse faciale en 3D sous un autre angle, en adoptant un système de cartes de représentation de la surface 3D. Nous avons ainsi proposé de projeter sur le plan 2D des informations liées à la topologie de la surface 3D, à l'aide d'un descripteur géométrique inspiré d'une mesure de courbure moyenne. Les problèmes de reconnaissance faciale et de reconnaissance d'expressions 3D sont alors ramenés à ceux de l'analyse faciale en 2D. Nous avons par exemple utilisé SIFT pour l'extraction puis l'appariement de points d'intérêt en reconnaissance faciale. En reconnaissance d'expressions, nous avons utilisé une méthode de description des visages basée sur les histogrammes de gradients orientés, puis classé les expressions à l'aide de SVM multi-classes. Dans les deux cas, une méthode de fusion simple permet l'aggrégration des résultats obtenus à différentes échelles. Ces deux propositions ont été évaluées sur la base BU-3DFE, montrant de

bonnes performances tout en étant complètement automatiques.

Enfin, nous nous sommes intéressés à l'impact des dégradations des modèles 3D sur les performances des algorithmes d'analyse faciale. Ces dégradations peuvent avoir plusieurs origines, de la capture physique du visage humain au traitement des données en vue de leur interprétation par l'algorithme. Après une étude des origines et une théorisation des types de dégradations potentielles, nous avons défini une méthodologie permettant de chiffrer leur impact sur des algorithmes d'analyse faciale en 3D. Le principe est d'exploiter une base de données considérée sans défauts, puis de lui appliquer des dégradations canoniques et quantifiables. Les algorithmes d'analyse sont alors testés en comparaison sur les bases dégradées et originales. Nous avons ainsi comparé le comportement de 4 algorithmes de reconnaissance faciale en 3D, ainsi que leur fusion, en présence de dégradations, validant par la diversité des résultats obtenus la pertinence de ce type d'évaluation.

Mots-clés : Analyse de visages en 3D ; Reconnaissance faciale en 3D ; Reconnaissance d'Expressions Faciales en 3D ; Approche basée Régions ; Cartes de représentation ; Cartes de Différence de Courbure Moyenne ; Dégradations.

Abstract

Historically and socially, the human face is one of the most natural modalities for determining the identity and the emotional state of a person. It has been exploited by computer vision scientists within the automatic facial analysis domain. Still, proposed algorithms classically encounter a number of shortcomings. They must be robust to varied acquisition conditions. Depending on the scenario, they must take into account intra-class variations such as expression, identity (for facial expression recognition), aging, occlusions. Thus, the 3D modality has been suggested as a counterpoint for a number of those issues. In principle, 3D views of an object are insensitive to lightning conditions. They are, theoretically, pose-independant as well.

The present thesis work is dedicated to 3D Face Analysis. More precisely, it is focused on non-textured 3D Face Recognition and 3D Facial Expression Recognition.

In the first instance, we have studied the benefits of a region-based approach to 3D Face Analysis problems. The general concept is that a face, when performing facial expressions, is deformed locally by the activation of muscles or groups of muscles. We then assumed that it was possible to decompose the face into several regions of interest, assumed to be either mimic or static. We have proposed a specific facial surface parametrization, based upon geodesic distance. It is designed to make region localization as robust as possible regarding expression variations. We have also used a region-based approach for 3D facial expression recognition, which allows us to compensate for errors relative to automatic landmark localization.

We also wanted to experiment with a Representation Map system. Here, the main idea is to project 3D surface topology data on the 2D plan. This translation to the 2D domain allows us to benefit from the large amount of related works in the litterature. We first represent the face as a set of maps representing different scales, with the help of a geometric operator inspired by the Mean Curvature measure. For Facial Recognition, we perform a SIFT keypoints extraction. Then, we match extracted keypoints between corresponding maps. As for Facial Expression Recognition, we normalize and describe every map thanks to the Histograms of Oriented Gradients algorithm. We further classify expressions using multi-class SVM. In both cases, a simple fusion step allows us to aggregate the results obtained on every single map.

Finally, we have studied the impact of 3D models degradations over the performances of 3D facial analysis algorithms. A 3D facial scan may be an altered representation of its real life model, because of several reasons, which range from the physical caption of the human model to data processing. We propose a methodology that allows us to quantify the impact of every single type of degradation over

the performances of 3D face analysis algorithms. The principle is to build a database regarded as free of defaults, then to apply measurable degradations to it. Algorithms are further tested on clean and degraded datasets, which allows us to quantify the performance loss caused by degradations. As an experimental proof of concept, we have tested four different algorithms, as well as their fusion, following the aforementioned protocol. With respect to the various types of contemplated degradations, the diversity of observed behaviours shows the relevance of our approach.

Keywords : 3D Facial Analysis ; 3D Facial Recognition ; 3D Facial Expression Recognition ; Regions-based Approach ; Representation Map ; Differential Mean Curvature Maps ; Degradations.

Publications

Les travaux effectués lors de mon doctorat ont fait l'objet de quatre publications dans des conférences internationales et deux dans une conférence nationale.

Conférences Internationales :

1. Huibin Li, Di Huang, Pierre Lemaire, Jean-Marie Morvan, Liming Chen, *Expression robust 3D face recognition via meshed-based histograms of multiple order surface differential quantities*, ICIP 2011
2. Pierre Lemaire, Di Huang, Joseph Colineau, Mohsen Ardabilian, Liming Chen, *3D Face Recognition in the presence of 3D model degradations*, Biosig 2011
3. Pierre Lemaire, Boulbaba Ben Amor, Mohsen Ardabilian, Liming Chen, Mohamed Daoudi, *Fully Automatic 3D Facial Expression Recognition using a Region-Based Approach*, MA3HO 2011, Workshop of ACMM 2011
4. Wael Ben Soltana, Mohsen Ardabilian, Pierre Lemaire, Di Huang, Przemyslaw Szeptycki, Liming Chen, Boulbaba Ben Amor, Hassen Drira, Mohamed Daoudi, Nesli Erdogmus, Lionel Daniel, Jean-Luc Dugelay, Joseph Colineau, *3D Face Recognition : A Robust Multi-matcher Approach to Data Degradations*, ICB 2012
5. Pierre Lemaire, Mohsen Ardabilian, Liming Chen, Mohamed Daoudi, *Fully Automatic 3D Facial Expression Recognition using Differential Mean Curvature Maps and Histograms of Oriented Gradients*, 3D Face Biometrics, Workshop of Face and Gesture 2013

Conférences Nationales :

1. Pierre Lemaire, Przemyslaw Szeptycki, Mohsen Ardabilian, Liming Chen, *Reconnaissance de visages en 3D orientée région*, Coresa 2010.
2. Pierre Lemaire, Wael Ben Soltana, Di Huang, Karima Ouji, Mohsen Ardabilian, Liming Chen, *Reconnaissance rapide, robuste et résistante aux leurres de visages en 3D*, WISG 2011.

Introduction

0.1 Contexte

En 1891 a lieu la création du premier fichier d'empreintes digitales par Juan Vucetich en Argentine. En France, le système Bertillon, correspondant au renseignement de mensurations de l'individu à identifier, est alors déjà en place depuis huit ans. À partir de 1912, les empreintes digitales ainsi que les données anthropométriques sont intégrées à la carte d'identité nationale de manière systématique par le biais du carnet anthropométrique. Cette carte d'identité est d'abord réservée aux nomades, puis elle est généralisée à l'ensemble de la population française qui le désire en 1921. L'article 8 du décret d'application de 1913 pour le carnet anthropométrique stipule :

> *Il doit, en outre, recevoir le signalement anthropométrique qui indique notamment la hauteur de la taille, celle du buste, l'envergure, la longueur et la largeur de la tête, le diamètre bizygomatique, la longueur de l'oreille droite, la longueur des doigts médius et auriculaires gauches, celle de la coudée gauche, celle du pied gauche, la couleur des yeux : des cases sont réservées pour les empreintes digitales et pour les deux photographies (profil et face) du porteur du carnet.*

En 1987, la carte d'identité cartonnée est remplacée par la carte nationale d'identité informatisée, dont le but est entre autres d'accélérer les recoupements relatifs aux empreintes digitales. En 2006 sont mis en circulation en France les passeports biométriques. En 2012, la proposition d'instaurer une nouvelle carte d'identité électronique, ainsi qu'un fichier central réunissant les données de cette future carte et celles du passeport biométrique est adoptée par l'état français. Celui-ci comprend la taille, la couleur des yeux, deux empreintes digitales et une photographie de face de l'individu concerné.

Ces évolutions dans l'utilisation et la généralisation des données biométriques ne sont pas restreintes au seul cas de la France. La prise des empreintes digitales est prévue par la législation de la communauté européenne pour la demande d'un passeport dans ses états membres. Aux États-Unis, les *mugshots*, célèbres portraits de face et de profil des individus en procédure d'arrestation, font partie de la culture populaire. En 2008, l'organisation des Jeux Olympiques de Pékin a fait pour la première fois l'usage de l'identification automatique du visage pour l'accès aux cérémonies d'ouverture et de fermeture [Ao *et al.* 2009]. En 2012, un système de reconnaissance faciale est proposé aux petits commerces de Nashville en lien avec Facebook pour

proposer des réductions en fonction des informations accessibles sur le profil des clients [Browser 2012]. En Inde est actuellement mis en place un système appelé "UID", qui identifie les résidents indiens grâce à leurs empreintes digitales et leur iris [Chakraborthy *et al.* 2011]. Le système revendiquait en octobre 2012 plus de 200 millions d'enregistrements. L'objectif est à terme de réguler l'accès aux prestations sociales et aux démarches administratives pour la population indienne (aujourd'hui recensée à plus d'un milliard de personnes).

De nombreux moyens d'identification biométrique ont vu le jour depuis les premiers pas des pionniers Bertillon, Pinkerton et Vucetich. Si leur utilisation fait toujours débat, force est de constater que leur nombre s'est considérament accru depuis la fin du XIX$^{\text{ème}}$ siècle. Citons entre autres, en plus des données anthropométriques, de la reconnaissance faciale et de l'identification par empreintes digitales : l'identification par l'iris, l'analyse comportementale et notamment la reconnaissance par la démarche, la reconnaissance par la géométrie des veines de la main...

Parmi ces modalités, la reconnaissance faciale a la réputation d'être socialement bien acceptée. Elle est naturelle –dans le sens où elle est pratiquée instinctivement et/ou historiquement par l'être humain– et non-intrusive. Elle a également l'avantage (selon le scénario) de ne pas requérir la coopération de la personne concernée. Sur le plan informatique, cette modalité est également facilitée par la très large démocratisation des capteurs optiques.

Mais en informatique, le visage humain n'est pas seulement un objet de biométrie. Les applications de l'analyse faciale s'étendent aux Interfaces Homme Machine (IHM), comprenant l'informatique affective [Jaimes & Sebe 2007] ; elles incluent la modélisation de visages à des fins de synthèse ; l'analyse d'expressions du visage améliore les performances des techniques de reconnaissance automatique de langage et de structure de la conversation [Otsuka *et al.* 2007] ; elle peut également trouver sa place dans des domaines relatifs à la sécurité, comme la détection de micro-expressions (détection de mensonges) ou de comportements anormaux, en plus d'améliorer potentiellement la qualité des méthodes automatiques de reconnaissance de personnes.

L'analyse faciale, tant sur l'aspect de la biométrie que celui de la reconnaissance d'expressions, est donc un important champ d'investigation de la vision par ordinateur et de la reconnaissance de formes. Les performances des algorithmes sont critiques en reconnaissance et en identification de personnes ; et les résultats obtenus par les algorithmes en reconnaissance d'expressions laissent espérer de possibles améliorations. Cependant, ces algorithmes sont constamment soumis aux conditions d'acquisition, pouvant affecter de façon variable la pose ou l'éclairage. Ces aspects demeurent des challenges dans le domaine de l'analyse de visages en 2D.

Depuis quelques années, le développement et la démocratisation progressive de capteurs d'images en 3 dimensions a facilité le développement du domaine de recherche qu'est l'analyse faciale en 3D. Certaines des problématiques majeures de l'analyse de visages en 2D, à savoir la détection, l'identification et la reconnaissance d'individus et la reconnaissance d'expressions, se sont logiquement retrouvés transposés à la 3D. L'avantage de la 3D est *a priori* sa robustesse aux problèmes d'éclairage et de pose. L'analyse de visages en 3D est par ailleurs soumise à des difficultés comparables à celles rencontrées par les méthodes 2D, à savoir la différenciation des variations intra-classe (expressions faciales, vieillissement...) et des similarités extra-classe (jumeaux, parents, sosies), ainsi que les occultations et les problèmes de qualité d'acquisition. Ajoutons des problématiques potentiellement différentes au niveau de la complexité calculatoire, et plus directement liées à la modalité 3D.

Que ce soit dans le cas de la reconnaissance et de l'identification de personnes, ou dans le cas de la reconnaissance d'expressions du visage, le défi consiste à différencier des variabilités *intra-classe* et *extra-classe*. L'identification de personne devant se faire indépendamment de l'expression, et la reconnaissance d'expressions devant se faire indépendamment de l'identité, les deux problèmes sont ainsi, en grande partie, complémentaires. Ils ne sont d'ailleurs pas contradictoires, et on peut espérer voir un jour des algorithmes capables de reconnaitre simultanément identité et expression d'un visage de manière robuste. En effet, et bien que les scénarios d'application faisant usage d'une reconnaissance faciale et ceux exploitant la reconnaissance d'expressions soient la plupart du temps dissociés, nous pensons qu'une approche unifiée puisse se justifier. Dans la mesure où la robustesse aux expressions est une des difficultés majeures de la reconnaissance faciale, nous supposons que la connaissance de l'expression puisse jouer un rôle positif dans la reconnaissance de l'identité. Le scénario complémentaire, c'est-à-dire l'amélioration des techniques de reconnaissance d'expression grâce à la connaissance de l'identité, est tout aussi valable. C'est d'ailleurs, à notre sens, l'asomption faite par les techniques basées sur l'exploitation de modèles déformables (partie 1.6.4).

La fiabilité de tels algorithmes doit être évaluée avec précision. À cet effet, diverses bases de données publiques comme, entre autres, FRGC (section 1.2.2.1) ou BU-3DFE (section 1.2.2.3) ont vu le jour, lesquelles sont exploitées selon des standards donnés et précisés dans la littérature scientifique.

Dans la suite de cet écrit, nous nous intéresseront de manière exclusive au domaine de l'analyse de visages en 3D, et plus précisément aux problématiques de reconnaissance et d'identification de personnes et à la reconnaissance d'expressions. Bien que des bases de données 4D publiques existent (3D dynamique), nous ne nous intéresserons pas dans ce manuscrit à ce problème.

0.2 Objectifs et contributions

0.2.1 Objectifs et contraintes

Dans cette thèse, nous avons abordé des problématiques courantes dans le domaine de l'analyse de visage en 3D, à savoir la reconnaissance de visages et la reconnaissance d'expressions. Les deux sujets étant liés, nous avons cherché des solutions et des manières d'envisager le problème qui puissent être communes aux deux problématiques. Ces sujets sont néanmoins vastes et demandent à être précisés.

Nous nous sommes intéressés à la modalité 3D, estimant que l'apport de la modalité 2D pouvait être ultérieure, indépendante et complémentaire à cette approche, par le biais de méthodes de fusions. Nous n'avons pas choisi de développer la fusion dans le cadre de cette thèse.

Plus spécifiquement, nous nous sommes intéressés à l'invariance aux expressions pour la reconnaissance faciale, et inversement à l'invariance à l'identité dans le cadre de la reconnaissance d'expressions du visage.

Nous avons également mis l'accent sur l'aspect entièrement automatique de nos méthodes. Dans nos travaux, nous avons essayé de faire un minimum appel à l'opérateur humain, notamment en ce qui concerne l'annotation de points d'intérêt manuels.

Nous avons également travaillé directement sur les données fournies dans les bases de données, c'est-à-dire déjà sous forme de maillages et d'images de profondeur.

Enfin, nous nous sommes intéressés au problème des dégradations des modèles 3D et leur impact sur les performances des algorithmes d'analyse faciale en 3D.

0.2.2 Approches proposées et contributions

Dans ce travail de thèse, nous avons proposé plusieurs approches.

- L'application d'une approche orientée régions aux problèmes de reconnaissance de personnes et de reconnaissance d'expressions du visage.
- Une représentation des visages en 3D par cartes. Ces cartes sont la projection dans le plan d'une mesure s'apparentant à la courbure moyenne, obtenue par un calcul de volumes sur l'objet 3D. Nous en montrons les applications à la fois pour la reconnaissance de personnes et la reconnaissance d'expressions.
- Une méthodologie pour tester la robustesse des algorithmes d'analyse faciale aux dégradations des modèles 3D, en vue d'une application pratique des algorithmes de l'état de l'art.

Cette thèse comprend les contributions principales suivantes :

- Une représentation de la surface faciale par une mesure basée sur les distances géodésiques à des points anatomiques localisés automatiquement. Cette me-

sure est invariante en pose, et nous montrons qu'elle est peu sensible aux expressions dans le cadre de la reconnaissance de personnes.
- Une approche régions pour compenser l'imprécision des landmarks localisés automatiquement dans le contexte de la reconnaissance d'expressions.
- Une représentation des images de profondeur s'apparentant à une mesures de courbure, emphvia un calcul intégral, que nous appellerons Carte de Différence de Courbure Moyenne.
- Une méthodologie pour l'analyse de l'impact des dégradations à l'acquisition des modèles en 3D sur les performances des algorithmes d'analyse 3D.

0.3 Contenu des différents chapitres

Ce mémoire est composé de quatre chapitres.

Le chapitre 1 est un état de l'art de l'analyse faciale en 3D. Dans cette section, nous présentons d'abord des connaissances et des résultats préalables à l'analyse de visages en 3D, comme les méthodes d'acquisition de visages en 3D, les bases de données couramment employées dans le domaine et le système FACS, qui décrit et décompose les relations entre les expressions du visage et les muscles faciaux. Dans la présentation des travaux de l'état de l'art, nous avons fait le choix de factoriser les approches afin de montrer les similitudes et les divergences entre différentes familles de méthodes de reconnaissance de personnes et d'expressions.

Le chapitre 2 présente une approche orientée régions, à la fois pour les problèmes d'identification de personnes et de reconnaissance d'expressions. Nous avons d'abord fait le choix de nous inspirer et de prolonger les travaux effectués dans [Ben Amor 2006]. Le visage est donc représenté comme un ensemble de régions, sur lesquelles nous effectuons individuellement des mesures de similarité. La sélection des régions employées s'appuie sur une étude du système FACS (section 1.3). Nous montrons que cette approche peut être appliquée aussi bien pour la reconnaissance de personnes que pour la reconnaissance d'expressions, en ayant fait le choix d'employer une paramétrisation de la surface 3D différente selon le cas de figure. Dans les deux cas, ces régions sont déterminées automatiquement à l'aide de points d'intérêts localisés par diverses méthodes de l'état de l'art.

Dans le chapitre 3, nous proposons une représentation de la surface 3D par cartes en 2D. Nos travaux sur l'approche régions au chapitre 2 ont montré des carences quand à la gestion de l'information sur l'ensemble du visage, c'est-à-dire quant à l'étude des relations entre ces régions. Si certaines méthodes ont été élaborées dans l'état de l'art pour la résolution de ce problème, nous avons préféré nous orienter sur une approche holistique. Dans un premier temps, nous présentons une méthode

de représentation de la surface 3D inspirée des mesures de courbures, ainsi que son mode de calcul. Nous montrons et évaluons ensuite ses applications aux problèmes de la reconnaissance d'expressions, et dans une moindre mesure au problème de la reconnaissance de personnes.

Le chapitre 4 présente enfin une méthodologie pour évaluer l'impact des dégradations sur les performances des algorithmes d'analyse faciale en 3D. Il s'agit d'une étude préliminaire indispensable à l'application en situation réelle des systèmes d'analyse automatique de visages en 3D. Dans un premier temps, nous définissons le type et l'origine des dégradations. Dans un deuxième temps, nous définissons un protocole permettant une évaluation quantifiée et scientifique de la robustesse des méthodes d'analyse faciale en 3D aux dégradations. Enfin, nous présentons les résultats d'une expérimentation où nous comparons les performances de différents algorithmes de l'état de l'art, en suivant le protocole que nous avons mis en place.

Chapitre 1

État de l'art

1.1 Introduction

Cet état de l'art sera organisé de la manière suivante. Dans un premier temps, nous allons donner un aperçu de l'existant en termes de capture de surfaces 3D. Nous allons également nous intéresser aux principales bases de données publiques disponibles. Dans un deuxième temps, nous allons nous intéresser à ce qui caractérise un visage, et notamment au système FACS. Nous nous intéresserons ensuite aux méthodes d'analyse de visage en 3D selon les angles d'attaque suivants : basées sur des points d'intérêt ; holistiques ; et hybrides ; en montrant à chaque fois comment ces techniques ont pu être exploitées à la fois en reconnaissance de personnes et d'expressions.

1.2 Acquisition de visages en 3D et bases de données

1.2.1 Acquisition de visages en 3D

Dans l'ensemble de ce travail de thèse, nous considérerons l'acquisition des visages en 3D comme effectuée, c'est-à-dire que nous travaillerons directement sur des images de profondeur ou des maillages 3D représentant le visage, soit de manière exclusive (visage isolé), soit comprenant des éléments de son environnement (cheveux, épaules, torse, éléments occultants, etc.). Ces modèles sont représentés en distances réelles, c'est-à-dire qu'entre deux points d'un même objet nous sommes capables de déterminer une distance en millimètres. Nous ne nous attarderons pas sur le problème de leur acquisition, qui est un sujet de recherche à part entière. Cependant, il nous est apparu important de mentionner les différentes méthodes de scan disponibles, pouvant avoir un impact sur la qualité du modèle et les traitements à lui apporter.

1.2.1.1 Type de données

Avant de s'intéresser à ces problèmes, il est important de préciser que les données sur lesquelles travaille l'ensemble de la communauté existent sous deux formes

distinctes, à savoir les maillages 3D et les images de profondeur.

- Une image de profondeur, (ou image 2.5D), correspond à une l'image, dotée de coordonnées discrètes en 2D (par convention X et Y), dont la valeur de chaque pixel est continue et correspond à la distance au capteur (coordonnée Z). Si les images de profondeur ont l'avantage d'être exploitables comme des images en 2D en assimilant la profondeur à un niveau de gris, elles ont l'inconvénient de varier selon le point de vue employé, c'est-à-dire la pose. Dans certains cas, elle présentent des auto-occultations, donc des données manquantes ou incomplètes.
- Les maillages 3D sont un nuage de points situés dans l'espace en 3 dimensions, et sont caractérisés par un ensemble d'arrêtes, qui relient chacune une paire de points, formant des faces. Leur définition est indépendante de toute projection, et les modèles ne souffrent pas du phénomène d'auto-occultations rencontré avec les images de profondeur. L'absence de paramétrisation naturelle et ou générale des maillages en 3D de visages interdit l'utilisation sans adaptation ou généralisation des algorithmes couramment employés dans le paradigme du traitement d'images en 2D.

La transformation entre ces deux représentations n'est pas bijective. S'il est toujours possible de convertir une image de profondeur en maillage sans perte d'information, l'inverse n'est pas vrai en raisons de phénomènes d'auto-occultation.

1.2.1.2 Moyens d'acquisition

De manière assez courante, l'acquisition de visages 3D se fait *via* les moyens suivants.

- **Balayage laser :** Ce type de scanner utilise un télémètre laser, qui permet de calculer la distance avec la surface étudiée. Le principe est de diriger un faisceau laser vers l'objet considéré, et de mesurer le temps aller-retour nécéssaire à une impulsion pour être réfléchie par l'objet étudié. Une précision de la mesure du temps de l'ordre de la picoseconde permet d'obtenir des relevés d'une précision inférieure au millimètre. Cette mesure est ponctuelle, ce qui contraint le scanner à effectuer un balayage, le faisceau étant alors dirigé dans une direction différente entre chaque mesure de distance, souvent à l'aide d'un système de miroirs. Le scanner laser a donc pour sortie une image de profondeur. Le fait que différents points de l'objet scanné doivent être capturés à divers instants peut être sujet à des distorsions causées par le mouvement. Ce type de scanner est également soumis à des erreurs de mesure lorsque la surface considérée est trop réfléchissante. Les effets sur la santé de scans répétés sur l'oeil font parfois l'objet de discussions. Ces scans sont généralement

FIGURE 1.1 – Capture de la topologie d'un terrain sur un lieu de fouilles archéologiques par un scanner 3D à balayage laser.

accompagnés par la prise d'une photo 2D afin de pouvoir la plaquer en tant que texture sur l'objet en 3D correspondant. Ce type de matériel est coûteux. Le Minolta vivid 900 est un exemple de scanner utilisé dans la constitution de bases de données 3D exploitées dans l'état de l'art comme FRGC v2 (section 1.2.2.1).

- **Projection de lumière structurée :** Le principe est de projeter un motif sur l'objet considéré depuis un point, et d'en observer les déformations depuis un autre. La difficulté est ici la mise en correspondance des points du motif initial sur l'objet scanné, notamment dans le cas d'occultations, de trous et plus généralement de discontinuités dans la surface étudiée. Un calibrage peut parfois se révéler nécessaire. Ce type de matériel d'acquisition a l'avantage de produire des captures quasi instantanées, évitant la distorsion de mouvement. En outre, ce type d'équipement est généralement moins coûteux que les scanners lasers.
- **Exploitation de vues 2D :** En associant entre eux les points de différentes vues 2D (photographies) d'un même objet tridimensionnel, il est possible par triangulation de reconstituer la profondeur de l'objet étudié. C'est le cas des scanners stéréoscopiques. Le problème est celui, largement répandu dans l'état de l'art, de l'appariement dense de surfaces (*dense matching*). Des al-

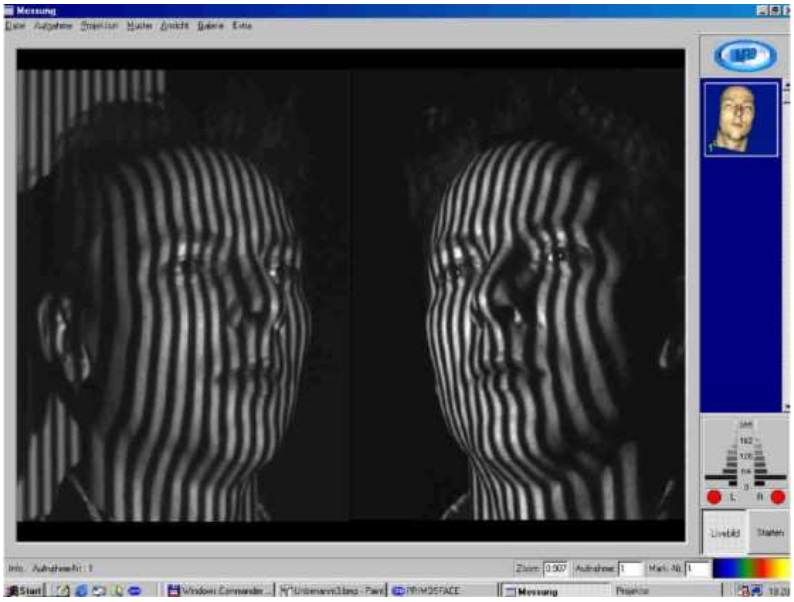


FIGURE 1.2 – Reconstruction d'un visage en 3D à l'aide d'un scanner à lumière structurée.

gorithmes comme SIFT [Lowe 2004], GLOH [Mikolajczyk & Schmid 2005] ou Daisy [Tola *et al.* 2010] ont pu être utilisés à ces fins. Dans le cas particulier de l'acquisition de visages en 3D, une connaissance approximative *a priori* de la géométrie du visage peut faciliter, comme dans le cas de l'acquisition par projection de lumière structurée, la résolution de ce problème. C'est l'objet de l'approche proposée par [Blanz & Vetter 2003]. L'avantage de cette méthode réside dans son coût et dans la simplicité de sa mise en œuvre.

Ces méthodes, s'appuyant sur des techniques variées, possèdent divers avantages et inconvénients, y compris dans la qualité du modèle obtenu. Nous verrons dans le chapitre 4 comment envisager ces variations pour le passage et l'évaluation des algorithmes d'analyse de visages en 3D à des applications pratiques.

Après s'être intéressé succintement aux méthodes couramment rencontrées pour l'acquisition de visages en 3D, nous allons nous pencher sur les bases de données existantes les plus utilisées dans le domaine de l'analyse de visages en 3D.

1.2.2 Les bases de données de visages 3D

Ces dernières années, plusieurs bases publiques ont fait leur apparition pour l'évaluation des méthodes d'analyse de visages en 3D. En fonction de l'objet pour lequel elles ont été envisagéesLeurs caractéristiques varient largement : nombre total de scansl ; nombre d'invidus différents ; nombre de scans fixe ou variable par personne ; espacement des scans d'une même personne dans le temps ; présence et

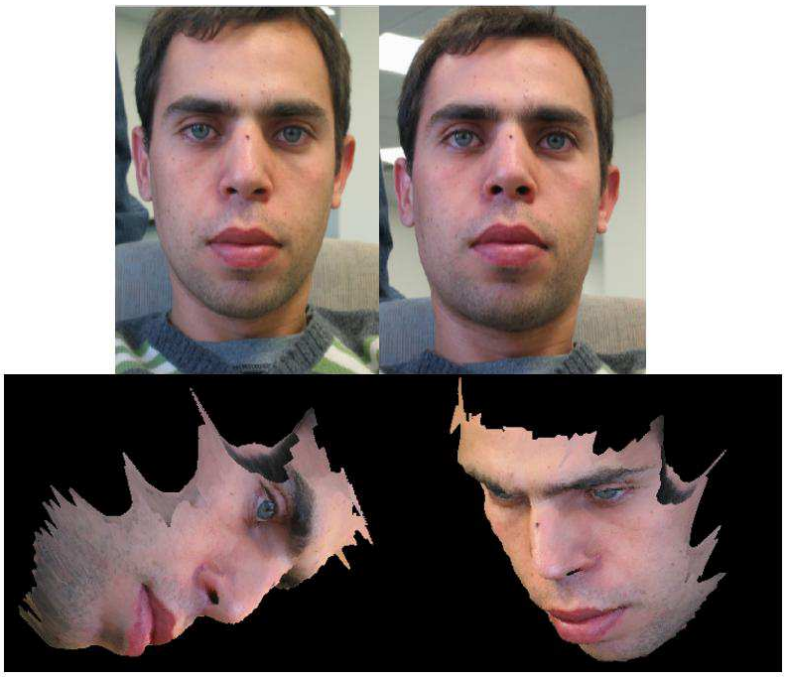


FIGURE 1.3 – Reconstruction stéréoscopique d'un visage en 3D.

intensité d'expressions faciales, maîtrisées ou non et avec quels moyens ; contraintes sur la distance au scanner ; résolution des modèles ; présence et maîtrise ou non des occultations ; captures partielles, variations de point de vue ; type de scanners employés et qualité des modèles ; présence d'annotations ; respect, maîtrise ou non d'une diversité d'âge et d'ethnie ; contribution volontaire, sciente ou non des sujets scannés : autant de critères qui caractérisent les bases de données, et qui ont leur importance au moment d'évaluer les algorithmes d'analyse faciale en 3D au regard d'un problème donné.

Nous citons dans les sections suivantes les bases les plus employées récemment dans l'état de l'art dans le domaine de l'analyse de visages en 3D.

1.2.2.1 FRGC

FRGC, pour Face Recognition Grand Challenge [Phillips *et al.* 2005], est une base de visages 3D publique crée par l'université de Notre-Dame. Elle est divisée en deux versions.

La version 1.0, parfois appelée UND, comprend 953 visages scannés correspondant à 277 personnes différentes. Tous les visages sont pris de face, avec une expression neutre, sans lunettes mais sans précaution particulière au niveau des cheveux ni des vêtements. Le nombre de modèles par personne n'est pas constant. Cette base inclut différents groupes ethniques, et différentes classes d'âge, bien qu'on

constate une prédominance de jeunes personnes entre vingt et trente ans. La parité homme femme est assez bien respectée. Les visages ont étés capturés par un scanner de type Minolta Vivid 700, et les visages sont fournis sous forme d'images de profondeur accompagnés d'une photo 2D. La qualité des visages pose souvent problème si elle est exploitée sans prétraitement, avec la présence de pics et de trous au niveau des parois nasales, de la bouche, des yeux et des parties chevelues du visage (barbe, cils, cheveux, moustache).

La version 2.0 contient 4007 visages correspondant à 466 personnes différentes, scannées à deux périodes différentes (automne 2003 et printemps 2004), selon les mêmes principes que la version 1.0 (pas de lunettes, vêtements et coiffures quelconques). Ici aussi, le nombre de scans par individu est variable, la parité et la diversité ethnique sont globalement bonnes. Certains visages sont expressifs et d'autres non, sans contrainte précise autre que la présence d'au moins un visage neutre par personne. Le scanner employé est un Minolta Vivid 900/910. Les visages comprennent nettement moins de défauts que la version 1.0, bien qu'on puisse encore constater quelques modèles distordus.

L'ensemble des personnes ayant contribué à la version 1.0 est en grande partie inclus dans la version 2.0. Le grand nombre de personnes différentes et de modèles présents fait de FRGC, et notamment FRGC 2.0, une base de référence pour l'évaluation des performance des algorithmes de reconnaissance de personnes.

1.2.2.2 Bosphorus

La base Bosphorus [Savran *et al.* 2008] comprend 4666 scans correspondant à 105 personnes différentes. C'est une base extrêmement contrainte au niveau de la pose, de l'expression et des occultations. La base est constituée de 60 hommes et 45 femmes, parmi lesquels 29 acteurs professionnels, majoritairement de type caucasien. La base est annotée manuellement d'origine, avec 24 points de repère sur le visage selon des critères anatomiques, et le pourtour du visage découpé (pas de cheveux ni vêtements visibles, sauf si spécifié). 18 hommes portent la moustache ou la barbe, et 15 autres sujets ne présentent pas un rasage de près. Pour chaque personne, nous avons à disposition :

- entre 1 et 3 visages neutres,
- un scan pour chaque expression prototypique (cf section 1.3)
- jusqu'à 28 scans correspondant à l'activation d'une ou d'un groupe d'Action Units (cf section 1.3)
- des rotations très contraintes au niveau de la précision angulaire : entre le visage de face et la vue de profil (rotation autour de l'axe vertical, en anglais *yaw*, geste de négation), et autour de l'axe des X (la personne hoche la tête

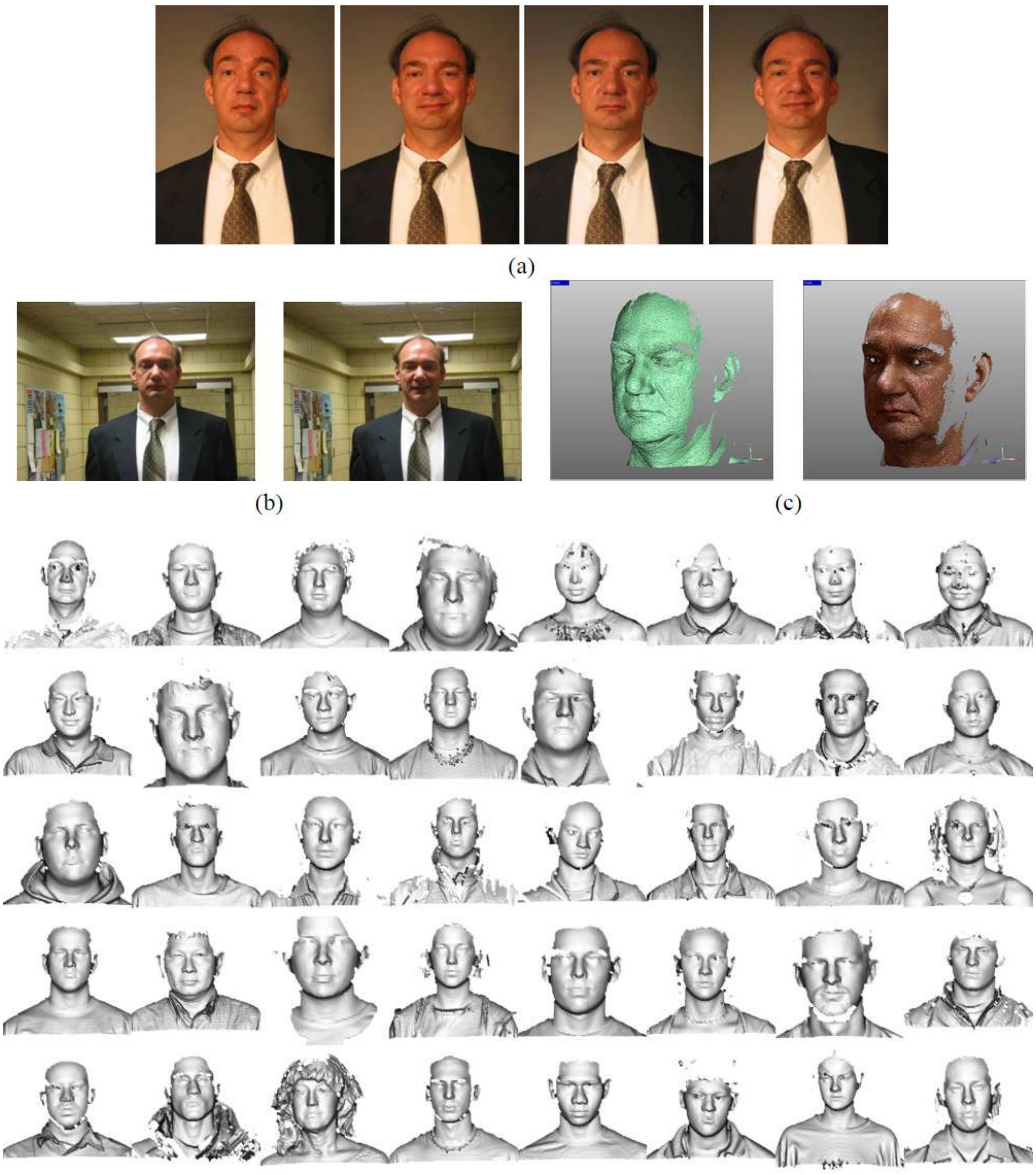


FIGURE 1.4 – Type de captures présentes dans la base de données FRGC. La partie supérieure montre des expressions contraintes en (a), non contraintes en (b), et les scans 3D correspondants en (c). La partie inférieure montre un échantillon des visages visibles dans la base en 3D pure.

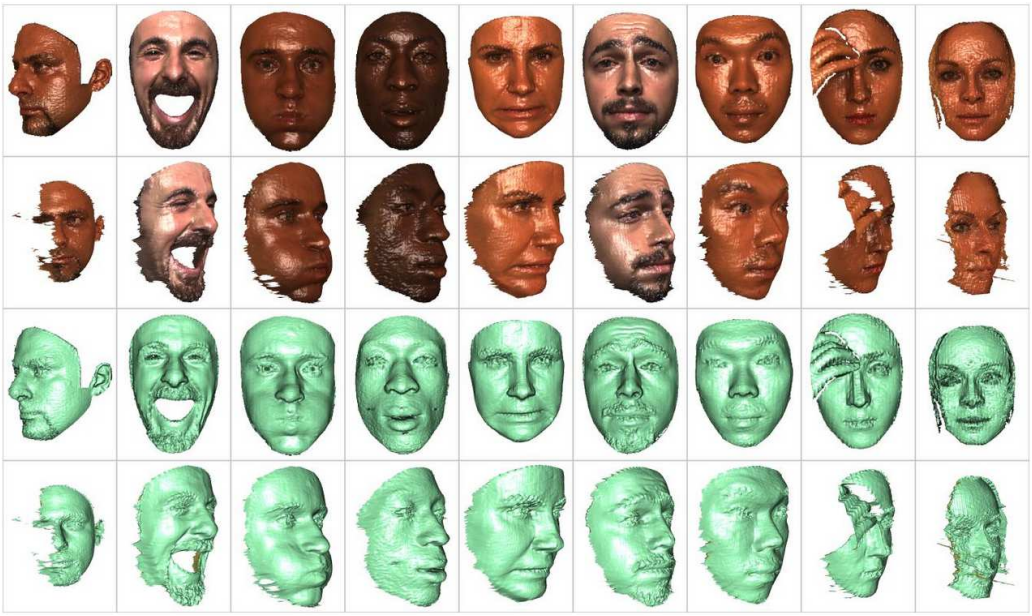


FIGURE 1.5 – Captures extraites de la base Bosphorus. Chaque colonne correspond à un même scan, avec et sans texture et sous un angle différent. Les expressions ainsi que les occultations sont contraintes.

de haut en bas, correspond au geste d'acquiescement, *pitch* en anglais), ainsi qu'une combinaison des deux gestes.

– des occultations elles aussi contraintes, à l'aide de la main sur des régions précises (œil droit, œil gauche, bouche) ainsi que le port de lunettes.

La base est fournie sous la forme d'images de profondeur, de maillages 3D qui en sont extraits, accompagnés de texture. Ces images ont été obtenues à l'aide de la projection de lumière structurée et d'un capteur unique.

Cette base est dans sa constitution très complète pour l'étude des expressions faciales et leur impact sur la reconnaissance de personnes, de même pour les variations de pose et les occultations. Cependant, si les modèles ne présentent pas de discontinuité telle que des pics ou des trous (en dehors de la bouche lorsqu'elle est ouverte) ni de distorsion de mouvement, d'importantes ondulations sur la surface 3D sont observables, notamment dans les régions chevelues du visage, mais aussi autour des régions avec une forte incidence comme les contours du nez.

1.2.2.3 BU-3DFE

BU-3DFE (pour Binghamtom University 3D Facial Expression) [Yin *et al.* 2006] est une base de données comprenant 2500 modèles pour 100 personnes différentes (56 femmes et 44 hommes). Elle est concentrée sur les expressions du visage et, à ce titre, est très contrainte. 25 scans dans une position parfaitement frontale sont disponibles

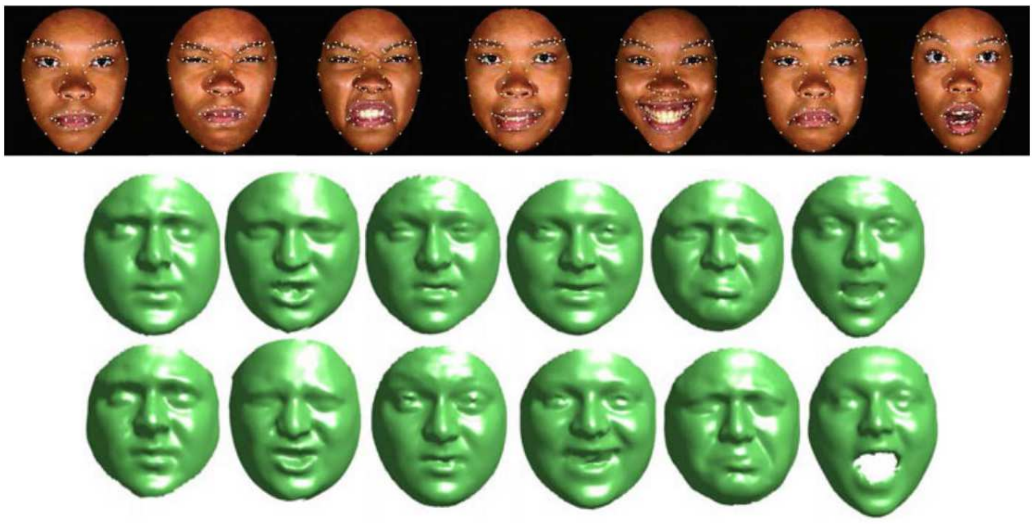


FIGURE 1.6 – Captures de la base BU-3DFE. En première ligne et de gauche à droite, les captures texturées d'une même personne correspondant aux expressions Neutre, Colère, Dégoût, Peur, Joie, Tristesse et Surprise au niveau maximum d'intensité. Les 83 points d'intérêt annotés manuellement de la base sont indiqués en blanc. En seconde et troisième ligne, les mêmes expressions moins le visage neutre, avec l'intensité minimum (2ème ligne) puis l'intensité maximum (troisième ligne).

pour chaque personne, incluant 1 modèle neutre et 4 modèles de chacune des 6 expressions prototypiques, correspondant à 4 niveaux d'intensité différents. L'origine ethnique des personne est variée et annotée. Les auteurs revendiquent également une diversité au niveau de l'âge des participants allant de 18 à 70 ans, bien que d'après nos observations une très large majorité des candidats semble être dans la fourchette de 20 à 30 ans. Les visages sont fournis au format VRML et sont disponibles en deux versions (modèle *RAW* ou recadré), accompagnés de l'annotation manuelle de 83 points d'intérêt définis sur une base anatomique. Deux photos de chaque côté prises avec un angle de 45 °, en plus d'une photo de face, sont également fournies. La qualité des visages est excellente.

Cette base est logiquement utilisée en reconnaissance d'expressions du visage.

Elle est accompagnée d'une base en 3D dynamique nommée BU-4D [Yin *et al.* 2008], dont nous ne nous servirons pas dans les travaux exposés dans cette thèse.

1.2.2.4 Autres bases

D'autres bases, moins utilisées dans les travaux récents de l'état de l'art, existent. Citons, entre autres, GavabDb [Moreno & Sánchez 2004] ; IV² ; MSU ; FSU ; 3D_RMA [Grgic & Delac 2012].

FIGURE 1.7 – Quelques exemples de *Charakterköpfe* sculptées par Franz Xaver Messerschmidt, mettant en avant (et exagérant) l'expressivité du visage humain.

1.3 Caractérisation du visage humain, et système FACS

Le visage humain et sa perception ont été un important sujet d'étude de la part des anatomistes et dans les sciences cognitives. Citons par exemple le modèle de Bruce et Young [Bruce & Young 1986], qui montre que le visage humain provoque chez l'homme un stimulus spécifique et distinct des autres tâches de reconnaissance d'objets. Ce diagnostic est appuyé par l'existence de pathologies affectant la perception des visages comme l'autisme ou surtout la prosopagnosie, cette pathologie affectant exclusivement la reconnaissance de visages. Une attention toute particulière est également prêté au visage dans les domaines artistiques et religieux. Le portrait est un thème majeur en peinture, dessin, photographie, sculpture. Certains artistes ont porté un soin tout particulier à la reproduction des expressions du visage. Citons Franz Xaver Messerschmidt et sa série des *Charakterköpfe* (les *têtes de caractère*). Le maquillage, les masques, les voiles ou les cagoules relèvent d'usages et de symboliques spécifiques à travers les cultures.

De manière générale, on distingue les traits invariants du visage, qui correspondent à l'identité, aux traits variants du visage, qui correspondent à l'expression, et qui sont liés à l'action des muscles faciaux. On compte environ 44 muscles faciaux, qui trouvent leur origine au niveau des os du crâne, et qui se fixent sur les tissus mous de la peau du visage. Les os du massif facial sont au nombre de 14, mais un seul d'entre eux est mobile de manière visible par rapport au reste, il s'agit de la mandibule.

La connaissance du modèle anatomique et musculaire de l'humain peut par exemple se révéler utile pour l'informaticien en animation, où il est possible de décomposer une expression en une somme d'actions musculaires, et de s'appuyer sur une modélisation de la boite crânienne. En analyse faciale, seuls les effets de telles actions musculaires sont visibles, et nous avons difficilement accès à la topologie osseuse ainsi qu'à l'implantation des tissus mous. Il apparaît en effet délicat de modéliser les os du massif facial à partir d'une photo du sujet de son vivant. Nous

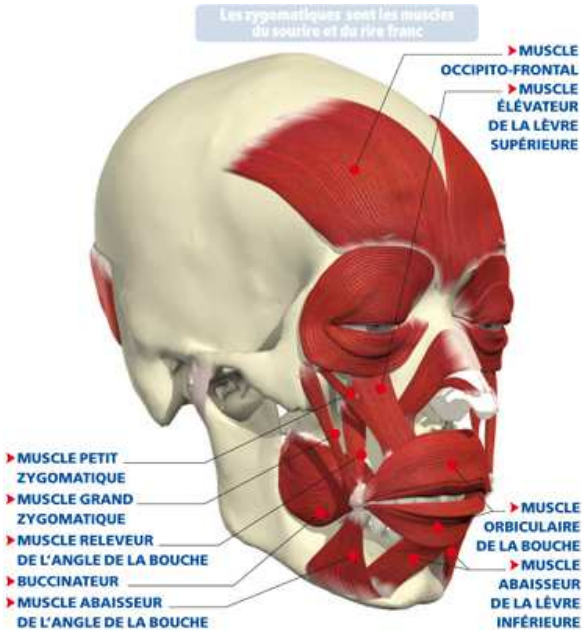


FIGURE 1.8 – L'ensemble des muscles intervenant dans l'expression du sourire.

sommes contraints, si nous voulons étudier l'action des muscles et les déformations qu'ils engendrent sur la surface faciale, de rester à un niveau d'observations, et non pas de modélisation, de l'activité de ces muscles.

En 1978, Ekman et Friesen ont introduit le *Facial Action Coding System* (FACS) [Ekman *et al.* 2002]. Il s'agit d'une nomenclature qui décompose les mouvements du visage en *Action Units* (AU). Plutôt que de s'intéresser à l'action de muscles faciaux activés de façon unitaire, ce système s'intéresse à l'activation de ces muscles par groupes, leur effet étant décrit de manière simple. Par exemple, l'AU 26 correspond à l'abaissement de la mâchoire inférieure et l'AU 2 à l'action de lever les deux sourcils extérieurs. Certaines expressions, caractérisée par l'activation d'un certain nombre d'AU, ne sont pas perçues de la même manière selon les ethnies. Le clin d'œil en est un exemple. Dans les cultures occidentales il s'interprête comme un signe de connivence ou de complicité, tandis qu'en Afrique de l'Ouest, il peut s'agir d'une demande des parents à leurs enfants de quitter la pièce. Son interprétation va de la demande romantique en Amérique du Sud à la grossièreté dans certaines régions d'Asie. Ekman a relevé, en plus de l'expression neutre, 6 expressions universelles, c'est-à-dire exécutées et perçues de manière identique quelle que soit l'ethnie [Ekman & Friesen 1971] :

- la joie,
- la surprise,
- la peur,

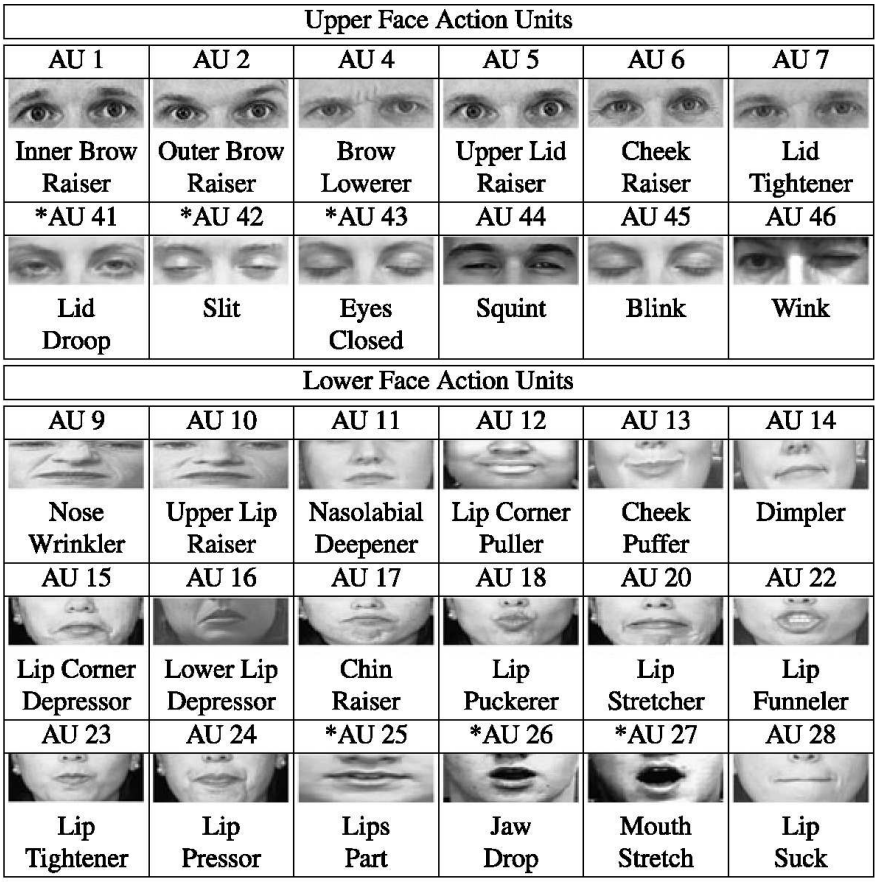


FIGURE 1.9 – Un exemple de l'effet d'activation d'AU sur le visage. Ces actions correspondent à la contraction ou à la détente d'un ou plusieurs muscles faciaux, et peuvent être décrites textuellement de manière simple. Elles peuvent rarement être interprétées émotionnellement en tant que telles, les émotions étant plus souvent décrites comme la composition de plusieurs AUs.

- le dégoût,
- la colère,
- la tristesse

Notons qu'Ekman, bien qu'affirmant l'universalité de ces expressions, a relevé pour certaines d'entre elles des variations dans la manière de les exécuter en termes d'AU activées.

De fait, le problème de la reconnaissance automatique d'expressions faciales a souvent été ramené dans l'état de l'art à celui de reconnaître ces 6 expressions prototypiques. C'est cette nomenclature qui a été adoptée dans les bases de données évoquées dans la partie 1.2.2.

Notons toutefois que l'universalité, et plus précisément l'universalité de la perception de ces 6 expressions a fait récemment l'objet de discussions

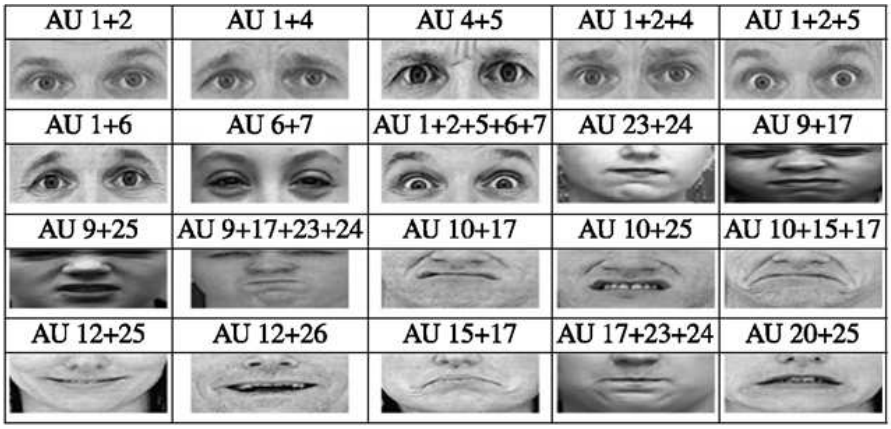


FIGURE 1.10 – Quelques exemples d'AUs activées simultanément.

[Jack *et al.* 2012]. Cette liste a par ailleurs subi des ajouts et des amendements [Matsumoto *et al.* 2009]. Dans ce travail, ainsi que dans l'ensemble de l'état de l'art dans lequel il s'inscrit pour l'instant, nous ferons abstraction de ce débat et nous nous confinerons à l'étude des 6 expressions prototypiques telles que proposées par Ekman.

L'utilisation du FACS dans l'analyse faciale peut se faire sous plusieurs angles. Pour la reconnaissance d'expressions faciales, elle permet de décomposer le problème de la reconnaissance d'expressions en un sous-problème potentiellement plus simple, qui est celui de la détection d'AU activées. Pour la reconnaissance faciale, la soustraction de l'ensemble des AU activées au visage permet de ne ramener le problème qu'à l'étude des traits invariants du visage.

1.4 Analyse de visages en 3D

Dans la suite de cet état de l'art, nous allons nous intéresser simultanément aux problèmes que sont la reconnaissance et l'identification de personnes, et la reconnaissance d'expressions faciales, sur la base de leurs modèles 3D.

Il est important de préciser le scénario général. Dans les deux problèmes, on suppose l'acquisition d'un modèle de visage en 3D dit inconnu (*probe*). Selon les cas, soit son identité, soit son expression, soit plus couramment les deux sont inconnus. Dans un scénario de reconnaissance de personnes, le système dispose en plus d'un ensemble de visages galerie *gallery* connus, dont on a préalablement effectué l'*enrolement*.

Dans le cas d'une vérification, le système dispose de l'identité de la personne qu'elle suppose avoir scannée (via un badge, un nom, un numéro de code), et doit vérifier que les marqueurs biométriques concordent avec cette identité. Un exemple

de scénario d'application possible est celui d'un portique d'entrée sécurisé, où l'utilisateur s'est enregistré à l'aide d'un badge ou de son passeport puis passe devant un scanner afin de vérifier qu'il ne s'agit pas d'une usurpation d'identité.

Dans le cas d'une reconnaissance, le système n'a pas de connaissance *a priori* de la personne scannée, le but est donc de l'identifier parmi la galerie de personnes connues. Un scénario d'application typique est celui de l'identification d'un délinquant parmi une liste de suspects, à partir d'une image extraite d'une séquence filmée par une caméra de surveillance par exemple.

Les indicateurs de performance vus couramment en reconnaissance faciale 3D, et plus généralement en biométrie sont les suivants :

- le *False Accept Rate* (FAR), ou taux de fausses acceptations : il est défini comme la probabilité que le système associe le visage inconnu à un visage ne correspondant pas dans la galerie. Il correspond au cas où un imposteur parvient à se faire passer pour une personne possédant un accès légitime.
- le *False Reject Rate* (FRR), ou taux de faux rejets : il correspond à la probabilité que le système identifie le visage *probe* comme un imposteur, alors qu'il est légitime.
- la *Receiver Operating Characteristic* (ROC), ou courbe ROC. Les deux mesures précédemment énoncées sont le résultat d'un compromis qui est fait entre la facilité d'accepter un visage comme légitime ou de le considérer comme un imposteur. Ce compromis se matérialise le plus souvent comme un seuil. La courbe ROC est la visualisation de ce compromis.
- l'*Equal Error Rate* (EER) : le taux auquel les FAR et FRR sont égaux. Il peut être facilement déduit de la courbe ROC. D'une manière générale, plus l'EER est faible et plus le système est performant.

Plus spécifiquement pour le problème de la reconnaissance faciale en 3D, on rencontre les indicateurs suivants :

- le *Rank-One Recognition Rate* : il correspond au taux de reconnaissance obtenu lorsque le FRR est nul. Pour mieux l'interpréter, il correspondrait à un scénario de reconnaissance où l'on partirait du principe que tout visage *probe* testé aurait déjà été enregistré dans la la galerie.
- le *Verification Rate at 0.1% FAR* : il correspond au taux de reconnaissance obtenu lorsque le FAR est de 0,1%. Cette mesure est couramment usitée dans les papiers de l'état de l'art en reconnaissance faciale en 3D.

En reconnaissance d'expressions en 3D, le scénario le plus courant est celui d'un visage *probe* sur lequel nous faisons l'assomption qu'il est expressif. Le but est alors

d'identifier la bonne expression. Les performances sont présentées sous la forme d'un taux de reconnaissance, ainsi que sous la forme d'une matrice de confusion. Les lignes de cette matrice représentent les occurrences des classes réelles et les colonnes représentent les occurrences des estimations. Un système de reconnaissance idéal ne devrait contenir des occurrences que dans sa diagonale.

1.5 Utilisation de points d'intérêt

De manière assez intuitive, l'utilisation de points clés de l'anatomie du visage semble présenter un intérêt au sens de l'analyse faciale. Elle s'inscrit également dans un contexte historique : un des tout premiers systèmes biométriques mis en place est le système Bertillon, qui mesure les angles et les distances entre différents points de l'anatomie humaine ; la reconnaissance d'empreintes digitales utilise largement la topologie des dermatoglyphes, dont la nature est ponctuelle. Le principal usage de points-clés de l'anatomie serait celui de leur relation spaciale (distance entre les yeux, taille du nez, ouverture de la bouche, angles entre le bout du nez, les coins des yeux et le lobe de l'oreille, etc), tant pour l'identification de personnes que la reconnaissance d'expressions faciales.

Cependant, nous allons le voir dans cette partie, il est possible d'extraire de tels points-clés de manière automatique, indépendamment d'une notion anatomique ou relative de manière évidente à la géométrie du visage (que ce soit en 2D ou en 3D). Dans un premier temps, nous allons nous intéresser aux points d'intérêt définis de manière anatomique, avant de nous intéresser aux points dits caractéristiques saillants.

1.5.1 Points anatomiques et fiduciaux

Les méthodes qui vont suivre s'appuient sur des points clés de l'anatomie, et leurs relations spaciales, pour effectuer leur tâche de reconnaissance. Dans cette partie, les points d'intérêt employés peuvent être exprimés, définis et localisés de manière relativement stable avec le langage, et sont intuitivement compréhensibles. Par exemple, le coin de la narine droite, le milieu du sourcil droit, le menton, etc. Dans la littérature, ils sont généralement référencés sous le nom de *landmarks*. Le plus souvent, ils sont localisés à l'aide d'un opérateur humain. Il existe cependant quelques méthodes comme [Zhao *et al.* 2009b], [Gupta *et al.* 2010] ou [Szeptycki *et al.* 2009] qui permettent d'en localiser automatiquement un certain nombre, avec plus ou moins de précision, l'annotation de l'opérateur humain servant de référenciel. La plupart du temps, l'état de l'art emploie des *landmarks* manuels en présupposant que la localisation automatique de tels points sera disponible à l'avenir. Toutefois, ce problème

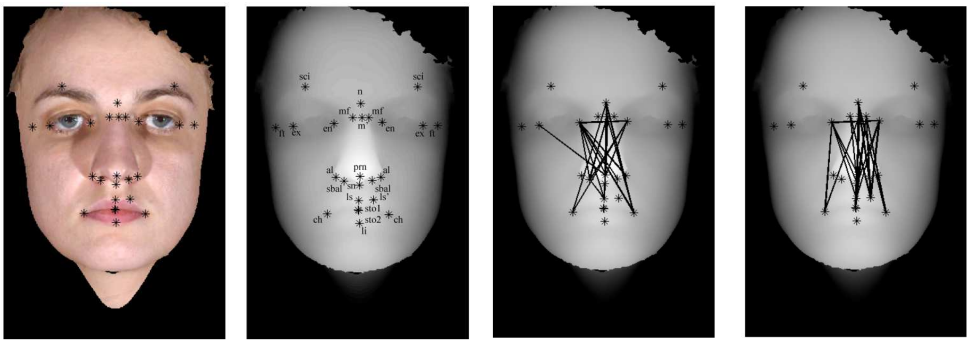


FIGURE 1.11 – La méthode proposée dans [Gupta *et al.* 2007] exploite les distances entre 25 points anatomiques du visage en 3D. Les 2 visages à gauche montrent l'emplacement de ces 25 points. Le 3ème visage montre les 20 distances euclidiennes les plus discriminantes, et le 4ème visage les 20 distances géodésiques les plus discriminantes.

reste aujourd'hui difficile. Ces méthodes de localisation automatique sont évoquées dans la partie 1.8.

1.5.1.1 En reconnaissance de personnes

[Wu *et al.* 2004] ont proposé l'emploi de *Local Shape Map* (*LSM*) pour décrire un point sur la surface faciale, et le représentent sous forme d'un histogramme 2D construit à partir de la projection des coordonnées 3D des points du voisinage du point délimité par une sphère. Le coefficient de corrélation est utilisé comme mesure de similarité entre la LSM de chaque paire de points, puis ces résultats sont classés à l'aide d'un système de vote. Cette méthode ne requiert pas l'emploi d'une technique de recalage.

Le travail de [Gupta *et al.* 2007] est basé sur l'emploi de 25 *landmarks* et utilise la distance euclidienne et la distance géodésique entre chaque paire de points comme vecteur pour représenter le visage (figure 1.11). Ces vecteurs sont ensuite classés *via* la méthode *Linear Discriminant Analysis* (LDA).

En 2008, [Castellani *et al.* 2008] ont introduit une méthode d'apprentissage générative en adaptant les Modèles de Markov Cachés (*Hidden Markov Model*, *HMM*) aux maillages 3D. La géométrie du voisinage d'un *landmark* est apprise par HMM, définissant une signature invariante aux changements de pose. Une telle description autorise l'appariement en comparant la signature des points correspondants grâce à une Estimation du Maximum de Vraisemblance. Les données d'appariement sont finalement utilisées dans un système de vote pour établir la classification.

[Daniyal *et al.* 2009] font l'apprentissage d'un *Point Distribution Model* (*PDM*) en utilisant des *landmarks* manuels sur 100 modèles de référence, puis localisent ces

points correspondant automatiquement sur les visages de test. Les distances euclidiennes entre chaque paire de points sont compilées dans un vecteur caractéristique, lequel est classé *via* LDA.

1.5.1.2 En reconnaissance d'expressions

Ces points anatomiques et leur relation spaciale ont également été exploités dans [Soyel & Demirel 2007] pour la reconnaissance d'expressions. Dans ce travail, les auteurs ont utilisé 6 distances caractéristiques à partir de 11 points fiduciaux comme entrées dans un réseau de neurones, pour classer 7 expressions (les 6 expressions prototypiques ainsi que le visage neutre).

Dans [Tang & Huang 2008], les auteurs ont suggéré l'emploi d'une méthode de sélection automatique de descripteurs, obtenue à partir de la distance euclidienne normalisée entre points d'intérêt. La sélection de descripteurs et la classification est effectuée à l'aide d'un algorithme du type Adaboost multi-classes.

Les auteurs de [Berretti *et al.* 2010a] ont proposé l'utilisation des descripteurs SIFT pour caractériser le voisinage des points d'intérêt. Dans leurs travaux, ils prennent en compte une partie des points d'intérêt annotés manuellement de la base BU-3DFE, ainsi qu'un ensemble de points d'intérêt extraits automatiquement à partir des précédents, à l'aide de règles géométriques simples. Le descripteur de l'ensemble de ces données étant de dimension 14336, ils opèrent une réduction de la dimensionnalité à l'aide d'un opérateur statistique nommé *minimal redundancy maximal-relevance* (mRMR). Enfin, ils effectuent la classification des expressions à l'aide de *multi-class SVM*.

Dans [Zhao *et al.* 2010], les auteurs utilisent un système baptisé SFAM pour la localisation automatique de points d'intérêt, s'appuyant à la fois sur les données 2D et 3D. Les positions relatives de ces *landmarks* sont ensuite classées à l'aide de fonctions de croyance.

Que ce soit en reconnaissance faciale ou en reconnaissance d'expressions, ces méthodes, au principe très intuitif, présentent l'inconvénient de représenter le visage de manière très ponctuelle. Elles reposent également sur l'utilisation de points d'intérêt anatomiques, dont la localisation automatique est aujourd'hui encore délicate. Elles sont ainsi dans la plupart des cas difficiles à mettre en œuvre dans un scénario d'application complètement automatisé.

1.5.2 Par points caractéristiques saillants.

En traitement et en analyse d'images, l'extraction et l'utilisation de points d'intérêt peuvent être indépendantes d'une interprétation humaine intuitive. Un exemple

emblèmatique en est l'algorithme d'extraction de points d'intérêt SIFT (pour *Scale Invariant Feature Transform*) [Lowe 2004], dont les points d'intérêt extraits ne semblent pas toujours présenter de signification visuelle pour l'humain.

À l'instar de ces méthodes, on trouve en analyse de visages en 3D un état de l'art présentant des caractéristiques similaires. Nous opposerons ainsi aux méthodes faisant usage de points d'intérêt anatomiques (y compris celles faisant usage de points d'intérêt anatomiques localisés automatiquement) celles faisant usage de points dits caractéristiques saillants, sans la base de connaissances anatomiques ou anthropométriques.

À notre connaissance, ce type de points d'intérêt n'a pas été utilisé en reconnaissance d'expressions. En effet, dans les méthodes qui suivent, les points caractéristiques sont ici déterminés de sorte à mettre en évidence des spécificités locales de la surface faciale. Ils sont ainsi très adaptés au problème de la reconnaissance de personnes, où les auteurs espèrent caractériser un ensemble de points du visage différent d'une identité à l'autre. Leur localisation n'est donc généralement pas stable d'une personne à l'autre, et il n'est *a priori* pas possible de les associer de manière déterministe à un muscle facial, ni même à une *Action Unit*.

Les méthodes décrites dans cette section concernent donc uniquement le cas de la reconnaissance faciale.

Dans [Mian *et al.* 2008] (figure 1.12), des points d'intérêt sont extraits automatiquement à l'aide d'une métrique s'apparentant à une mesure de courbure, et mettant en évidence des vallées ou des crêtes (*valley* ou *ridge* selon les termes employés pour la classification HK). La surface au voisinage de ces points est alors décrite dans le repère local, en projetant la surface sur une grille de 20x20 pixels pour en extraire une image de profondeur, cette image étant elle-même ramenée à un vecteur de dimension inférieure grâce à une Analyse en Composantes Principales (ACP) et un apprentissage effectué sur la galerie de visages neutres. Les points d'intérêt sont enfin appairés à la manière de SIFT, et une triangulation est générée afin de comparer les graphes obtenus selon diverses caractéristiques : longueur moyenne d'une arête, distance entre graphes *gallery* et *probe* après recalage rigide, etc.

[Huang *et al.* 2011] proposent de représenter le visage 3D sous forme de cartes en 2D, extraites à partir de l'image de profondeur. Les nombreuses cartes sont générées à l'aide de variantes de LBP (*Local Binary Pattern*), en faisant varier dans ce filtre le voisinage (distance et nombre de points), et l'ordre des composantes binaires. SIFT est ensuite appliqué pour l'extraction de points d'intérêt puis leur appariement. Les scores de similarité (typiquement, le nombre de points appairés) de chaque carte sont ensuite fusionnés. L'auteur montre la pertinence de telles cartes en comparaison avec l'algorithme SIFT appliqué directement à l'image de profondeur.

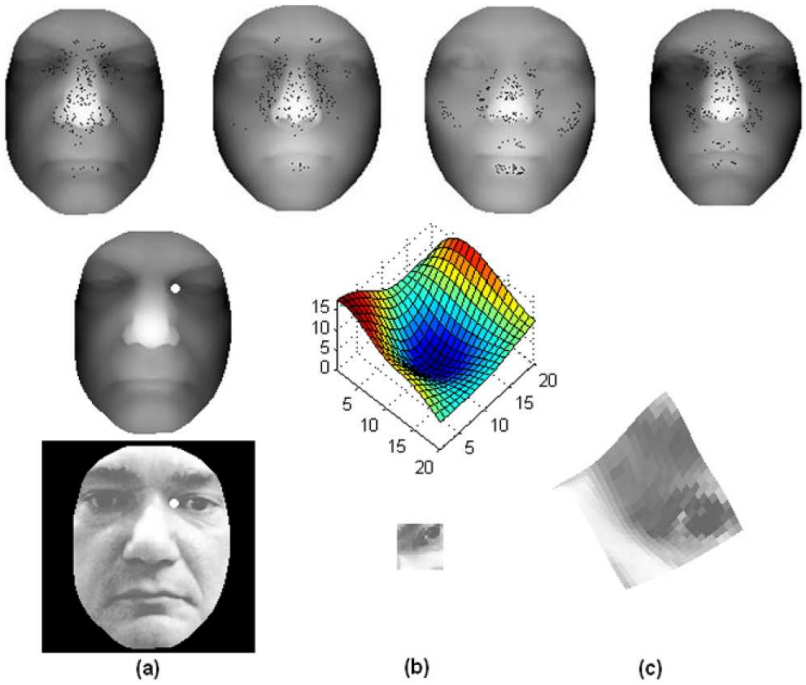


FIGURE 1.12 – Un exemple de l'utilisation de points caractéristiques dans la méthode de [Mian *et al.* 2008]. La première ligne montre la localisation de ces points caractéristiques sur le visage. Les 2 lignes suivantes montrent la manière dont ces points sont décrits, dans le repère local, à la fois au niveau de leur texture et de leur topologie en 3D.

Dans [Maes *et al.* 2010], les auteurs ont étendu l'algorithme SIFT au maillage 3D, et l'ont baptisé Mesh-SIFT. Au contraire de l'information classique de gradient extraite par l'opérateur original, mesh-SIFT encode les histogrammes locaux de valeurs de shape-index. Les points d'intérêt extraits sont des minima et des maxima locaux de la courbure moyenne (calculée à partir du maillage) à plusieurs échelles.

À noter que ces méthodes, bien qu'étant ici classées comme locales, sont parfois classées comme holistiques (cf. partie 1.6), en considérant que le traitement est appliqué de manière indistincte à l'ensemble de la surface faciale.

Un des avantages de ce type de méthode est son efficacité sur le plan du stockage, un visage étant assimilé à un ensemble de points aux caractéristiques restreintes. Une critique que l'on peut néanmoins formuler à l'encontre de ces méthodes est leur potentielle complexité calculatoire. En effet, en supposant que N visages soient représentés par un ensemble de M points caractéristiques chacun, il sera nécessaire d'opérer $N \times M \times M$ comparaisons point à point pour déterminer l'appariement d'un seul visage à l'ensemble de la base de données. En contrepartie, ces opérations sont habituellement très simples.

1.6 Appariement de surfaces et méthodes holistiques

Ces méthodes basées sur la localisation de points d'intérêt ou le voisinage de points d'intérêt sont en fait complémentaire des méthodes holistiques. Pour reprendre la définition de [Smuts 1926] et l'article de wikipedia correspondant, l'holisme est :

> la tendance dans la nature à constituer des ensembles qui sont supérieurs à la somme de leurs parties.

En terme d'analyse faciale en 3D, celà reviendrait à envisager la surface 3D comme un ensemble, muni de sa topologie, en opposition aux méthodes précédemment exposées qui assimilent la surface faciale à un ensemble de points d'intérêt ainsi que leur voisinage, et limitant la comparaison entre visages à une comparaison entre surfaces locales.

1.6.1 Iterative Closest Point

Dans cette catégorie, la méthode généralement reconnue comme *baseline* est Iterative Closest Point (ICP) [Zhang 1992] (figure 1.13). Il s'agit d'une méthode de recalage de surfaces rigides, fournissant un score de similarité qui est en fait la distance au sens des moindres carrés entre les deux surfaces appairées. Le principe est de calculer la transformation rigide, notée (R, t) (Rotation, Translation) qui,

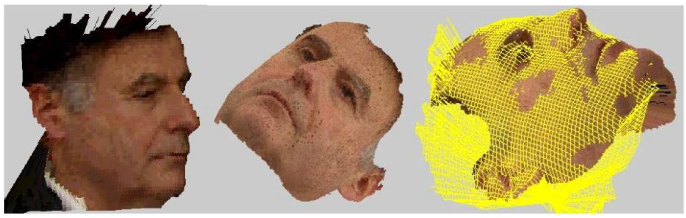


FIGURE 1.13 – À droite, un exemple de recalage ICP, extrait de [Lu *et al.* 2004]. Le visage de référence (au milieu) est représenté avec sa texture, le visage recalé (à gauche) est représenté *via* son maillage, en jaune sur l'image de droite.

appliquée au nuage de points X, minimise sa distance avec le nuage de points P. Cet algorithme a été employé à l'origine pour la reconstitution d'objets 3D à partir de vues partielles.

Concrètement, le fonctionnement de l'algorithme est itératif. Chaque itération suit le schéma suivant :

- À chaque point de X ou d'un sous-ensemble de X, on lui associe le point le plus proche de P selon la mesure de distance choisie.
- On calcule la transformation rigide (rotation et translation) appliquée à X qui minimise la distance pour l'ensemble des paires précédemment déterminées.
- On applique cette transformation. La distance entre P et X est obtenue en faisant la moyenne des distances point à point entre chaque point de X et son plus proche correspondant sur P.

Dans le cas général, la transformation est calculée à l'aide d'une *SVD* (Décomposition en Valeurs Singulières) appliquée à la matrice de corrélation déduite des paires de points appairés.

Comme on peut le voir, cet algorithme correspond à la recherche d'un minimum local à chaque nouvelle itération. [Zhang 1992] ont démontré sa convergence. Cet algorithme est cependant, par nature très dépendant des conditions d'initialisation.

Diverses variantes ont été proposées, comme la disqualification des paires associant un point en commun ; l'utilisation d'un sous-ensemble, tiré aléatoirement ou non, de points pour constituer les paires sur lesquelles sont basées le recalage ; l'utilisation de mesures de distance autres que la distance euclidienne ; etc.

Cet algorithme a été adapté à de nombreuses reprises pour la reconnaissance de personnes en 3D. Citons notamment [Medioni & Waupotitsch 2003], [Lu *et al.* 2004], [Wang *et al.* 2006b].

ICP, ainsi que ses dérivées, est en fait une heuristique et fournit un résultat non-optimal. Une de ses grandes faiblesses, qui justifie l'existence de nombreuses variantes, est sa sensibilité aux conditions d'initialsation. Il s'agît par ailleurs d'une technique de recalage rigide, étant donc sensible, dans sa forme initiale, simultané-

ment et sans distinction à l'identité et à l'expression. Il est donc délicat de l'utiliser en l'état dans un scénario de reconnaissance d'expressions classique (c'est-à-dire indépendamment de l'identité).

1.6.2 Représentation par un sous-espace

Les techniques de représentation du visage 3D par sous-espace, telles que pratiquées fréquemment dans le domaine de la 2D, ont été également expérimentées sur les images de profondeur de visages en 3D. Citons notamment les populaires Analyse en Composantes Principales (ACP, ou *PCA* en anglais) ou Analyse Discriminante Linéaire (*LDA*). [Achermann *et al.* 1997] ont testé l'approche *eigenface* sur des images de profondeur. En 2003, [Hesher *et al.* 2003] ont employé l'ACP sur diverses tailles d'images et avec un nombre varié de vecteurs propres. En 2004, [Heseltine *et al.* 2004] ont testé diverses mesures de distance entre *eigenfaces* comme mesure de similarité, pour en conclure que la distance de Mahalanobis est supérieure aux distances euclidienne et *arccosinus*. Dans [Heseltine *et al.* 2004], les mêmes auteurs ont conduit une expérimentation similaire en représentant leurs visages via *LDA*.

Notons également l'approche de [Mpiperis *et al.* 2007b] (figure 1.14) pour la reconnaissance de visages en 3D, qui effectue d'abord un mappage du visage en 3D sur une surface en 3D à l'aide d'une représentation polaire géodésique, centrée sur le bout du nez. Les méthodes appliquées au visage sur la surface en 2D sont par la suite assez standard.

Dans [Wang *et al.* 2010], les auteurs effectuent un alignement fin entre visages basé sur une méthode spécifique, puis soustraient tout simplement les images de profondeur entre deux modèles à comparer, créant ainsi ce qu'ils appellent des *SSDM* (différence entre images de profondeur) et *SDM* (valeur absolue de la carte précédente). Le problème est ramené à l'étude de cartes en 2D symbolisant la différence entre visages, et il est résolu à l'aide de méthodes du type boosting. Les auteurs revendiquent une très grande efficacité calculatoire.

Les auteurs de [Gong *et al.* 2009] ont, quant à eux, travaillé uniquement sur la reconnaissance d'expressions. Ils ont encodé directement les images de profondeur comme la somme de *Basic Facial Shape Components* (BFSCs) et d'*Expressional Shape Components* (ESC). Le principe est de représenter la partie non-mimique du visage comme la somme d'autres visages neutres à l'aide de la transformée de Karhunen-Loeve, en supposant la base d'apprentissage comme suffisamment représentative et complète. À l'aide de régions, les expressions sont classées *via* SVM.

Ces méthodes requièrent souvent une normalisation minutieuse des modèles 3D. Elles sont en effet sensibles aux problèmes d'échelle de par leur manière de représen-

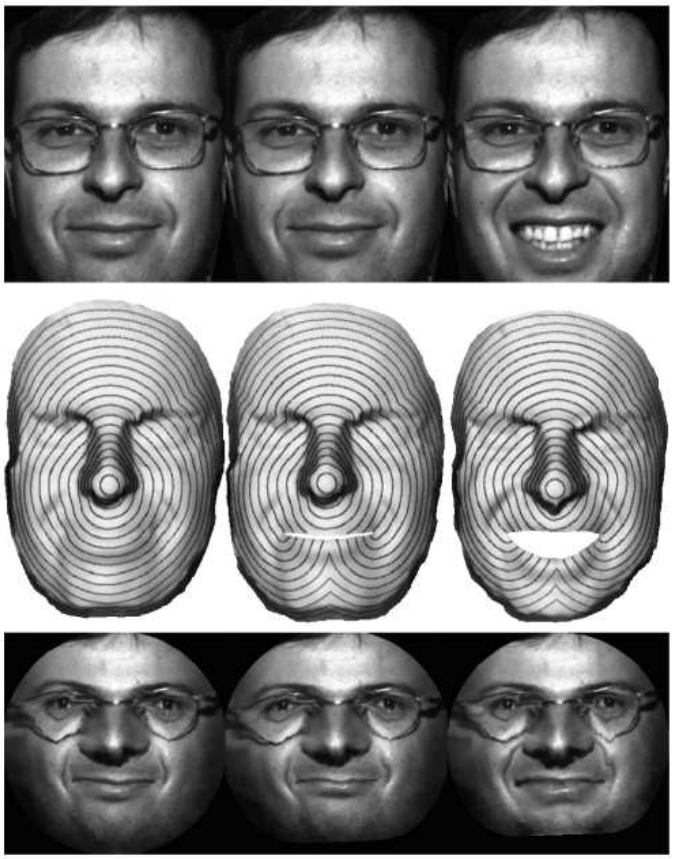


FIGURE 1.14 – La méthode proposée dans [Mpiperis *et al.* 2007b] ramène le visage à une représentation en 2D grâce à une représentation géodésique polaire, dans laquelle un traitement spécifique est réservé au problème de la bouche ouverte. Une fois ramenés au disque en 2D, les visages sont décrits à l'aide d'une Analyse en Composantes Principales.

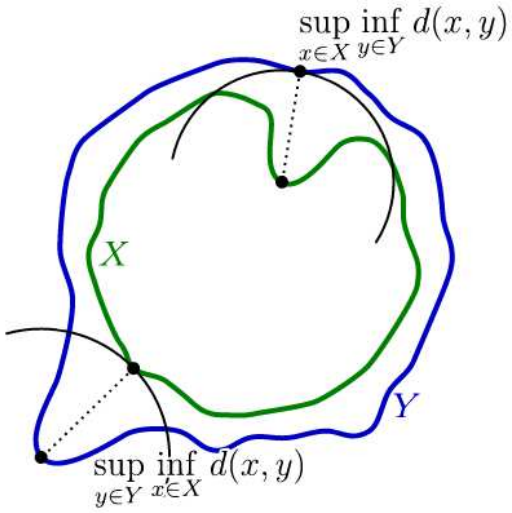


FIGURE 1.15 – Exemple de calcul de la distance d'Hausdorff entre deux formes 2D.

ter le visage. Une autre critique que l'on peut formuler à leur encontre est qu'elles se basent toutes d'une manière ou d'une autre sur la supposition qu'il existe un modèle moyen des données représentées, en l'occurrence du visage en 3D. Cette approche est discutable en ce qui concerne le visage humain.

1.6.3 D'autres mesures de distance entre surfaces rigides

D'autres mesures de distance que la somme des distances point à point entre deux surfaces ont été expérimentées dans l'état de l'art. De même qu'avec ICP, le principe reste d'aligner deux surfaces *via* une transformation rigide, puis d'en déduire un score de similarité.

La distance d'Hausdorff (figure 1.15), utilisée entre deux formes en traitement d'image, permet de quantifier leur dissimilarité. Sa définition est simple et ses utilisations sont très générales. Elle a été introduite pour la comparaison de visages en 3D par [Achermann & Bunke 2000], dans un scénario de reconnaissance de personnes. Supposons que nous avons deux ensembles de points A et B, la distance $H(A,B)$ entre les deux ensembles est définie par

$$H(A,B) = \max\left(h(A,B), h(B,A)\right)$$

où h est définie par :

$$h(A,B) = \max_{a\in A}\left(\min_{b\in B}\|a-b\|\right)$$

Dans [Achermann & Bunke 2000], le principe est étendu de la 2D à la 3D, et appliqué directement sur les images de profondeur. [Pan *et al.* 2003] ont comparé

les résultats obtenus par cette méthode à ceux d'ACP sur la base 3D_RMA, ces derniers se montrant à l'avantage de la distance d'Hausdorff.

[Lee & Shim 2004] ont pondéré la distance d'Hausdorff afin de calculer un score de similarité, qu'ils ont fusionné avec des informations de courbure (courbures maximale, minimale et gaussienne). [Russ *et al.* 2004] ont également employé la distance de Hausdorff sur des images de profondeur, dans un scénario de reconnaissance de personnes.

[Queirolo *et al.* 2010] ont utilisé une mesure d'interpénétration dans un scénario de reconnaissance faciale. Après un alignement grâce à une transformation rigide, les auteurs ont utilisé non pas la distance entre deux points appairés, mais ils ont incrémenté un compteur chaque fois que la distance entre ces deux points appairés était supérieure à un certain seuil donné, ou que leurs normales à la surface divergeaient d'un angle supérieur à un seuil. Le compteur, ramené au nombre de points considérés, forme un coefficient dit d'interpénetration. Les auteurs ont calculé un tel coefficient sur diverses régions du visage.

Dans chacun de ces cas, une étape d'alignement est nécessaire. Celle-ci est généralement effectuée *via* ICP, mais [Queirolo *et al.* 2010] ont utilisé une heuristique comme le recuit simulé pour déterminer finement les coefficients de rotation et de translation. D'après les auteurs, cette technique donne plus de chances d'arriver à un minimum local que l'algorithme ICP. Ces méthodes restent toutefois assez similaires dans leur concept et dans leur application à ICP. Celà explique, à notre sens, l'absence d'applications connues au problème de la reconnaissance d'expressions.

1.6.4 Modèle déformable

Le calcul de similarité entre surfaces recalées par transformation rigide s'est rapidement montré limité, surtout en regard des déformations du visage liées aux expressions. Plusieurs approches ont favorisé la mesure de ces déformations, éventuellement à partir d'un modèle génerique. L'idée est, ici, d'étudier les déformations à appliquer à un modèle pour épouser la géométrie d'un autre ; les paramètres de cette déformation servent alors à décrire le modèle *probe*, et sont utilisés pour la comparaison entre visages. Le but recherché étant, *in fine*, d'être capable de dissocier la part de ces paramètres correspondant à l'expression de celle liée à l'identité.

[Lu & Jain 2005] (figure 1.16) ont proposé, pour la reconnaissance faciale, l'emploi de *Thin-Plate Splines* (TPS). Cette méthode, dérivée de la mécanique, permet de calculer l'énergie nécessaire à plier une surface. Plusieurs courbes (splines) sont ainsi considérées entre des *landmarks* positionnés manuellement sur le visage, localisées sur des portions non rigides du visage comme les joues. Un traitement spécifique est également appliqué à la région de la bouche. Un groupe de contrôle et d'appren-

tissage est constitué pour déterminer les paramètres de déformations appliqués aux splines en fonction des expressions. Les visages *probe* sont ensuite comparés à chaque visage *gallery* en utilisant une transformation rigide pour les régions non-mimiques du visage, et en étudiant l'énergie de déformation des régions mimiques en regard des paramètres appris pour les expressions courantes.

Le principe de cette méthode a été étendu dans [Kakadiaris *et al.* 2007] (figure 1.17), toujours pour la reconnaissance de personnes. Dans ce travail, les auteurs ont généré un modèle moyen, baptisé l'AFM (*Annotated Facial Model*), muni d'une paramétrisation UV, c'est-à-dire conformale. La représentation conformale, au contraire d'une représentation géodésique, conserve les angles tandis qu'elle déforme les distances. En l'occurrence, elle permet de disposer d'un *mapping* dense entre une surface 3D et une image correspondante en 2D. Dans un premier temps, ce modèle générique est aligné sur le modèle étudié (utilisant consécutivement des *spin images*, ICP et une méthode de recuit simulé), puis il est déformé pour en épouser les contours. De fait, la paramétrisation UV et les régions déterminées sur le modèle générique sont applicables au modèle *probe*. La suite de l'algorithme se base sur des représentations par sous-espace appliquées au domaine 2D, dont les descripteurs comprennent des cartes normales, des ondelettes de Haar et une méthode baptisée *Pyramid Transform*. Les auteurs revendiquent une grande efficacité calculatoire, considérant que les représentations en 2D sont rapides, et que a majeure partie des calculs se situe autour de l'étape d'ajustement et de déformation entre le modèle générique et le modèle 3D étudié.

Dans [Mpiperis *et al.* 2008], les auteurs proposent une approche similaire en déformant un modèle générique pour l'adapter au modèle étudié. L'idée est de minimiser une énergie de déformation. Toutefois, contrairement à l'approche précédente, cette méthode nécessite l'utilisation de *landmarks* manuels. Autre différence, ce sont directement les paramètres de déformation qui sont exploités : on ne ramène pas le visage à un autre espace de représentation. Ces paramètres sont encodés à l'aide de modélisations bilinéaires symétriques et asymétriques, la première encodant simultanément l'identité et l'expression tandis que la seconde est dédiée à l'expression. Des expérimentations ont été conduites sur la base BU3D-FE.

En 2012, [Fang *et al.* 2011] ont proposé une amélioration du *framework* présenté dans [Kakadiaris *et al.* 2007], appliqué aux expressions du visage, à la fois sur des images 3D statiques et sur des séquences d'images 3D dynamiques (dites de 4D). De même que dans [Kakadiaris *et al.* 2007], l'idée est d'adapter l'AFM à un visage inconnu. Après une normalisation d'échelle, un ensemble de PDMs (*Point Distribution Models*) est établi à l'aide de la représentation conformale du modèle générique, en utilisant divers descripteurs pour la surface (position, normales à la

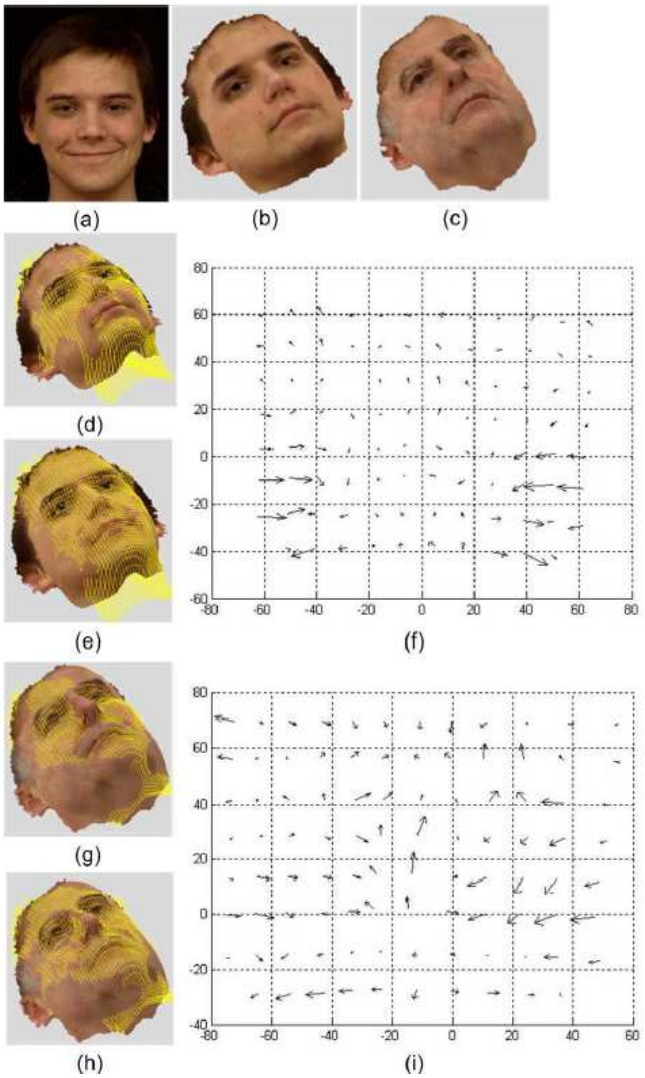


FIGURE 1.16 – Étude des déformations entre 2 visages selon la méthode exposée dans [Lu & Jain 2005]. L'étude se porte sur l'énergie nécessaire à déformer le maillage autour de points d'ancrages disposés en grille.

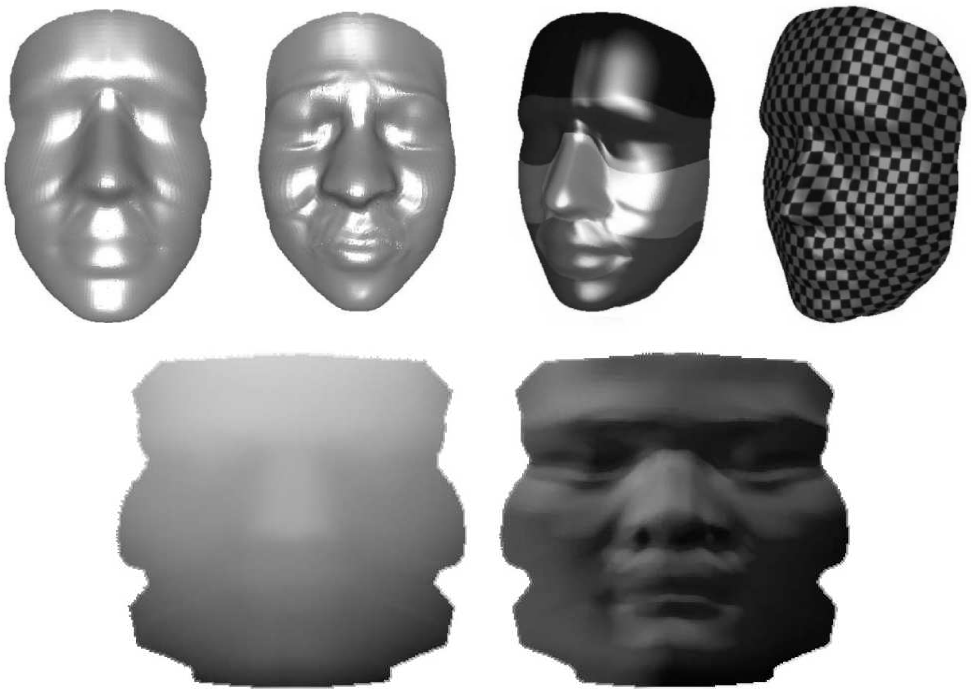


FIGURE 1.17 – Le principe de l'AFM (*Annotable Deformable Model*) [Kakadiaris *et al.* 2007] est de recaler un modèle 3D générique sur le modèle *probe* par déformations successives. Dans la première rangée, les deux premières images montrent ce recalage : la première correspond à l'état initial du modèle déformable et la seconde à son état une fois le recalage terminé. Ce modèle générique est doté d'une segmentation en régions (3ème image) et d'une représentation conformale (4ème image). Cette paramétrisation permet de ramener le problème à celui de la reconnaissance faciale en 2D (2ème rangée), à l'aide par exemple de caractéristiques géométriques (à gauche) ou de l'orientation des normales (à droite).

surface, courbure, ondelettes). Nous n'aborderons pas ici la méthode suivie pour la reconnaissance d'expressions en 4D.

D'après leurs auteurs, ces méthodes montrent des résultats très prometteurs, tant en termes de performances qu'en termes de temps de calcul. Elles sont néanmoins très complexes sur le plan calculatoire, et semblent sensibles au point d'initialisation. Par ailleurs, certaines de ces méthodes s'appuient sur une représentation conformale de la surface faciale, laquelle est avec les techniques actuelles très dépendante des conditions d'initialisation et de la topologie de la surface.

1.6.5 Courbes de niveau

Une dernière famille notable de méthodes étudiant les déformations élastiques d'un visage à l'autre est basée sur l'exploitation de courbes de niveau. Ces méthodes ne sont toutefois pas à proprement parler holistiques. En effet, elles exploitent un ensemble de courbes indépendamment pour aboutir à une décision sur la base d'une fusion entre les résultats obtenus.

[Samir *et al.* 2006] ont représenté le visage comme un ensemble de courbes de niveau, équipotentielles par rapport à la coordonnée en Z sur l'image de profondeur, pour la reconnaissance de personnes. Ces courbes sont ensuite représentées puis comparées dans un espace aux propriétés riemanniennes, justifiant l'existence d'une norme entre ces courbes. Les auteurs affirment ainsi que chaque point du chemin reliant un ensemble de courbes à un autre définit un visage. La distance entre deux courbes est une nouvelle fois assimilée au coût de déformation de l'une à l'autre. Cette méthode est toutefois sensible à la pose et aux occultations ou aux trous, puisqu'elle exploite des courbes fermées. Le premier problème a fait l'objet d'un travail de [Drira *et al.* 2009], les auteurs travaillant sur des courbes iso-géodésiques plutôt que sur des courbes isométriques. La position de l'origine et la topologie de la surface posant toujours problème en cas de données lacunaires ou d'occultations, les auteurs ont ensuite proposé de travailler sur des courbes radiales et non fermées, en les comparant cette fois à l'aide d'un algorithme de programmation dynamique [Drira *et al.* 2010] (figure 1.18). Dans chacun de ces scénarios, diverses méthodes de fusion simples (Borda-Count, moyenne...) ont été utilisées pour le score global de reconnaissance.

Ce *framework* a également été exploité pour la reconnaissance d'expressions. Dans [Maalej *et al.* 2011a] (figure 1.19), les auteurs ont comparé des courbes iso-géodésiques à un ensemble de visages de référence pour chaque expression, au voisinage de points d'intérêts localisés manuellement sur la base BU-3DFE. Des algorithmes de type *Multiboost* et SVM multiclasses ont été utilisés pour effectuer la classification.

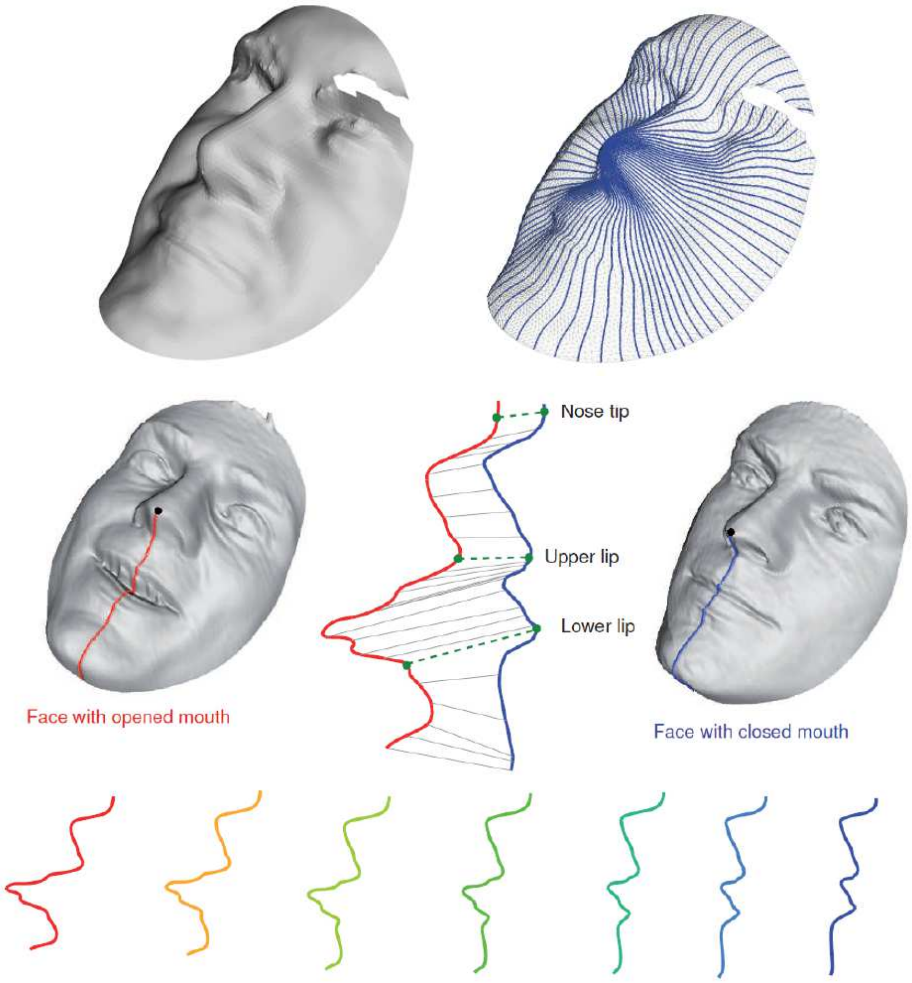


FIGURE 1.18 – Dans [Drira *et al.* 2010], le visage 3D est représenté sous la forme d'un ensemble de courbes radiales partant du bout du nez (ligne du haut). Le chemin de déformation pour passer d'une courbe à celle lui correspondant sur un autre visage est ensuite déterminé à l'aide d'une énergie de déformation, définissant une métrique de distance. La distance entre deux visages est donnée par l'ensemble des distances individuelles de chacune de leurs courbes.

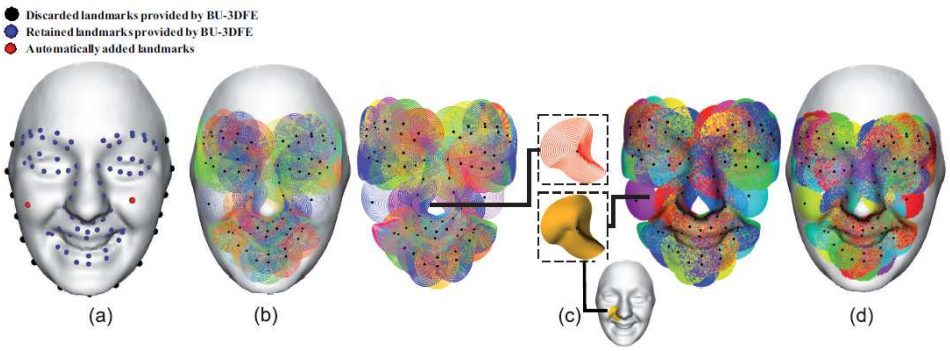


FIGURE 1.19 – Dans [Maalej *et al.* 2011a], le visage 3D est représenté comme un ensemble de courbes géodésiques au voisinage de points d'intérêt annotés manuellement. De même que dans le *framework* exposé en figure 1.18, une distance est calculée entre chaque courbe correspondante d'un visage à l'autre. À la différence d'un scénario de reconnaissance faciale, ces distances sont utilisées directement dans un classifieur pour la reconnaissance d'expressions du visage.

Dans [Mpiperis *et al.* 2007a], les auteurs ont calculé, dans un scénario de reconnaissance faciale, l'intersection entre différentes sphères et la surface faciale pour définir un ensemble de courbes de contour sur le visage. Ces contours sont ramenés à une représentation en 2D. Ils sont ensuite décrits à l'aide d'algorithmes de la littérature du domaine de l'analyse de formes en 2D tels que les moments de Hu, les descripteurs de Fourier elliptiques et le *Curvature Scale Space* (CSS). Une distance entre les descripteurs des courbes plutôt que les courbes directement est ainsi calculée.

Ces méthodes sont séduisantes sur le plan mathématique, mais sont en l'état très dépendantes des conditions d'initialisation (localisation du bout du nez, pose dans certain cas) ainsi que de la topologie de la surface faciale. Un problème évoqué couramment par les auteurs est l'ouverture de la bouche, qui nécessite des traitements spécifiques dans plusieurs des travaux évoqués dans cette section.

1.7 Méthodes hybrides et fusion.

Nous avons montré dans les parties précédentes que les méthodes d'analyse de visage en 3D pouvaient s'appuyer sur des informations locales du visage (partie 1.5), ou sur une comparaison globale (holistique) entre les visages (partie 1.6). Il existe des méthodes intermédiaires, que nous appellerons ici hybrides, qui considèrent le visage comme une somme de régions sur lesquelles un appariement global est effectué de manière individuelle. Nous nous intéresserons également aux méthodes de fusion, qui rassemblent les données de divers types de méthodes pour produire leurs résultats.

1.7.1 Méthodes hybrides globales - locales

Dans les méthodes exposées dans les sections 1.6.4 et 1.6.5, nous avons pu constater l'usage de régions, c'est-à-dire de sous-ensembles du visage, pour améliorer les performances d'algorithmes holistiques. L'AFM [Kakadiaris *et al.* 2007] combine les résultats d'une description effectuée sur quatre régions distinctes du visage. Les méthodes basées sur des courbes de niveau comme [Drira *et al.* 2010] n'utilisent en fait que des représentations partielles du visage pour les combiner ensuite. Nous pouvons également noter dans [Huang *et al.* 2011] l'octroi de poids différents accordés à l'appariement des points d'intérêt selon leur localisation sur le visage. Enfin, citons pour la reconnaissance d'expressions les auteurs de [Maalej *et al.* 2011a], qui ont analysé le voisinage de points d'intérêt à l'aide d'une méthode employée habituellement pour une comparaison globale de visages.

Ces travaux ne sont pas isolés et il existe une famille de méthodes faisant usage de techniques qu'on pourrait qualifier d'holistiques, appliquées localement pour ensuite être fusionnées afin d'obtenir un score global. À notre connaissance, de telles méthodes n'ont été utilisées que dans le cadre de la reconnaissance de personnes. Nous proposerons néanmoins une telle approche pour la reconnaissance d'expressions dans la partie 2.4.

Parmi ces méthodes, nous pouvons citer [Amor *et al.* 2006] (figure 1.20). Dans ces travaux, les auteurs ont divisé manuellement un visage référence en deux parties, une partie statique (front, yeux et région nasale) et une partie mimique (joues, bouche). Le visage *probe* est appairé via ICP uniquement à la partie statique du visage référence, définissant ainsi les parties mimique et statique du visage *probe*. Ce dernier est ensuite recalé aux visages *gallery* via ICP en utilisant uniquement la région statique, puis un score ICP (sans recalage) est calculé pour les parties statique et mimique du visage. La fusion des deux scores définit le score global attribué au visage.

Les auteurs de [Faltemier *et al.* 2008] vont plus loin dans ce sens. Ils calculent les scores d'appariement ICP de 28 régions, définies manuellement comme l'intersection de sphères de différents diamètres avec le visage. Les centres de ces sphères sont localisés par leurs coordonnées X et Y par rapport à celles du bout du nez, induisant une certaine sensibilité à la pose et donc à l'étape de normalisation. Les auteurs ont testé plusieurs méthodes de fusion et rapportent qu'une fusion des rangs (Borda Count modifiée) est optimale selon leur expérimentation.

Dans [Spreeuwers 2011], les auteurs emploient 60 régions, sur lesquelles ils appliquent une classification du type PCA-LDA. La localisation des régions est effectuée grâce à un système de paramétrisation dit intrinsèque du visage, correspondant de fait aux coordonnées cylindriques obtenues grâce à une méthode de normalisation

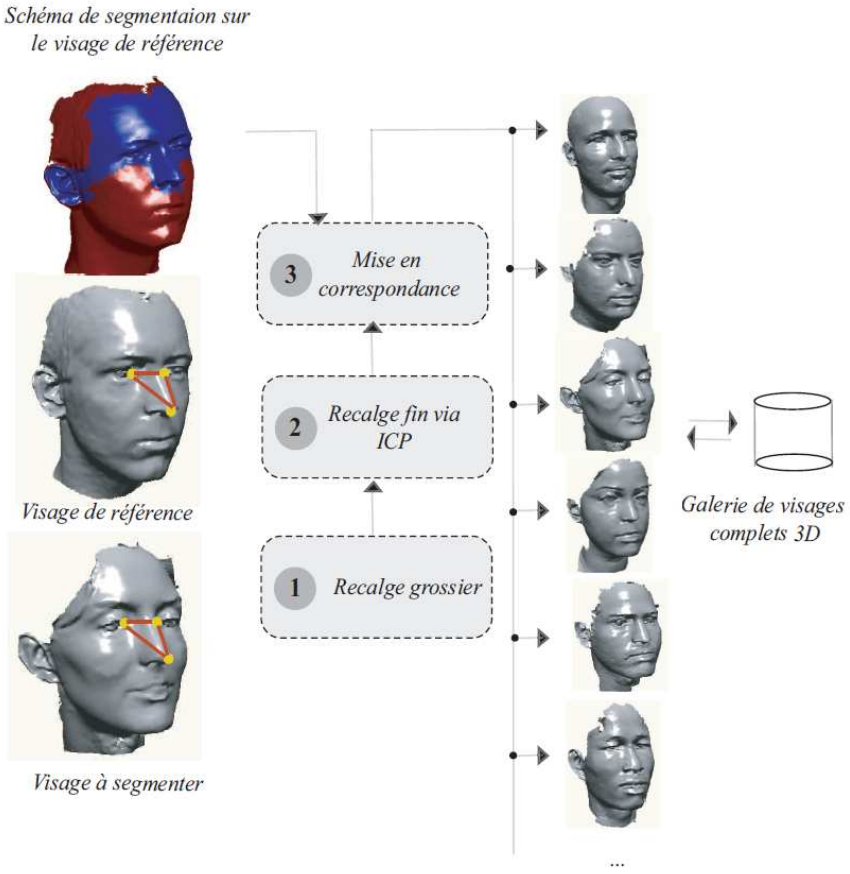


FIGURE 1.20 – Aperçu de la méthode employée dans [Amor *et al.* 2006]. Par rapport à ICP, Le visage est recalé partiellement grâce à une subdivision du visage en 2 régions mimiques et statiques. Le score global est une somme pondérée des distances ICP entre région mimique et région statique, en utilisant le seul recalage sur la région statique.

de la pose sophistiquée. La fusion est, selon les cas (vérification ou identification) effectuée à l'aide d'un vote majoritaire (rang) ou par le biais de seuils appliqués aux scores.

La difficulté dans ces approches réside en grande partie dans la détermination et la localisation des régions employées. D'après notre étude, toutes les méthodes abordées dans cette partie sont basées sur l'emploi de régions déterminées de manière arbitraire.

1.7.2 Fusion

Si dans la section précédente, nous avons évoqué uniquement des méthodes basant leur principe sur l'application d'une même technique à différentes régions du visage avant une fusion des différents résultats obtenus indépendamment, il est possible de fusionner les résultats obtenus de différentes sources et méthodes. Plusieurs approches dans l'état de l'art ont appliqué cette technique pour la reconnaissance de personnes. Nous n'avons malheureusement pas connaissance de telles techniques appliquées à la reconnaissance d'expressions. Cette différence est sans doute lié à la relative abondance de techniques correspondant à la reconnaissance faciale comparativement à la reconnaissance d'expressions.

Les auteurs de [Gökberk *et al.* 2005] ont comparé les performances de 5 systèmes de reconnaissance faciale différents, qui sont ICP, une représentation de la surface par les normales, une représentation par les profils et des approches PCA et LDA appliquées à l'image de profondeur. Ils ont ensuite testé deux méthodes de fusion entre ces différentes approches, une approche dite parallèle et une approche dite hiérarchique, au niveau des rangs.

Dans [Mian *et al.* 2007], les auteurs utilisent d'abord un *prescreener* (une méthode de rejet prématuré des visages) basée sur SIFT et épaulée par une représentation sphérique de la surface faciale. Une approche région est alors appliquée aux visages restants, à l'aide d'ICP. La fusion des distances ICP au niveau des différentes régions est effectuée au niveau des scores.

Les auteurs de [Li & Zhang 2007] ont annoté manuellement 43 points sur les visages de la base GavabDb (cf. section 1.2.2.4), formant ainsi un maillage 3D représentant le visage dans une structure stable. Ils ont ensuite exploité différents attributs de ce maillage, comme les distances géodésiques, les angles, les aires, les distances euclidiennes et les courbures. Pour chacun de ces attributs, qu'ils appellent signature, ils ont attribué un poids aux différentes mesures le composant en s'inspirant du calcul de la distance de Mahalanobis. Ils ont ensuite appliqué une fusion au niveau score dont les paramètres sont extraits des vecteurs de la matrice de covariance de chacune des signatures.

Dans [Ben Soltana *et al.* 2010], les auteurs ont décrit les images de profondeurs par le biais de différentes mesures (courbures moyenne, gaussienne, maximum et minimum, orientation des vecteurs normaux, tangents et binormaux). Ces mesures étant corrélées, ils ont d'abord réduit la dimensionnalité via *LDA*, puis ils ont classé les visages via kNN pour les différents descripteurs de la surface du visage. Ils ont ensuite effectué une fusion hybride entre score et rang, dont le scénario est défini par une heuristique, en l'occurrence un algorithme génétique.

1.8 Méthodes annexes à l'analyse de visages en 3D et localisation de points d'intérêts

De nombreuses publications évoquées dans les paragraphes précédent ont proposé des méthodes plus ou moins sophistiquées pour normaliser la pose, recadrer le visage et localiser certains points d'intérêt. Notamment, la localisation du bout du nez y est récurrente. Certains articles fondent également leurs résultats sur la localisation de points d'intérêts définis anatomiquement. On trouve également dans la littérature des mentions concernant le rééchantillonnage des visages en 3D, la suppression de trous et de pics, le lissage des surfaces et leur retriangulation.

Le problème du prétraitement des visages est ainsi une constante dans le domaine de l'analyse de visages en 3D, que ce soit pour la reconnaissance de personnes ou la reconnaissance d'expressions. Ce problème, ainsi que celui de la localisation automatique des points d'intérêt anatomiques, a donc été spécifiquement abordé dans un certain nombre de techniques de l'état de l'art.

Dans [Szeptycki *et al.* 2009], les auteurs ont proposé une méthode permettant de localiser automatiquement 15 points d'intérêt, testée sur la base FRGC. Dans leur méthode, un premier prétraitement de la surface faciale est détaillé, et fait usage d'un filtre médian modifié sur l'image de profondeur, suivi d'une méthode d'interpolation pour remplir les éventuels trous. La recherche des points d'intérêt se base sur une classification HK de la courbure à différentes échelles. En premier lieu sont localisés le bout du nez et les coins intérieurs des yeux grâce à des seuillages successifs. Un modèle générique est ensuite utilisé pour localiser grossièrement les 12 points d'intérêt restants, puis leur localisation est affinée par seuillages successifs.

Dans [Gupta *et al.* 2010], les auteurs ont proposé une approche de reconnaissance faciale basée sur la localisation automatique de points d'intérêt du visage. Les 10 points sont localisés de manière successive, en se servant de la position des points précédemment détectés. Le bout du nez est d'abord détecté à l'aide d'un recalage ICP, puis d'une classification de courbures HK et au niveau d'un maximum de la courbure gaussienne. Les points suivants sont localisés à l'aide de détecteurs

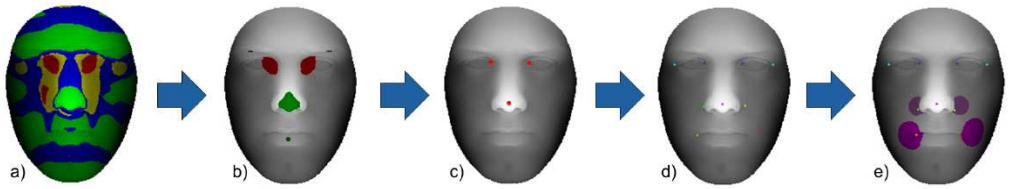


FIGURE 1.21 – Visualisation de la méthode de [Szeptycki *et al.* 2009] pour localiser les points d'intérêt du visage, par seuillages successifs sur la classification H-K (figure *a*) des courbures du modèle 3D. Les régions colorées sur les visages *b*, *c*, *d* et *e* correspondent aux zones de recherche des différents points d'intérêts.

de contour et sur la base de minimums locaux dans les mesures de courbures. Dans certains cas, une méthode utilisant un apprentissage appelée 2D+3D EBGM est utilisée. La recherche est également réalisée avec l'aide de la texture du visage, et les procédés sont *ad-hoc* pour chacun des points recherchés.

Dans [Zhao *et al.* 2009a], les auteurs ont construit un modèle statistique basé sur un apprentissage, appelé SFAM pour la localisation automatique de points d'intérêt. La description de chaque point du visage est ramenée à son repère local, puis son voisinage est représenté par ses valeurs de texture, profondeur, courbure et son masque d'occultation. Une ACP est calculée pour réduire la dimensionalité nécessaire pour décrire chaque point. Puis une optimisation basée sur un modèle statistique est effectuée pour localiser les points d'intérêt, en prenant en compte simultanément leur configuration spatiale et leurs descripteurs locaux. Cet algorithme a été testé avec 15 points d'intérêt sur la base FRGC et 19 points sur la base BU3D-FE.

1.9 Conclusion

Dans ce chapitre, nous avons présenté un état de l'art en analyse des visages en 3D. Cet état de l'art nous a permis d'avoir une vision globale des approches proposées pour la reconnaissance faciale en 3D ainsi que la reconnaissance d'expressions du visage. Les méthodes évoquées dans ce chapitre sont résumées dans les tableaux 1.1 et 1.2. La catégorisation choisie pour les regrouper permet d'en dégager quelques grandes familles. Elle met aussi en avant des similitudes entre reconnaissance faciale et reconnaissance d'expressions.

Néanmoins, nous avons pu constater un déséquilibre entre les méthodes proposée. Il est d'une part cardinal, la reconnaissance faciale étant un sujet largement plus abordé que la reconnaissance d'expressions. Il est d'autre part relatif à l'autonomie des méthodes vis à vis de l'opérateur humain. La plupart des méthodes de reconnaissance faciale sont entièrement automatiques, tandis que la moitié des méthodes de reconnaissance d'expressions présentées dans ce chapitre s'appuie sur des *landmarks*

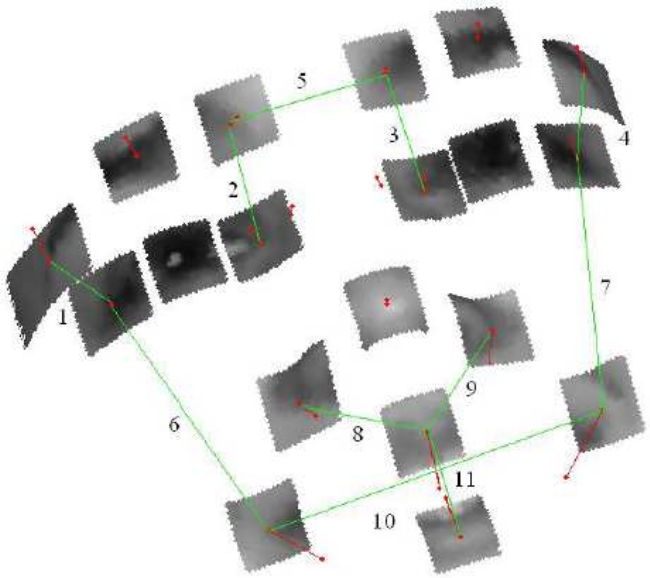


FIGURE 1.22 – Localisation de points d'intérêt par la méthode proposée dans [Zhao *et al.* 2011]. Dans ce papier, les auteurs caractérisent les points d'intérêt par leur voisinage (topologie et texture) ainsi que par leur relation spatiale, *via* un modèle statistique.

localisés automatiquement. Enfin, la variété des méthodes employées en reconnaissance d'expressions est moindre. Si il existe des méthodes de reconnaissance faciale pour chacune des catégories exposées dans ce chapitre, certaines techniques ne sont pas du tout exploitées en reconnaissance d'expressions faciales.

Les méthodes de reconnaissance d'expressions sont rarement complètement automatiques. Quand elles le sont (AFM), elles exploitent des mécanismes complexes et sophistiqués. En dehors de [Zhao *et al.* 2009a], aucune de ces méthode n'utilise les *landmarks* déterminés automatiquement.

En reconnaissance faciale, la méthode ICP a été améliorée par des approches hybrides et l'utilisation de régions. Cependant, si l'utilisation de régions doit améliorer la robustesse de l'appariement vis à vis des expressions, leur localisation est aussi, bien souvent, dépendante de la paramétrisation, c'est-à-dire de la méthode d'appariement. Cependant, la détermination de ces régions semble être souvent faite dans une paramétrisation sensible à la normalisation du visage (pose, initialisation).

Ces deux points sont abordés dans le chapitre 2, à travers une approche basée sur l'emploi de régions.

Les méthodes utilisant des points caractéristiques saillants affichent, comme en 2D, des résultats performants et prometteurs. Les techniques exploitant des cartes de représentation n'ont toutefois pas de pendant en reconnaissance d'expressions.

TABLE 1.1 – Résumé des méthodes de reconnaissance faciale exposées dans ce chapitre. *I.H.* correspond à *Intervention Humaine*.

Référence	Type	Mots Clés	I. H.
[Wu *et al.* 2004]	*Landmarks*	LSM	Oui
[Gupta *et al.* 2007]	*Landmarks*	LDA	Oui
[Castellani *et al.* 2008]	*Landmarks*	HMM	Non
[Daniyal *et al.* 2009]	*Landmarks*	PDM	Oui
[Mian *et al.* 2008]	Pts. Carac.	Classif. HK	Non
[Huang *et al.* 2011]	Pts. Carac.	MS-eLBP ; SIFT	Non
[Maes *et al.* 2010]	Pts. Carac.	Mesh-SIFT	Non
[Medioni & Waupotitsch 2003]	Rec. Rigide	ICP	Non
[Lu *et al.* 2004]	Rec. Rigide	ICP	Non
[Wang *et al.* 2006b]	Rec. Rigide	ICP	Non
[Achermann *et al.* 1997]	Sous-esp.	Eigenface	Non
[Hesher *et al.* 2003]	Sous-esp.	ACP	Non
[Heseltine *et al.* 2004]	Sous-esp.	Mahalanobis ; Eigenfaces	Non
[Mpiperis *et al.* 2007b]	Sous-esp.	Repr. Polaire Géo.	Non
[Wang *et al.* 2010]	Sous-esp.	SSDM ; SDM	Non
[Achermann & Bunke 2000]	Hol. Rigide	Distance de Hausdorff	Non
[Pan *et al.* 2003]	Hol. Rigide	Distance de Hausdorff	Non
[Lee & Shim 2004]	Hol. Rigide	Distance de Hausdorff	Non
[Russ *et al.* 2004]	Hol. Rigide	Distance de Hausdorff	Non
[Queirolo *et al.* 2010]	Hol. Rigide	Interpénétration	Non
[Lu & Jain 2005]	Mod. Déform.	TPS	Non
[Kakadiaris *et al.* 2007]	Mod. Déform.	AFM	Non
[Mpiperis *et al.* 2008]	Mod. Déform.	Mod. bilinéaire	Non
[Samir *et al.* 2006]	Courbes	Équipotentielles en Z	Non
[Drira *et al.* 2009]	Courbes	Géodésiques	Non
[Drira *et al.* 2009]	Courbes	Géodésiques	Non
[Drira *et al.* 2010]	Courbes	Coupes radiales	Non
[Mpiperis *et al.* 2007a]	Courbes	Courbes de contour	Non
[Amor *et al.* 2006]	Hybride	R-ICP	Non
[Faltemier *et al.* 2008]	Hybride	Multiples ICP	Non
[Spreeuwers 2011]	Hybride	Coord. cylindriques	Non
[Gökberk *et al.* 2005]	Fusion	PCA ; LDA	Non
[Mian *et al.* 2007]	Fusion	Fusion des scores	Non
[Li & Zhang 2007]	Fusion	Mahalanobis	Oui
[Ben Soltana *et al.* 2010]	Fusion	Algo génétique	Non

TABLE 1.2 – Résumé des méthodes de reconnaissance d'expressions faciales exposées dans ce chapitre. *I.H.* correspond à *Intervention Humaine*.

Référence	Type	Mots Clés	I. H.
[Soyel & Demirel 2007]	*Landmarks*	Réseau de neurones	Oui
[Tang & Huang 2008]	*Landmarks*	Adaboost	Oui
[Berretti *et al.* 2010a]	*Landmarks*	Descripteurs SIFT	Oui
[Zhao *et al.* 2009a]	*Landmarks*	SFAM	Non
[Wang *et al.* 2010]	Sous-esp.	SSDM ; SDM	Non
[Gong *et al.* 2009]	Sous-esp.	BFSC ; ESC	Non
[Fang *et al.* 2011]	Mod. Déform.	AFM ; PDM	Non
[Maalej *et al.* 2011a]	Courbes de niveau	Courbes iso-géodésiques	Oui

Étant holistiques, elles peuvent constituer une réponse aux méthodes basées sur la la localisation de points anatomiques, à la représentation trop parcellaire du visage. Elles peuvent également se révéler plus simples sur les plans calculatoire, de mise au point et de mise en œuvre, que les méthodes holistiques exploitant un modèle déformable.

Nous avons jugé intéressant d'explorer une approche holistique par cartes représentatives, dont l'objectif est d'être peu coûteuse et de répondre simultanément aux problématiques de reconnaissance faciale et de reconnaissance d'expressions. Ce point est abordé dans le chapitre 3.

Enfin, si des efforts ont été poursuivis par la communauté scientifique sur le sujet précis des occultations et des scans partiels de visages, peu d'études ont été menées à notre connaissance quant à la qualité de l'acquisition des modèles 3D et leur conséquence sur les performances des algorithmes d'analyse faciale. L'étude de cette problématique devrait pourtant être préliminaire à l'application sur le terrain de tels algorithmes. Nous nous y sommes intéressés dans le chapitre 4.

Chapitre 2

Approche régions

2.1 Introduction

La solution de la segmentation a souvent été employée dans les méthodes de l'état de l'art. Comme nous l'avons vu dans la partie 1.6.4, un système comme l'AFM [Kakadiaris *et al.* 2007], ou les solutions apportées par Faltemier [Faltemier *et al.* 2008] (voir partie 1.7.1) ou [Mian *et al.* 2007] (voir partie 1.7.2) ont montré des résultats intéressants en reconnaissance de personnes. Elles sont en concurrence avec d'autres méthodes, holistiques comme locales. En reconnaissance d'expressions, on préfère souvent baser les méthodes sur l'étude de points d'intérêt et de leur relation dans l'espace ou leur voisinage, ou employer des méthodes holistiques.

Dans ce chapitre, nous allons étudier en deux temps les apports, ainsi que les faiblesses d'une approche régions aux problématiques d'analyse faciale. Nous nous intéresserons tout d'abord à l'apport d'une approche régions pour le problème de la reconnaissance de personnes, tout particulièrement en présence d'expressions du visage. Nous nous intéresserons ensuite dans la section 2.4 au problème complémentaire, celui de la reconnaissance d'expressions.

2.2 Comment envisager une approche régions

Par région du visage, nous entendrons dans la suite de cette section un sous-ensemble de la surface faciale. Dans le contexte de l'analyse faciale, la subdivision du visage en plusieurs régions suppose d'effectuer les tâches d'appariement et/ou de comparaison de surface parallèllement (région par région, par exemple), puis de fusionner les résultats pour obtenir un classement global. Cette manière de procéder permet *a priori* une plus grande robustesse aux occultations dans les scénarios d'analyse faciale. On peut en effet supposer que si une région est occultée, la (probable) mauvaise qualité de sa description par l'algorithme d'analyse faciale sera partiellement ou totalement compensée à l'étape de fusion. Nous n'étudierons néanmoins pas directement cet aspect dans ce chapitre. L'approche régions permet également, potentiellement, d'améliorer les performances globales des algorithmes en

mettant l'accent sur les régions saillantes, c'est-à-dire les régions où les algorithmes parviennent à obtenir des informations plus discriminantes.

Toutefois, ce sont la localisation et la stabilité des régions qui sont en question.

Intuitivement, il semblerait logique de plaquer ces régions sur des parties anatomiques du visage : par exemple, les pommettes, les joues, le menton, une sphère de 2cm centrée sur le nasion, etc. Les zones du visage déformées par l'action des muscles faciaux sont également une alternative possible et répondent intuitivement aux problèmes rencontrés en analyse faciale. Cependant, nous avons vu en partie 1.3 que ces régions du visage n'ont pas toujours de relation évidente avec la géométrie du visage, que ce soit en 2D ou en 3D. La localisation de telles régions définies anatomiquement fait plutôt appel à l'expérience humaine ou à la localisation d'éléments sous-cutanés.

Dans les problèmes d'analyse faciale automatique, la raison anatomique n'est pas indispensable à la définition d'une région sur le visage. Le but recherché est la capacité de l'algorithme à maximiser les différences extra-classe, et à minimiser les différences intra-classe. Partant de ce principe, nous pouvons supposer qu'il nous sera utile de viser une stabilité intra-classe desdites régions, plutôt que de viser une stabilité sur le plan anatomique. De même, leur usage devra viser à maximiser les différences extra-classe.

Dans les sections suivantes, leur définition et leur localisation seront dictées par ces objectifs.

2.3 Reconnaissance faciale

2.3.1 Objectifs visés et aperçu de la méthode

Dans cette partie, nous nous limiterons à la reconnaissance de personnes en 3D, sans texture. Comme nous avons pu le voir en partie 1.7.1, une approche régions a déjà été proposée à plusieurs reprises dans ce domaine. Ce qui nous intéresse plus précisément ici est la résistance aux expressions du visage. Voulant particulièrement mettre l'accent sur cet aspect, les algorithmes présentés ici ont été évalués sur une base de données spécialisée dans les expressions du visage, la base Bosphorus (1.2.2.2). Les différences intra-classe seront donc au niveau de l'expression, et les différences extra-classe seront au niveau de l'identité, c'est-à-dire de la morphologie de la personne étudiée.

Nous nous proposons d'adresser ce problème selon le *framework* suivant. Dans un premier temps, le visage en 3D est prétraité, puis nous en extrayons des points caractéristiques. On effectue une paramétrisation de la surface 3D à l'aide des distances géodésiques. Nous apparions ensuite le visage *probe* au visage *gallery*, préalablement

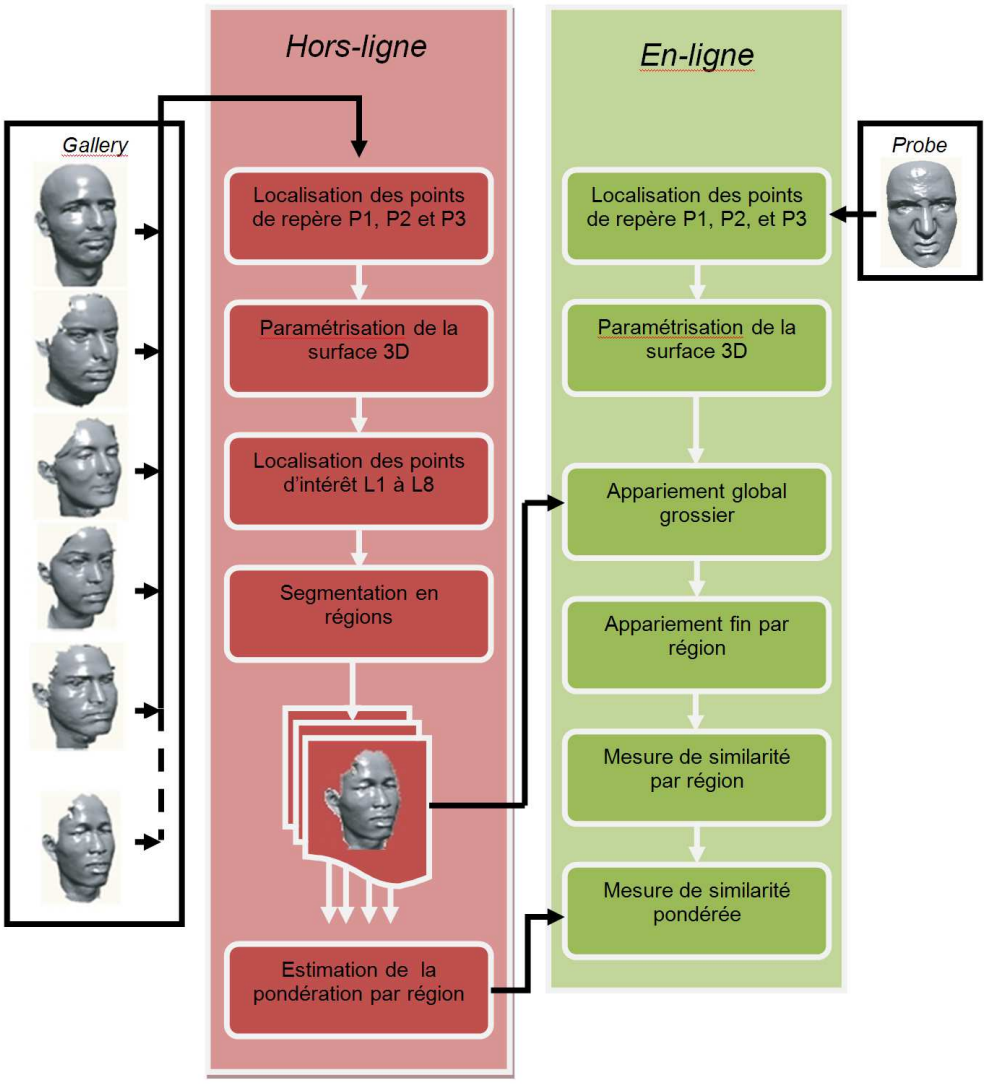


FIGURE 2.1 – *Framework* de notre méthode.

et automatiquement segmenté en régions, à l'aide de notre paramétrisation. Enfin, le score de similarité entre 2 visages est obtenu comme une somme pondérée des scores sur chacune des régions. Le score associé à une région est calculé sur la base d'une méthode de recalage rigide.

Nous diviserons la section suivante en deux parties. La première présentera le *framework* employé. La seconde présentera les performances de l'algorithme sur la base Bosphorus.

2.3.2 Une représentation du visage basée sur des distances géodesiques

En reconnaissance de personnes en 3D, l'étude des expressions faciales revient à l'étude des déformations causées par ces dernières. Comme nous l'avons vu dans la section 1.3, les expressions peuvent être décomposées sous forme de l'activation d'unités d'action (*Action Units*, AU), telles que décrites par [Ekman *et al.* 2002] au sein du FACS (voire la partie 1.3). Chaque AU peut se décrire comme la contraction ou la détente d'un ou plusieurs muscles du visage. L'action des muscles impliqués dans la plupart des AU étant locale, on peut donc supposer l'intérêt d'une approche régions. L'objectif est principalement d'obtenir une comparaison de surfaces performante sur les régions stables du visage, c'est-à-dire non affectées par des *Action Units*. Ainsi, il semble important que la localisation des régions soit stable sur les régions non mimiques, quand bien même le reste du visage serait déformé par l'activation d'AU. De même, il semble intéressant qu'un nombre minimal de régions soit affecté lors de l'activation d'une AU ou d'un groupe d'AUs lié(s) à une expression. La stabilité de ces régions en extra-classe (d'une personne à une autre) n'a par contre pas d'utilité au sens de la reconnaissance de personnes *a priori*. Elle peut au contraire témoigner d'une différence de morphologie.

Globalement, l'ensemble de l'approche de reconnaissance basée régions peut être scindé en deux phases, l'une hors-ligne et l'autre en-ligne (figure 2.1). La première phase est consacrée au traitement et à l'analyse des modèles de la *gallery*. Pendant la deuxième phase, en-ligne, un modèle *probe* est appairé avec les modèles de la *gallery*, et les mesures de similarité sont calculées. Une mesure de similarité entre deux visages est la pondération des mesures de similarité par région, dont les coefficients ont été calculés suite à un apprentissage qui sera abordé dans la section 2.3.5.

Dans les tâches de localisation des régions, d'appariement et de comparaison de surfaces, un élément clé est la paramétrisation du modèle 3D. Elle doit permettre d'associer à chaque point du modèle 3D des coordonnées uniques. Issue d'un scanner 3D, la paramétrisation naturelle est le repère euclidien dont l'origine correspond au capteur, et dont les axes X et Y correspondent respectivement aux axes X et Y de l'image de profondeur collectée. Cependant, cette paramétrisation n'est pas robuste aux changements de pose du visage scanné : une rotation et une translation sur l'un ou l'autre des axes se répercuteront, pour le même point du visage, en un changement de coordonnées dans l'espace de représentation euclidien. Si nous définissons un repère centré sur le visage en 3D, nous annulons l'impact d'une transformation rigide sur la pose du visage. De la même manière, la méthode de représentation des distances peut avoir un impact sur la sensibilité de notre paramétrisation aux variations d'expression et de morphologie. Ce sera le sujet de la partie 2.3.4.

2.3.3 Prétraitement et localisation des points de repère

Les objectifs du prétraitement des modèles 3D sont de minimiser l'influence de la qualité de l'acquisition à l'étape de la reconnaissance. Les données sont en effet généralement des images de profondeur et non pas des modèles 3D complets, ce qui implique que certaines parties du visage peuvent être manquantes. Ils comportent souvent également des pics et des trous, et font presque systématiquement l'objet d'un bruit d'acquisition. Pour assurer l'invariance à la pose, un modèle maillage est généré. Afin de supprimer les pics, on applique un filtre médian aux points dont les coordonnées en Z sont détectées comme aberrantes. Afin de corriger les trous, c'est-à-dire de déterminer les coordonnées en Z des points manquants sur l'image de profondeur, on opère une régression linéaire sur l'estimation des coefficients polynomiaux biquadratiques. Les détails de cette partie correspondent à des travaux publiés dans [Szeptycki *et al.* 2009], et déjà évoqués en partie 1.8.

La localisation de points de repère sur le visage est nécessaire à deux éléments dans notre approche. D'une part, une paramétrisation unique de chaque point d'un espace en 2D ou en 3D requiert au minimum la connaissance de trois points non alignés. D'autre part, la segmentation du visage, c'est-à-dire la localisation des zones de déformations liées aux AU ainsi que des zones statiques du visage, repose sur la localisation d'invariants du visage. Cette étape doit être invariante en pose et vis-à-vis des conditions d'éclairage. De préférence, elle ne doit pas utiliser l'information texturale. Comme points de repère, nous avons choisi le bout du nez et les coins intérieurs des yeux. Ces points sont des invariants du visage, c'est-à-dire que leur présence et leur localisation est toujours bien définie d'un point de vue topologique, indépendamment de la morphologie et de l'expression.

L'approche implémentée de localisation de ces points de repère anthropométriques est basée sur l'utilisation des courbures moyenne et gaussienne, et un modèle générique. Par seuillages successifs, et en étudiant les courbures moyennes et gaussiennes à différents rayons, nous localisons les coins intérieurs des yeux (notés P_1 et P_2) et le bout du nez (noté P_3), en affinant progressivement la précision de la localisation. Le détail de cet algorithme est disponible dans [Szeptycki *et al.* 2009].

2.3.4 Paramétrisation de la surface faciale 3D

L'étape de paramétrisation vise à accorder des coordonnées identiques à un même point physique, dans deux modèles 3D faciaux différents d'un même individu. On suppose que les points de repère anthropométriques P_1 à P_3 (coins intérieurs des yeux et bout du nez) sont localisés de manière précise. Ils sont naturellement distincts et non-alignés. Chaque point p d'un modèle facial 3D est alors décrit de manière unique

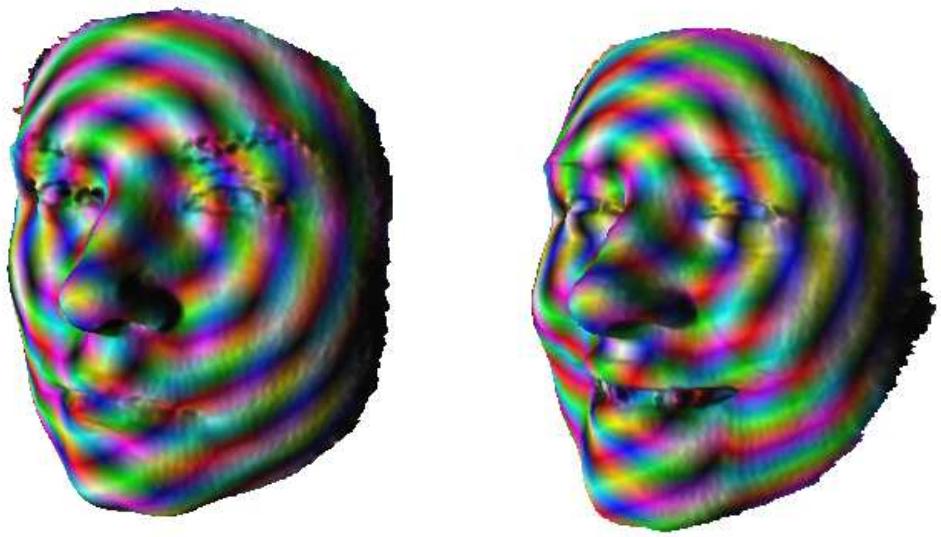


FIGURE 2.2 – La paramétrisation du visage à partir des points d'intérêt P_1 à P_3, utilisant les distances géodésiques. Ici, les composantes RVB correspondent respectivement aux distances géodésiques à P_1, P_2 et P_3.

par ses distances géodésiques (d_1, d_2, d_3) à ces 3 points (figure 2.2). Cette paramétrisation comporte l'avantage d'être invariante aux déformations rigides, c'est-à-dire aux translations et aux rotations. Contrairement à la représentation euclidienne, cette paramétrisation n'exige pas de recalage relatif à la pose. Elle est appliquée à chaque visage de la *galery*, mais aussi à chaque visage *probe*. À un point p de coordonnées (d_1, d_2, d_3) d'un visage *gallery*, on apparie p' de coordonnées (d'_1, d'_2, d'_3) du visage *probe* tel que la distance D entre (d_1, d_2, d_3) et (d'_1, d'_2, d'_3) soit minimale. Dans nos travaux, D est une distance euclidienne.

2.3.5 Segmentation en régions

La segmentation du visage constitue l'une des étapes centrales de l'approche de reconnaissance par régions. C'est à cette étape que les régions statiques et mimiques du visage sont localisées automatiquement, pour chaque visage de la *gallery*.

La segmentation d'un visage selon des critères purement anatomiques est délicate, étant donné qu'aucun champ d'action d'un groupe musculaire ou d'une AU ne peut, semble-t-il, être localisé simplement et directement grâce à la forme ou à la texture du visage. Nous avons ainsi choisi de déterminer nos régions de manière empirique, à l'aide de cartes dites de potentiel de déformation. Le potentiel de déformation est ici la mise en correspondance, via la paramétrisation précédemment exposée, d'une mesure de courbure entre 2 visages différents. En l'occurrence, la mesure de courbure dont nous nous sommes servis est le *Shape-Index*

[Koenderink & van Doorn 1992] avec un rayon de 25mm, indiquant la topologie de la surface au voisinage d'un point du modèle 3D. En appariant de la sorte, pour une même personne, plusieurs visages soumis à l'activation d'une ou plusieurs AU, nous pouvons mettre en évidence à l'aide du potentiel de déformation les régions statiques (le *Shape-Index* varie peu d'un visage à l'autre, c'est-à-dire que sa variance est faible) et les régions mimiques (le *Shape-Index* varie largement d'un visage à l'autre, c'est-à-dire que sa variance est grande).

A la suite de nos observations sur le potentiel de déformation à l'aide de la base Bosphorus (voir partie 1.2.2.2), nous avons décidé de découper le visage en 8 régions (figure 2.3), correspondant grossièrement :

- au nez,
- à l'œil gauche,
- à l'œil droit,
- à la pommettes gauche,
- à la pommettes droite,
- au front,
- à la partie gauche de la mâchoire,
- à la partie droite de la mâchoire.

Ces régions correspondent à des zones globalement stables (nez, front, pommettes), ou globalement mimiques (yeux, machoire), selon ce que nous avons observé du potentiel de déformation sur plusieurs visages. La figure 2.3 donne un exemple de visages sur lesquels nous avons pu nous appuyer afin de définir cette segmentation, les régions vert-jaune correspondant à des régions stables et les régions rouges correspondant à des zones mimiques selon notre paramétrisation. La séparation de la bouche en deux régions distinctes est supposée permettre une meilleure robustesse aux problèmes d'expressions partielles, de modèles partiels ou d'occultations.

Afin de segmenter la surface faciale 3D de la sorte, nous localisons d'abord les points de repère L_1 à L_8 correspondant aux représentants des huit régions. Ceux-ci sont localisés grâce à leurs distances géodésiques à P_1, P_2 et P_3, à l'aide d'un modèle générique. Ensuite, chaque point du visage est associé respectivement à la région S_i (centrée en L_i) à l'aide de deux paramètres, w_i et R_i avec $i \in [1, 8]$. Plus précisément, tout point p du visage appartient à la région S_i, représentée par le point de repère L_i, lorsque

$$s_i = \min(s_j) \text{ avec } j \in [1, 8]$$

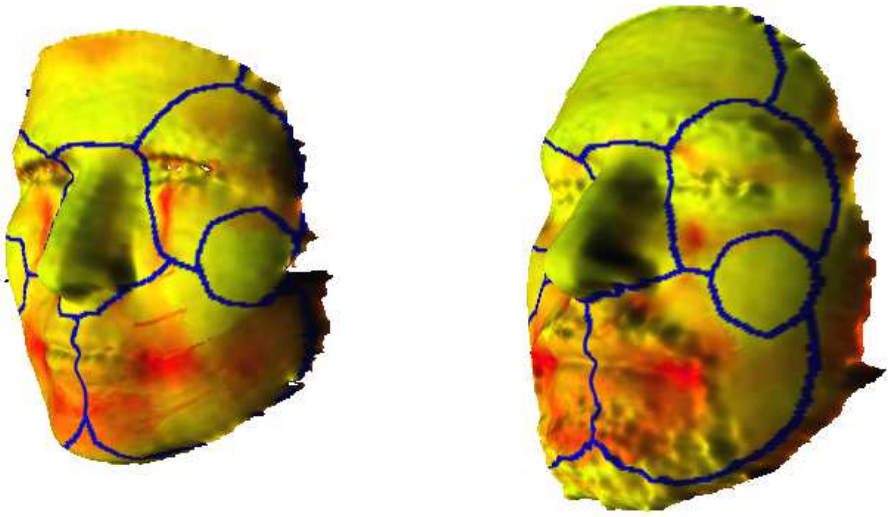


FIGURE 2.3 – Exemples de segmentation du visage en régions, sur des cartes de potentiel de déformation. la couleur correspond aux différences dans la valeur du *Shape-Index* (calculé sur la base d'un rayon de 25mm) par rapport au modèle expressif.

où s_i est calculé de la manière suivante :

$$\begin{aligned} \text{Si } w_i \times d_i > R_i \ , \quad & \text{alors} \quad s_i = \infty \\ & \text{sinon} \quad s_i = w_i \times d_i \end{aligned}$$

où d_i est la distance géodésique du point p au point de repère L_i.

Une fois la segmentation en régions effectuée, on calcule un score de similarité pour chaque région. Cette étape est relativement indépendante de la segmentation en régions du visage. Si dans cette section nous présentons des résultats relatifs à l'algorithme de référence ICP (Iterative Closest Point) [Zhang 1992], afin d'évaluer les bénéfices de la segmentation du visage, il est envisageable d'utiliser d'autres mesures de similarité dans cette méthode.

Pour chaque point p du visage *gallery* dans une région donnée, nous apparions le point p' correspondant du visage *probe* suivant notre paramétrisation. Une itération d'ICP permet de minimiser la distance euclidienne au sens des moindres carrés entre les deux régions. C'est-à-dire qu'on calcule la rotation et la translation qui minimisent cette distance. Le score de similarité par région est la moyenne des distances euclidiennes point à point entre le visage *gallery* et le visage *probe* au sein de cette région.

Finalement, le score de similarité entre deux visages est une mesure de similarité

pondérée de la forme suivante :

$$M = \sum_{i=1}^{8} W_i \times d(S_i, S'_i)$$

avec

$$d(S_i, S'_i) = \frac{1}{\text{card}(S_i)} \times \sum_{p_j \in S_i, p'_j \in S'_i} d_{euclid}(p_j, p'_j)$$

où $\text{card}(S_i)$ correspond au nombre de points p compris dans la région S_i.

La détermination des poids lors de cette fusion de scores est obtenue à l'aide d'une base d'apprentissage. Les poids sont calculés de sorte à minimiser l'*Equal Error Rate* (EER). Les détails concernant cet algorithme de fusion sont disponibles dans [Soltana *et al.*].

2.3.6 Résultats expérimentaux

L'expérimentation a été menée sur la base de visages Bosphorus (1.2.2.2). Cette base est constituée des images de profondeur de 105 individus, pour la plupart des acteurs professionnels, avec 1 à 3 modèles neutres par individu. Pour chaque individu, la base comporte 34 modèles avec déformations, dont 28 obtenus par l'activation d'AU isolées et 6 par expression. La base présente également des modèles avec occultations (lunettes ou mains) ou rotations. Nous n'avons pas utilisé ces derniers dans l'expérimentation présentée dans cette partie. Cette base est réputée difficile, car centrée sur les expressions faciales.

Concernant notre expérimentation, 29 individus ont été sélectionnés aléatoirement, avec au moins un modèle neutre par individu. Ces modèles neutres constituent notre *gallery*. Par la suite, 80 modèles *probe* ont été sélectionnés aléatoirement pour évaluer les performances de notre algorithme et les bénéfices de la segmentation en régions présentée dans cette partie. 74 d'entre eux présentent des expression faciales. La moitié de ces modèles est utilisée pour estimer la pondération optimale par région, l'autre moitié constituant les modèles de test.

Nous avons comparé les améliorations incrémentales de notre approche (approche D) avec l'approche *baseline* ICP. Les améliorations successives de l'approche R-ICP [Ben Amor 2006] sont obtenues :

- en additionnant simplement les scores par région mais sans apprentissage (approche B),
- en fusionnant les scores par la méthode Borda Count modifiée présentée dans [Faltemier *et al.* 2008] (approche C),
- et finalement en utilisant la pondération optimale des scores basée sur l'EER,

Approche	Taux d'identification (%)
Baseline ICP (A)	78.5
Régions sans pondération (B)	82.5
Régions Borda Count (C)	87.5
Régions avec apprentissage (D)	92.5

TABLE 2.1 – Résultats de l'expérimentation : améliorations successives de notre approche comparées à l'approche *baseline* ICP

telle que présentée dans [Soltana *et al.*].

Les résultats sont présentés dans le tableau 2.3.6. On remarque une amélioration progressive du taux de reconnaissance avec le raffinement des méthodes de fusion. Dans tous les cas, notre approche de segmentation du visage en régions ainsi que notre paramétrisation permettent une amélioration nette du taux de reconnaissance par rapport à l'algorithme ICP en présence d'expressions.

2.3.7 Critique

Si une amélioration nette des résultats est constatée dans cette expérimentation, nous devons tempérer sa portée.

Une première difficulté est que le système final est nettement plus complexe que l'algorithme *baseline*, ce qui implique un risque de surspécialisation. Ce risque est également lié au fait que, lors de nos études, nous n'avons pas trouvé de moyen algorithmique naturel de déterminer l'emplacement et la localisation des régions. Une autre observation que nous avons pu faire lors de cette expérimentation est liée aux poids accordés aux différentes régions. Les régions les plus petites (en terme de superficie) se sont en effet vues accorder les pondérations les plus faibles par la méthode d'attribution des poids. La supposition selon laquelle les régions non-mimiques auraient du recueillir une pondération plus importante ne s'est pas vérifiée d'une manière franche, laissant penser qu'une bonne partie des informations saillantes pour la reconnaissance faciale est localisée dans les régions mimiques.

2.4 Reconnaissance d'expressions

Nous nous intéressons maintenant au problème complémentaire, qui est celui de la reconnaissance d'expressions. Ici, l'intra-classe correspond à des personnes d'identité variable montrant la même expression, tandis que l'extra-classe comprend les autres expressions, parfois pour la même personne. Par conséquent, nous cherchons à résoudre le problème complémentaire à partir des mêmes outils : comparaison de

surface et système FACS, utilisation de régions notamment.

Dans cette partie, nous nous pencherons sur le problème de la reconnaissance d'expressions faciales en 3D, sans nous soucier de celui des occultations. Plus précisément, nous aborderons un problème standard de l'état de l'art, qui est celui de déterminer l'expression d'une personne inconnue parmi les six expressions prototypiques, telles que définies par Ekman [Ekman & Friesen 1971] (voir partie 1.4).

2.4.1 Pourquoi envisager une approche régions

Comme nous l'avons vu, et c'est une constante entre les problèmes de reconnaissance de personnes et d'expressions, une approche régions présente intuitivement de l'intérêt dans le cas où il est possible de rencontrer des occultations, ou d'avoir une représentation partielle du visage. Dans cette partie, nous allons voir que la reconnaissance d'expressions implique d'autres propriétés qui justifient potentiellement l'utilisation d'une approche régions.

Typiquement, les méthodes de reconnaissance d'expressions en 3D utilisant des points d'intérêt peuvent être classées en deux courants majeurs.

- Le premier extrait des données liées à la localisation de points d'intérêt, le plus souvent correspondant à des données anatomiques. Les éléments de littérature correspondant sont détaillés en partie 1.5.1.2.
- Le second extrait des caractéristiques locales à divers endroits sur le visage, comme des paramètres d'estimation de surface, des descripteurs SIFT, ou des comparaisons de surfaces utilisant une représentation géodésique. Ces méthodes sont décrites dans les parties 1.5.1.2 et 1.6.5.

Dans quasiment chacun de ces cas de figure, les méthodes et leurs évaluations sur la base BU-3DFE (voir partie 1.2.2.3) se sont appuyées sur l'utilisation de points fiduciaux dont la localisation était dépendante d'un opérateur humain. Dans ce cas, une des solutions pour un passage à un système totalement automatique serait de se baser sur une méthode d'extraction automatique de ces points d'intérêt. De telles méthodes existent pourtant (voir partie 1.8). Si elles n'ont été que rarement employées, il est probable que ce soit lié à la difficulté pour ces dernières de fournir une localisation d'une précision et d'une stabilité suffisante, comparé à ce qu'offre l'opérateur humain.

Dans la section qui suit, nous cherchons à compenser l'impact de l'imprécision des méthodes de reconnaissance d'expressions automatique, à l'aide d'une approche basée régions.

2.4.2 Résumé de la méthode.

Notre méthode peut être décomposée en deux phases majeures. Elles sont toutes deux précédées par le prétraitement des données 3D, étant entendu que certains modèles souffrent de bruit dans la plupart des bases de données de visages en 3D. Ce prétraitement est similaire à celui exposé dans la partie 2.3.3. Nous normalisons également la pose de tous nos modèles en les alignant via ICP à un modèle choisi aléatoirement. Nous effectuons finalement un remaillage de nos modèles 3D de sorte que leur triangulation soit propre et consistante.

Précisons également que dans cette partie, nous avons utilisé la paramétrisation euclidienne pour représenter le visage 3D. La paramétrisation présentée dans la partie 2.3.4 étant en effet destinée à minimiser l'impact des expressions, nous ne l'avons pas jugée adaptée au problème de la reconnaissance d'expressions du visage.

La première étape est une phase hors-ligne. Nous avons besoin de générer un modèle moyen pour chaque expression prototypique. Ces modèles nous serviront de visages de référence, auxquels seront comparés les modèles *probe* pendant la phase en-ligne. Nous avons également besoin de faire l'apprentissage du modèle utilisé dans la localisation automatique des points d'intérêt.

La deuxième étape est une phase en ligne. Dans un premier temps, nous localisons automatiquement les points d'intérêt sur les visages *probe*, à l'aide du modèle précédemment entrainé. Ensuite, sur la base de ces points d'intérêt, nous divisons le visage en plusieurs régions. Ces régions ont été déterminées à la suite d'une étude sur le système FACS (partie 1.3), et sur des considérations morphologiques. L'étape suivante est l'alignement de chacune de ces régions sur les modèles moyens générés lors de la phase hors-ligne, à l'aide de la méthode ICP. De cet alignement découle un score de similarité entre chaque région et chaque modèle moyen. Ces scores sont alors concaténés pour obtenir un vecteur caractéristique (méthode du *Lipschitz-Embedding* [Hjaltason & Samet 2003]). Finalement, nous utilisons ce vecteur caractéristique pour effectuer l'apprentissage ainsi que le test de notre classifieur sur les six expressions prototypiques.

L'ensemble de notre *framework* est disponible dans la Figure 2.4.

Dans la suite de cette partie, nous discutons les détails de chacune des étapes de notre méthode.

2.4.2.1 Phase hors-ligne

Concernant la détection automatique de points d'intérêt, nous avons utilisé le système SFAM (*Statistical Facial feAture Model*) proposé par [Zhao *et al.* 2009b]. Brièvement, ce système consiste en un modèle statistique, qui fait l'apprentissage

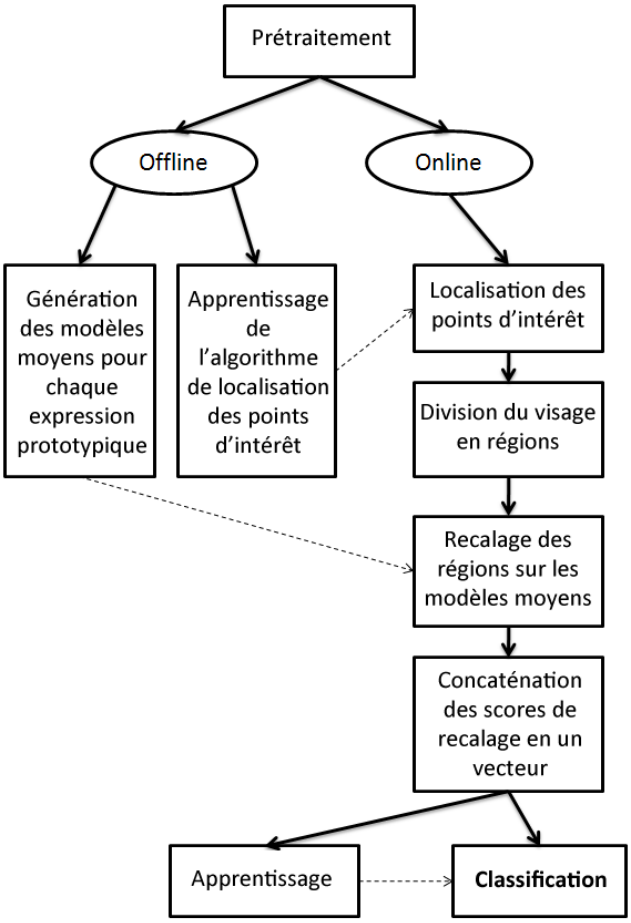


FIGURE 2.4 – *Framework* de notre méthode.

simultanément des variations globales de la morphologie du visage, ainsi que de celles locales autour des points d'intérêt. Cette méthode est décrite plus avant en partie 1.8. Elle nécessite un apprentissage hors-ligne, basé sur des *landmarks* localisés manuellement. Nous utiliserons donc cette méthode dans les deux parties hors-ligne et en-ligne de notre algorithme.

Pour les besoins de notre application, il est également nécessaire de générer des modèles moyens correspondants à chaque expression prototypique, ainsi qu'à l'expression neutre. À cet effet, nous avons tiré aléatoirement un modèle 3D pour chaque expression. Nous alignons alors chacun des modèles restants de notre base hors-ligne et de même expression via ICP. Suite à celà, chaque point de notre modèle original est associé à un nuage de points provenant de divers modèles. Le barycentre de chacun de ces nuages est utilisé comme coordonnée du nouveau point correspondant pour notre nouveau modèle générique, ce dernier conservant la triangulation du maillage de référence. Un lissage est finalement nécessaire. Pour le réaliser, nous projetons le modèle obtenu sur ses coordonnées X et Y afin de générer l'image de profondeur correspondante. Nous effectuons alors un remplissage des trous par interpolation bicubique ainsi qu'un lissage sur les coordonnées en Z par un filtre gaussien. Dans nos expérimentations, nous avons fixé le rayon de ce filtre à 3mm.

2.4.3 Phase en-ligne

La localisation automatique des points d'intérêt va nous permettre, lors de la phase en ligne, de comparer localement les modèles *probe* aux modèles moyens correspondant aux expressions prototypiques.

Il a été montré que les déformations du visages dues aux expressions, ou plus généralement l'activation d'*Action Units* (AU) affectent la surface faciale à travers des déformations non rigides. Ainsi, avoir la capacité de mesurer la quantité de déformations en certaines zones de la surface faciale devrait nous permettre de détecter l'activation d'une AU ou d'un groupe d'AUs. Comme dans la partie précédente (2.3), la principale difficulté est que la localisation de la région du visage affectée par le changement d'état d'un muscle ou d'un groupe de muscles uniquement emphvia la topologie ou la texture du visage, demeure un problème ouvert et difficile. De même, la capacité d'une machine à mesurer localement les déformations causées par l'activation d'une AU ou d'un groupe d'AU reste un challenge, ce problème étant notamment affecté par la morphologie de l'individu étudié. En outre, bien que les expressions prototypiques soient reconnues comme étant universelles, la magnitude de la déformation engendrée par l'activation d'une AU pour une expression donnée semble être différente d'une personne à une autre, ou d'un groupe ethnique à un autre. Ainsi, nous sommes dans l'obligation de résoudre le problème à une modéli-

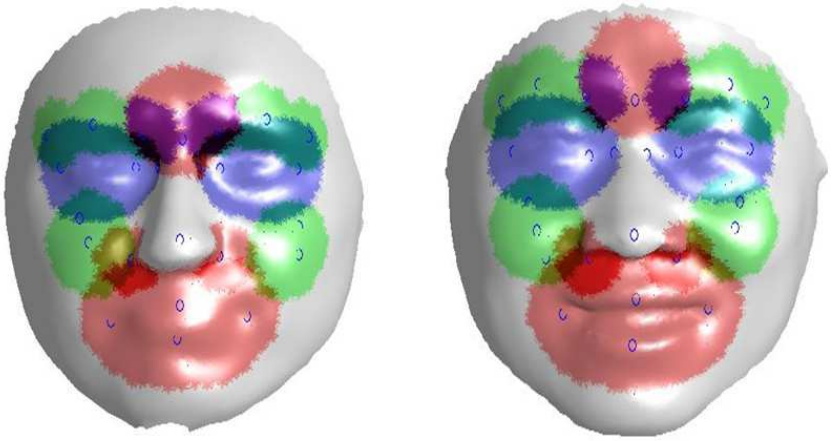


FIGURE 2.5 – Des exemples de segmentation de visages sur des modèles issus de la base BU3D. Les régions que nous avons choisies sont *1 : Narine gauche, 2 : Narine droite, 3 : Bouche, 4 : Oeil gauche, 5 : Oeil droit, 6 : Sourcil gauche, 7 : Sourcil droit, 8 : Joue gauche, 9 : Joue droite, 10 : Nasion et partie supérieure du visage.*

sation plus simple, et donc potentiellement moins puissante.

Dans un premier temps, nous divisons le visage en régions d'intérêt morphologiques et génériques plutôt qu'en correspondance avec des AUs spécifiques. Ce choix nous permet de nous appuyer sur des points fiduciaux (par ex. les coins des yeux, les narines, les coins de la bouche), que nous sommes capables de détecter relativement précisément (partie 1.8), plutôt que sur des considérations purement anatomiques. Ces régions sont sensées être saillantes vis à vis des déformations d'un visage soumis à expression. Une étude du système FACS (voire partie 1.3), d'interprétations artistiques des expressions [Flores 2005] et des résultats issus d'autres approches de l'état de l'art [Maalej *et al.* 2011b] nous ont conduit à diviser le visage en 10 régions d'intérêt (cf figure 2.5). Ces 10 régions regrouppent, à nos yeux, les zones d'effet des principaux muscles faciaux impliqués dans les expressions faciales, tout en demeurrant relativement simples à décrire et à localiser à l'aide des *landmarks* dont nous disposons.

Les régions que nous avons retenues sont les suivantes :

- *Narine gauche,*
- *Narine droite,*
- *Bouche,*
- *Oeil gauche,*
- *Oeil droit,*
- *Sourcil gauche,*
- *Sourcil droit,*

- *Joue gauche,*
- *Joue droite,*
- *Nasion et partie supérieure du visage.*

En pratique, nous avons effectué la segmentation en régions comme étant l'intersection entre la surface faciale et des sphères ou des ellipsoïdes de divers rayons, centrés autour des points d'intérêts automatiquement localisés. À noter que pour la région des joues, nous avons défini le centre des sphères comme le milieu du chemin géodésique reliant les coins extérieurs des yeux aux coins extérieurs de la bouche. Les rayons des ellipsoïdes sont fonctions de l'écartement entre leurs foyers, ce qui semble respecter les proportions anatomiques et morphologiques. Le fait qu'il existe des régions se recoupant et des zones du visage n'appartenant à aucune région ne pose pas problème, étant donné que l'étape d'alignement et de calcul de distance est indépendant d'une région à l'autre.

Nous avons utilisé ICP comme méthode d'alignement et de comparaison de surfaces. À l'inverse d'autres algorithmes comme la paramétrisation UV [Szeptycki *et al.* 2010] [Mpiperis *et al.* 2008] ou les chemins géodésiques [Maalej *et al.* 2011b], l'avantage principal d'ICP est sa capacité à compenser les erreurs d'alignement. Sous certains paramètres, ICP compense également les inconsistances aux bornes de la région envisagée. Ainsi, même si la région à recaler a été localisée imprécisément, ICP peut dans une certaine mesure la comparer à la zone correspondante sur le visage de référence. Par ailleurs, afin de limiter les problèmes liés à des contours de régions inconsistants, nous utilisons ICP entre la région d'intérêt et l'ensemble du visage référence.

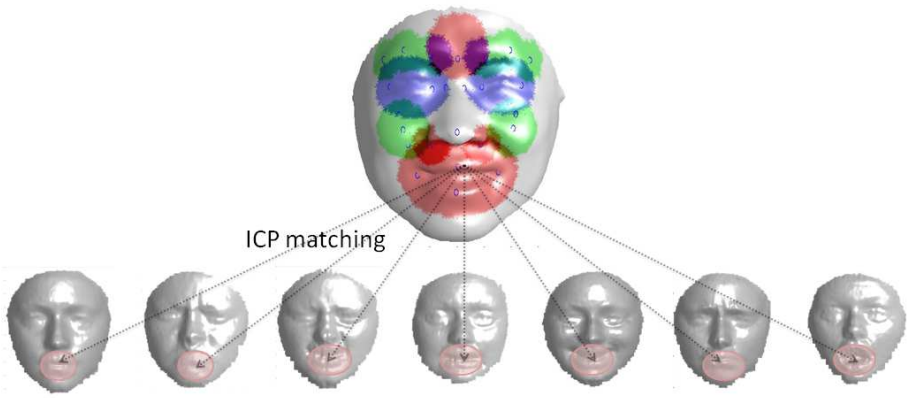
La distance au sens des moindres carrés fournie par ICP nous montre, pour chaque région, sa proximité avec chaque visage référence, et donc chaque expression prototypique (Figure 2.6). Ainsi, nous pouvons considérer la concaténation de l'ensemble de ces distances ICP –de chaque région à chaque expression– comme un descripteur pour le visage *probe*. Enfin, nous exploitons ce descripteur de manière tout à fait standard par le biais d'un classifieur, tel que les *Support Vector Machines* (SVM) [Franc & Hlavac 2002], pour les étapes d'apprentissage et de classification.

Dans la partie suivante, nous fournissons des résultats correspondant au classifieur *multi-class SVM*.

2.4.4 Résultats expérimentaux

Dans cette partie, nous montrons les résultats de l'expérimentation que nous avons menée de notre *framework* de reconnaissance d'expressions sur la base BU-3DFE (voir partie 1.2.2.3).

Le protocole d'expérimentation suivi est celui qui est décrit dans



region	neutral	anger	disgust	fear	happiness	sadness	surprise
1 (l. nostril)	4,2375	5,7211	5,8814	5,1220	**4,6714**	4,9906	4,5894
2 (r. nostril)	5,3418	2,9698	1,6227	5,7317	**2,0129**	3,3094	3,3974
3 (mouth)	11,1553	12,6463	2,3620	15,7645	**1,5372**	6,9487	10,0663
...	...	...	...	...	**...**	...	...
10 (u. part)	1,5536	2,0065	2,9002	1,1653	**1,7238**	2,0859	2,0850

FIGURE 2.6 – Un exemple des résultats du matching ICP par région pour un visage *probe* montrant une expression de Joie. Le tableau ci-dessus montre les distances moyennes ICP de chaque région du modèle *probe* vis-à-vis de chaque modèle référence

[Gong *et al.* 2009]. Jusqu'alors, le protocole généralement employé était le suivant : 40 personnes étaient utilisées pour l'apprentissage des paramètres de l'algorithme. Ensuite, pour la classification, les 60 personnes restantes sont utilisées dans un schéma devalidation croisée (apprentissage et test) répété 10 fois indépendamment. [Gong *et al.* 2009] ont démontré que ce protocole précis était trop sensible à la variance, étant donné que dans ce schéma, seules 6 personnes différentes sont utilisées pour le test (soit 72 visages de test). Les auteurs préconisent donc de répéter la validation croisée non pas 10, mais 1000 fois indépendemment. Ils ont ainsi reproduit un certain nombre d'expérimentations de l'état de l'art et montré des variations importantes dans les résultats obtenus selon leur protocole. Nous nous conformerons à ce protocole.

En tant que classifieur, nous avons utilisé les SVM multi-classe [Franc & Hlavac 2002]. Dans nos tests, les SVM étaient paramétrés avec un noyau linéaire et utilisaient la norme $L1$, avec un *trade-off* entre erreur et marge réglé à 1000. Ces paramètres ont été fixés empiriquement et l'ajustement de leur valeur n'a pas montré un grand impact sur les résultats, illustrés par le tableau 2.2. AN correspond à la Colère, DI au Dégoût, FE à la Peur, HA à la Joie, SA à la Tristesse et SU à la Surprise.

TABLE 2.2 – Matrice de confusion moyenne obtenue avec la méthode présentée dans ce chapitre.

%	**AN**	DI	FE	HA	SA	SU
AN	**69.4**	3.5	3.8	0	23	0.3
DI	10.4	**78.2**	4.3	1.3	0.5	5.3
FE	11.2	17.1	**42.8**	18.3	2.5	8.0
HA	0	2.5	8.3	**88.8**	0.3	0.2
SA	13.8	0.3	2.2	0.8	**82.9**	0.1
SU	0	0.5	3.9	1.3	1.8	**92.5**
Average	**75.8**					

TABLE 2.3 – Taux de reconnaissance moyen dans les travaux de Berretti [Berretti *et al.* 2010b], Gong [Gong *et al.* 2009], Wang [Wang *et al.* 2006a], Soyel [Soyel & Demirel 2007], Tang [Tang & Huang 2008] et la méthode proposée dans ce chapitre.

%	Berretti	Gong	Wang	Soyel	Tang	Méthode proposée
Avg	77.54	76.22	61.79	67.52	74.51	75.76

En raison des problèmes pointés dans [Gong *et al.* 2009] avec le protocole généralement employé dans l'état de l'art, nous ne fournirons ici qu'une comparaison avec les méthodes se pliant au même protocole, dans le tableau 2.3. Pour cette raison, nous n'avons volontairement pas mentionné les résultats de [Mpiperis *et al.* 2008] et [Maalej *et al.* 2011b]. Dans le tableau 2.3, Berretti correspond à [Berretti *et al.* 2010b], Gong à [Gong *et al.* 2009], Wang à [Wang *et al.* 2006a], Soyel à [Soyel & Demirel 2007], Tang à [Tang & Huang 2008]. Notons que les résultats reportés ici pour certaines des méthodes sont ceux recalculés par [Gong *et al.* 2009] dans le cadre de leur protocole expérimental.

Le tableau 2.3 montre que la méthode exposée dans cette section fait état de performances comparables aux approches de l'état de l'art, alors qu'elle est totalement automatique, au contraire des autres méthodes présentées dans ce tableau. Nous devons tempérer ces résultats en raison du manque de données comparables pour certaines méthodes *a priori* performantes, comme [Mpiperis *et al.* 2008] ou [Maalej *et al.* 2011b]. Aussi, notre méthode, bien que globalement performante, souffre de sérieuses lacunes concernant l'expression de Peur (FE), et dans une moindre mesure la Colère (AN). La raison sous-jacente est probablement que la réalisation de ces expressions déforme moins la surface faciale que des expressions telles que la Surprise, la Joie et le Dégoût. Une autre supposition que nous pouvons faire est que ces expressions, bien qu'engendrant des déformations particulières mesurées, sont caractérisées par une modification dans leur configuration spaciale.

C'est-à-dire, dans le déplacement relatif d'une région entière par rapport aux autres. Dans notre système, une modification de la configuration spaciale des régions n'aura que peu d'impact sur le score définitif, notre système n'étudiant que les déformations à un niveau local. Cette critique peut également être formulée à l'encontre des méthodes ne prenant en compte que le voisinage de points d'intérêts, comme [Berretti *et al.* 2010b] ou [Maalej *et al.* 2011b].

2.5 Résumé

Dans ce chapitre, nous avons vu comment envisager une approche régions dans les problèmes d'analyse de visage en 3D. Dans chacun des cas, nous avons défini une méthode automatique de localisation de ces régions, c'est-à-dire que l'humain n'intervient ni pour leur localisation, ni pour la détermination d'un ensemble de points d'intérêt permettant de les localiser. Dans la partie 2.3, nous avons également proposé une paramétrisation de la surface du visage en 3D robuste en pose. Dans la partie 2.4, nous avons mis en avant l'utilisation de régions pour compenser les erreurs inhérentes aux systèmes de localisation de points d'intérêt.

Nous avons pu mettre en avant une amélioration des performances vis-à-vis des approches holisitques rigides, grâce à leur application sur des sous-ensembles du visage 3D, suivie de l'application d'une technique de fusion adaptée. Cette approche a permis, selon les cas, une meilleure robustesse aux expressions, ou une meilleure robustesse à l'identité. Elle est fondée sur l'approche de Ekman, qui est caractérise l'expression humaine comme la somme de l'activation d'*Action Units*.

Cependant, nous avons relevé plusieurs écueils au cours de cette étude. Le premier est que la détermination et la localisation de telles régions ne sont pas triviales. Dans nos travaux, nous n'avons pas su faire usage de méthodes d'apprentissage pour les déterminer, mais nous avons fait appel à une expertise humaine. Le deuxième écueil est, selon nous, un risque de sur-spécialisation des techniques proposées. La tentative d'améliorer les résultats obtenus dans ce chapitre fait surtout apparaître l'idée selon laquelle, outre l'application d'une fusion pour combiner les résultats obtenus spécifiquement par région, il est sans doute bénéfique de lui adjoindre une méthode capable d'étudier les relations entre ces régions, et donc un niveau supplémentaire de sophistication.

Ainsi, il nous est apparu plus intéressant, pour la suite de nos travaux, de nous orienter vers des méthodes holistiques et systématiques comme celles proposées dans le chapitre 3.

Chapitre 3

Représentation par Cartes de Différence de Courbure Moyenne

3.1 Introduction

Comme nous avons pu le voir dans le chapitre 1, les deux grands courants de travaux dans le domaine de l'expression faciale s'appuient d'une part sur la localisation de points d'intérêts, et d'autre part sur des méthodes holistiques basées sur l'application de modèles déformables. Citons par exemple [Mpiperis *et al.* 2008], [Gong *et al.* 2009] et [Fang *et al.* 2011]. Le fait de se baser sur la localisation précise de points d'intérêt reste aujourd'hui problématique en raison de la qualité des algorithmes de localisation et de leur complexité, malgré la proposition que nous avons faite dans la partie 2.4. Un autre défaut de ce type de méthodes est la représentation assez partielle et ponctuelle des caractéristiques du visage. Par ailleurs les méthodes holistiques basées sur des modèles déformables sont basées sur des modèles sophistiqués dont la mise en œuvre et l'apprentissage sont complexes.

Cependant, dans l'état de l'art de la reconnaissance d'expressions en 2D, certaines méthodes holistiques proposent des approches systématiques, efficaces et peu exigeantes en termes d'apprentissage et de mise au point. Dans la suite de ce chapitre, nous nous inspirerons pour la reconnaissance d'expressions des travaux exposés dans [Dahmane & Meunier 2011]. Dans cet article, les auteurs décrivent le visage sous la forme d'une concaténation d'Histogrammes de Gradients Orientés *(Histograms of Oriented Gradients*, HOG), appliqués par régions selon une grille qui subdivise le visage en différents patches. La concaténation de ces histogrammes indépendants par subdivision est ensuite utilisée directement comme descripteur du visage dans un classifieur standard, comme les multi-class SVM [Franc & Hlavac 2002].

Par ailleurs, nous avons vu dans le chapitre 1 que [Huang *et al.* 2011] avait appliqué avec succès une approche SIFT [Lowe 2004] classique à des cartes représentatives extraites des images de profondeur pour la reconnaissance de personnes. Le principe est d'appliquer un opérateur mettant en avant les détails du visage par rapport aux images de profondeur, jugées peu distinctes et trop peu informatives telles quelles. Cependant, comme nous allons le voir dans la partie suivante, cette méthode souffre

de quelques lacunes, potentiellement dommageables dans un contexte d'analyse faciale en 3D.

Dans ce chapitre, nous allons définir un nouvel opérateur, et générer un ensemble de cartes que nous appellerons *Cartes de différence de courbure moyenne* (CDCM). Nous évaluerons ces cartes à la fois dans un contexte de reconnaissance d'expressions et de reconnaissance de personnes, en les exploitant de manière différente selon le scénario.

3.2 Une représentation intégrale de la surface 3D

Comme l'ont montré les auteurs de [Huang *et al.* 2011], dans un contexte de reconnaissance de visages en 3D, l'application d'une technique d'extraction de points d'intérêt telle que SIFT [Lowe 2004] directement sur les images de profondeur ne fournit pas de résultats probants. Afin de résoudre ce problème, les auteurs ont proposé l'application à l'image de profondeur de diverses variantes de Local Binary Patterns (LBP), appelées MS-eLBP (pour *Multi-Scale extended Local Binary Patterns*). L'idée est de générer une ou plusieurs cartes depuis l'image de profondeur à l'aide d'un opérateur, puis d'appliquer à ces cartes les méthodes de reconnaissance d'objets en 2D, en l'occurrence en reconnaissance faciale. Le score final de similarité est le résultat d'une fusion des scores obtenus sur chaque carte individuellement.

3.2.1 L'algorithme MS-eLBP comme base de départ

Dans cette partie, nous décrivons et analysons l'algorithme MS-eLBP tel que proposé par [Huang *et al.* 2011] pour la reconnaissance faciale.

3.2.1.1 Local Binary Patterns

LBP [Ojala *et al.* 2002] est un algorithme non paramétrique destiné originalement à décrire des textures d'images en 2D. Une de ses propriétés est sa tolérance aux variations d'illumination et sa simplicité algorithmique. Dans l'algorithme LBP de base, chaque pixel p_x reçoit un label sur la base d'un seuil appliqué aux pixels de son voisinage direct 3x3. Le principe est de parcourir successivement chacun des 8 pixels voisins $p_x^{1..8}$ de p_x. Si $p_x^i < p_x$, alors p_x^i se voit assigner le label $l_i = 1$. Sinon, il reçoit le label 0. La concaténation des labels $l_{1..8}$ forme un octet, traduit en valeur hexadécimale. L'ensemble des p_x d'une image décrite de cette manière forme ainsi une image de taille équivalente en niveau de gris. Appliqué à une image de profondeur, cet opérateur permet de distinguer de manière stable certaines topologies. Cependant, l'effet est limité à un voisinage de 1 pixel, et ne rend pas nécessairement

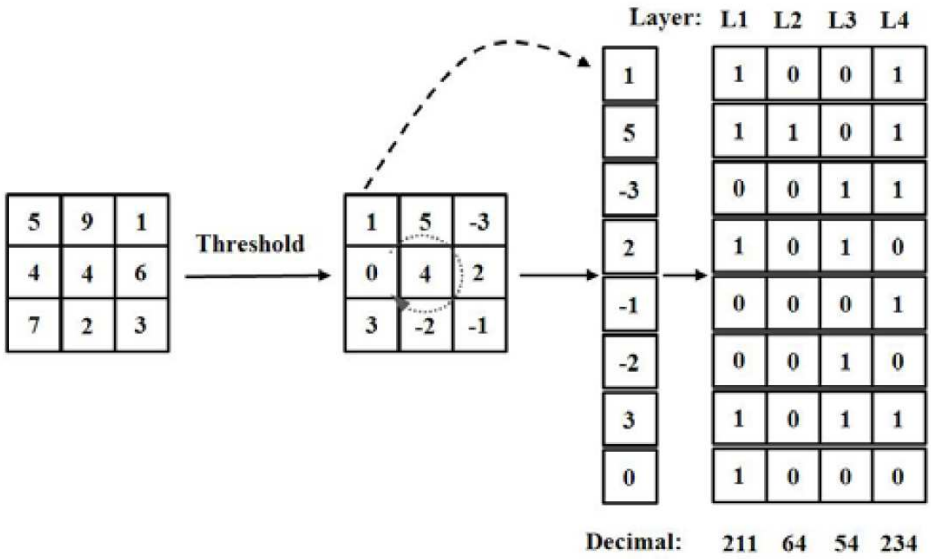


FIGURE 3.1 – Principe de fonctionnement de l'algorithme eLBP sur 4 niveaux. À gauche, la valeur des pixels que nous cherchons à décrire en représentation décimale. La différence entre le pixel central et les pixels voisins (diagramme du m !ilieu) est ensuite encodée en binaire signé, sur 4 bits en l'occurrence (tableau à droite). On ramène enfin ce tableau à une représentation hexadécimale en encodant les colonnes. Le niveau L1 correspond au signe de la différence et donc à l'opérateur LBP original. Le niveau L4 correspond à la parité de la différence (bit de poids faible) et les niveaux L2 et L3 aux bits de poids plus élevés.

compte de la topologie à une échelle supérieure. Elle est également dépendante de la résolution de l'image de profondeur. Les différentes propositions formulées dans [Huang *et al.* 2011] sont donc destinées à améliorer le pouvoir descriptif de LBP tout en conservant la même philosophie.

3.2.1.2 Géneralisation par extended-LBP

Les bits concaténés par LBP encodent le signe de la différence entre un pixel et ses voisins. L'idée suivie par les auteurs de [Huang *et al.* 2011] est d'encoder plus d'informations en décomposant la différence entre un pixel et ses voisins dans sa notation binaire. Ainsi, l'utilisation du bit de parité permet de former une nouvelle carte, et ainsi de suite pour les bits suivants. Un exemple de ce principe est donné en figure 3.1. La combinaison de plusieurs niveaux de eLBP permet d'encoder plus d'informations sur la topologie de la surface que l'application directe de LBP.

Expérimentalement, les auteurs ont trouvé que 4 niveaux supplémentaires étaient suffisants pour encoder la majorité de l'information.

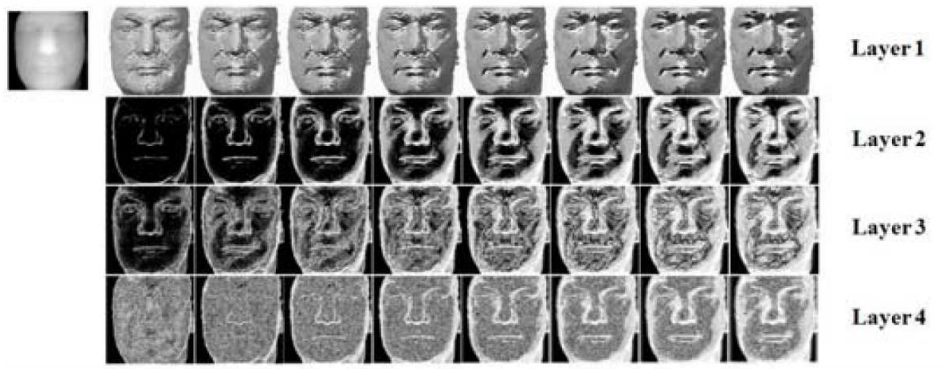


FIGURE 3.2 – Combinaison entre divers rayons (stratégie multi-échelles) et la approche eLBP sur un visage en 3D. De gauche à droite, on augmente le rayon d'un pixel et de bas en haut le niveau envisagé.

3.2.1.3 Stratégie multi-échelles

En plus d'enrichir les niveaux d'information concernant le voisinage direct, les auteurs de [Huang *et al.* 2011] ont appliqué une stratégie multi-échelles. Au lieu de considérer uniquement des pixels connexes, ils ont pris en compte des pixels à diverses distances du pixel central. Ils ont également envisagé des cas où plus de 8 pixels seraient considérés, passant donc l'information encodée à 12 ou 16 bits, voire plus.

Au total, dans cette approche, 140 cartes en 2D (niveaux de gris) ont pu être générées pour chaque visage, afin d'être exploitées ensuite par l'algorithme SIFT. La figure 3.2 montre l'application de cette stratégie à l'image de profondeur d'un visage en 3D.

3.2.2 Interprétation

De notre point de vue, en comparaison des images de profondeur originale, ces cartes ne retiennent que les détails de la surface 3D, à différents niveaux, tandis qu'elles permettent de s'affranchir de la topologie, très générale, du visage 3D. Cette dernière est particulièrement utile dans une situation de détection de visage, mais elle ne présente *a priori* que peu d'intérêt dans un scénario de reconnaissance faciale. LBP ainsi que ses dérivés permettent de mettre en avant des spécificités de la surface à un niveau très local. Dans certaines circonstances, une simple différence de bit engendre un changement fondamental de la valeur associée au pixel. Par exemple, si la seule différence entre p_x et son 8ème voisin p_x^8 passe de -1 à 1, alors la valeur associée à p_x par l'opérateur diffère de 128.

Cependant, la méthode proposée par [Huang *et al.* 2011] est potentiellement sen-

sible au bruit étant donné qu'une différence de valeur d'un pixel peut amener une différence considérable dans la valeur des cartes. On peut également noter que l'opérateur LBP traduit différemment des changements d'orientation, étant tributaire de l'ordre dans lequel le voisinage du pixel central est évalué. Par ailleurs, si la mise en avant de spécificités très ponctuelles est adaptée à un scénario de reconnaissance faciale, elle nous apparaît moins appropriée à un scénario de reconnaissance d'expressions, qui est un sujet qui nous intéresse également. Ainsi, nous cherchons un moyen de mieux mettre en avant, en regard de l'approche MS-eLBP, la topologie de la surface à une échelle plus large.

3.2.3 Cartes de Différence de Courbure Moyenne

Dans [Pottmann *et al.* 2007], les auteurs ont proposé une méthode de calcul intégral pour approximer la mesure de courbures. L'idée générale est de baser l'étude sur le calcul de volumes, d'aires ou de périmètres, plutôt que de mettre la surface en équation par le biais de tenseurs. Par exemple, le calcul de la courbure moyenne est particulièrement efficace. Pour un point p sur la surface d'un volume V, l'intersection $V_b(r,p)$ entre une sphère b de rayon r centrée en p et V donne :

$$V_b(r,p) = \frac{2\pi}{3}r^3 - \frac{\pi}{4}H(p)r^4 + O(r^5) \qquad (1)$$

où $H(p)$ correspond à la courbure moyenne au point p.

De l'équation (1), nous pouvons voir qu'il existe une corrélation directe entre $V_b(r,p)$ et $H(p)$. Cependant, H est homogène à r^{-1}. Étant donné que, pour la génération de cartes de niveaux, nous attendons des indices, nous définissons $h(r,p)$ tel que

$$h(r,p) = \frac{3}{4\pi r^3}V_b(r,p)$$

$h(r,p)$ est un indice, dont la valeur est comprise entre 0 et 1.

Plus r est petit, plus l'approximation de $H(p)$ formulée par $h(r,p)/r$ est précise. Néanmoins, comme le soulignent les auteurs de [Pottmann *et al.* 2007], en raison de la nature discrète des données manipulées dans des calculs réels, un rayon plus petit amène un plus grand niveau de bruit. De fait, nous pouvons tirer avantage de grands rayons, étant donné qu'ils mettent l'accent sur différentes échelles et niveaux sur l'objet 3D.

Nous avons étendu cette approche par l'utilisation d'un rayon extérieur r_o et d'un rayon intérieur r_i, et nous avons défini $h(r_i, r_o, p)$ comme

$$h(r_i, r_o, p) = \frac{3}{4\pi r_o^3}V_b(r_o,p) - \frac{3}{4\pi r_i^3}V_b(r_i,p)$$

avec $0 \leq r_i < r_o$.

L'objectif est d'être capable de ne pas prendre en compte les plus petits rayons, de sorte que le comportement de notre opérateur se rapproche légèrement plus de celui d'un filtre passe-bande au regard des échelles sur le visage 3D. Dans la suite de ce chapitre, nous appellerons Carte Différentielle de Courbure Moyenne (CDCM) $H_d(Im, r_i, r_o)$ une carte obtenue en calculant $h(r_i, r_o, p)$ à chaque point p de l'image de profondeur Im.

Sur un plan calculatoire, le volume $V_b(r, p)$ peut être approximé par un ensemble de cubes, considérés comme des volumes unitaires. Une image de profondeur d'un visage peut être regardée comme un volume, en appréciant le demi-espace situé derrière la surface du visage comme faisant partie du volume de la tête. Ainsi, $H_d(Im, r_i, r_o)$ peut être calculé efficacement et directement depuis l'image de profondeur, en faisant correspondre la résolution de l'image de profondeur et la résolution choisie pour les cubes unitaires. Par construction, un tel opérateur est invariant à la pose, tant que n'apparaissent pas des phénomènes d'auto-occultation.

3.3 Applications à la reconnaissance d'expressions du visage en 3D

Dans cette partie, nous présentons l'application des Cartes de Différence de Courbure Moyenne à la reconnaissance d'expressions du visage. Cette approche est nouvelle dans le domaine de la reconnaissance d'expressions du visage en 3D, en ce qu'elle constitue une approche holistique, complètement automatique et n'étant pas basée sur un modèle déformable. Notre objectif est, par ce moyen, de proposer une technique de reconnaissance d'expressions du visage en 3D dont la mise en œuvre et la mise au point est plus immédiate et plus accessible que celle des modèles déformables, tout en représentant la surface faciale de manière plus complète que les approches basées sur l'utilisation de points d'intérêt.

3.3.1 Méthode employée

De même que dans les travaux de [Huang *et al.* 2011], notre idée est de générer diverses CDCM $M_I(r_i, r_o)$, avec diverses valeurs de r_i et de r_o à partir d'une image de profondeur. Nous générons ensuite un vecteur descripteur pour chaque carte individuellement. Dans ce travail, nous avons choisi de reproduire l'approche simple et efficace proposée dans [Dahmane & Meunier 2011]. Enfin, l'étape de décision emploie l'ensemble des descripteurs précédemment produits. Il existe divers moyens de fusionner et de classer les descripteurs. Dans nos travaux, nous avons décidé de concaténer directement les vecteurs sans appliquer de technique de réduction de la

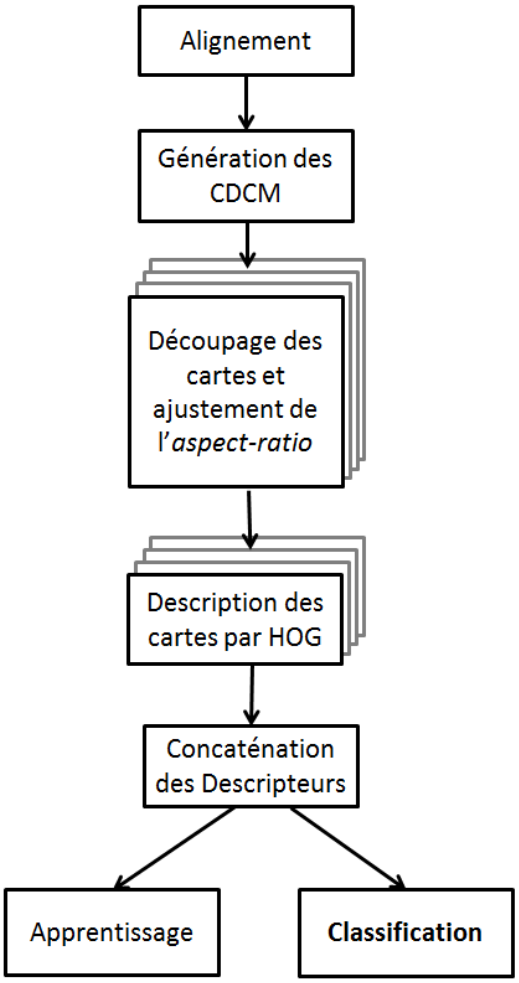


FIGURE 3.3 – *Framework* de notre méthode de reconnaissance d'expressions faciales basée sur les cartes de différence de courbure moyenne.

dimensionnalité, puis d'appliquer une méthode classique du *machine learning* avec les SVM multi-classe [Franc & Hlavac 2002], dans un souci de simplicité.

Le *framework* de cette approche est disponible dans la Figure 3.3.

3.3.1.1 Extraction des cartes

Dans un premier temps, nous rééchantillonnons les modèles 3D et générons l'image de profondeur correspondante. Le choix de la taille (en dimensions réelles) de la grille appliquée aux coordonnées X et Y est important, car il s'agit d'une balance entre la précision des calculs et le temps de calcul. Dans nos travaux, nous avons fixé cette grille à 0.8 mm. Les images de profondeur sont générées à l'aide d'une interpolation bilinéaire sur le maillage 3D, lui même obtenu à partir d'une triangulation de

TABLE 3.1 – Rayons utilisés dans notre approche de reconnaissance d'expressions (en mm).

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
r_i	0	3	$3\sqrt{2}$	6	$6\sqrt{2}$	12	$12\sqrt{2}$
r_o	$3\sqrt{2}$	6	$6\sqrt{2}$	12	$12\sqrt{2}$	24	$24\sqrt{2}$

Delaunay. L'ensemble a pour effet de lisser la surface et d'éviter la présence de trous. Nous générons ensuite les cartes de différence de courbure moyenne pour différents rayons, avant de procéder au recadrage du visage. L'objectif est de réduire l'impact des artefacts liés à la bordure de l'image. Les ensembles de rayons $S_k = (r_i^k, r_o^k)$ que nous avons utilisés sont détaillés dans le tableau 3.1. Ils ont été choisis de telle sorte à être répartis en octave, à l'image des rayons utilisés par la méthode SIFT. Les limites supérieures et inférieures des rayons choisis ont été choisies de sorte à mettre en évidence visuellement les expressions faciales au détriment des détails, et leur intérêt est confirmé expérimentalement dans les résultats présentés en 3.3.2.

3.3.1.2 Normalisation

L'étape suivante est le recadrage et la normalisation des cartes de visages. Nous recadrons en premier les visages avec une sphère centrée sur le bout du nez, dont le rayon a été fixé à 80 mm, à l'instar de la plupart des méthodes de l'état de l'art. Sur la base de données BU-3D, le bout du nez était détecté comme le point le plus proche du capteur dans une *bounding-box* vers le centre de l'image. Ce recadrage a pour but de concentrer nos algorithmes sur les parties informatives du visage humain, et de limiter les artefacts liés aux bords de l'objet 3D original dans le calcul des CDCM. Une fois ce recadrage appliqué sur les données 3D, nous restreignons l'image de la carte aux limites du visage. À ce stade, nous avons observé une grande disparité de la taille des images, liée à la morphologie des individus dans notre base de données. Selon la taille (en dimensions réelles) de leur visage et de leur forme, nous avons observé des tailles de cartes variées (en pixels) ainsi que des ratios hauteur sur largeur variés.

Expérimentalement, nous nous sommes rendus compte que déformer les images des personnes en les écrasant horizontalement ou verticalement par le biais d'une modification du format de l'image (le rapport entre largeur et hauteur) produit une amélioration sensible des performances de notre méthode. Nous avons donc redimensionné les CDCM pour obtenir des images de taille 240x200, telles que les limites du visage recadré à 80mm coincident avec la bordure de ces images. Nous avons

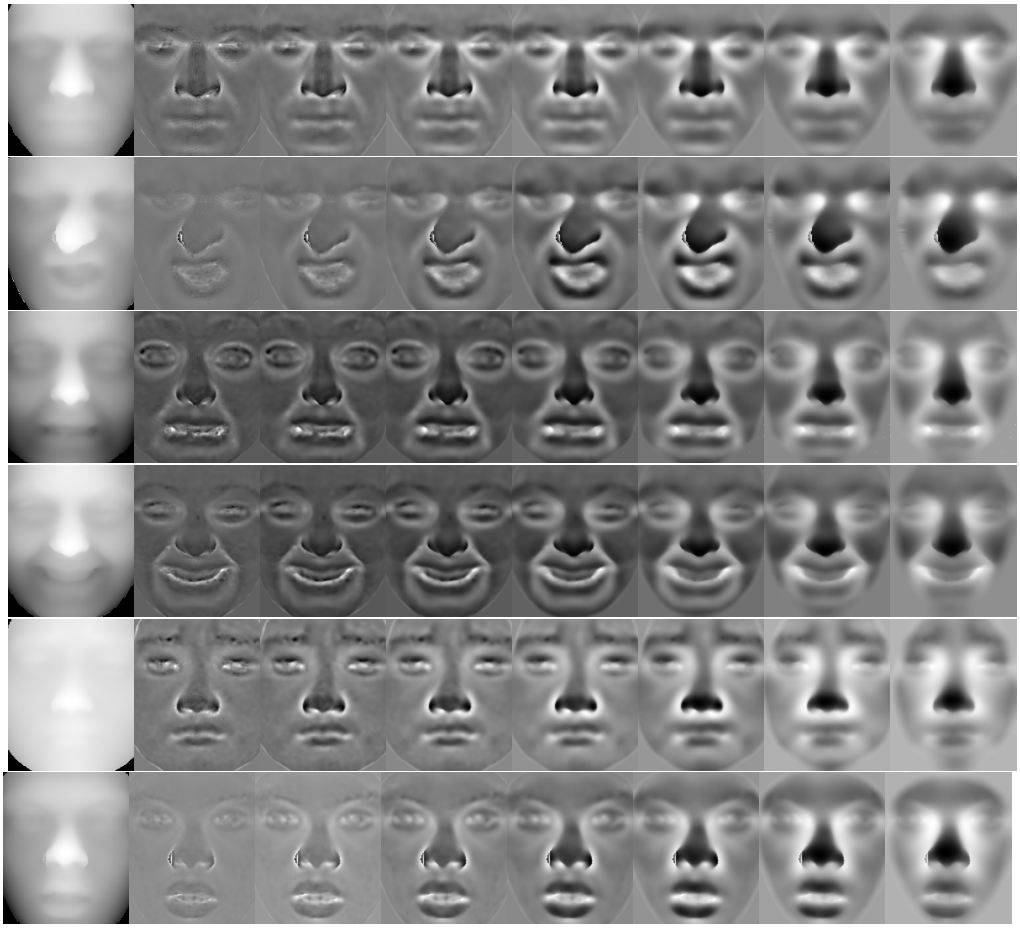


FIGURE 3.4 – Exemples d'application des CDCM après normalisation appliquées à des visages 3D expressifs. Les rayons correspondent à ceux évoqués dans la partie 3.3.1.1.

ensuite utilisé un *offset* de 8 pixels sur les coordonnées en X, pour mieux restituer l'information des sourcils, et nous avons enfin recadré cette image à une dimension de 180x150 pixels. Sur l'ensemble des modèles que nous avons pu observer visuellement, cette normalisation retient les informations importantes du visage expressif (sourcils - yeux - nez - bouche) tout en évitant autant que possible ses extrémités.

La figure 3.4 montre un exemple de ces CDCM normalisées, extraites depuis la base de données BU-3DFE en suivant la méthode décrite au paragraphe précédent.

3.3.1.3 Description

Une fois les différentes cartes générées, il est temps de procéder à la description et à la classification. À cette intention, nous avons employé simplement la méthode exposée dans [Dahmane & Meunier 2011]. Dans cet article, les auteurs ont d'abord

divisé les images de visage en différentes régions à l'aide d'une grille régulière. Ils ont ensuite extrait de chacune de ces subdivisions les Histogrammes de Gradients Orientés (Histograms of Oriented Gradients, HOG). Le principe de cette méthode est de calculer la direction du gradient en chaque point d'une image en niveaux de gris, puis d'échantillonner l'ensemble des directions. L'intensité du gradient dans un intervalle de direction donné est compilée en chaque point dans l'histogramme correspondant. De la sorte, un descripteur dont la dimension correspond au nombre de directions envisagées est généré pour chaque subdivision du visage. Ces descripteurs sont alors concaténés pour représenter l'ensemble du visage, puis ils sont utilisés dans une méthode d'apprentissage automatique (*machine learning*). Dans notre méthode, afin de rester simple et de ne pas nécessiter une étape d'apprentissage intermédiaire, nous avons également concaténé directement les descripteurs extraits de CDCM différentes (correspondant à diverses valeurs de (r_i, r_o)) comme méthode de fusion. Ces descripteurs sont par la suite exploités à l'aide d'un classifieur.

3.3.1.4 Classification

Pour ce dernier, nous avons décidé d'employer l'algorithme multi-class SVM [Franc & Hlavac 2002], fréquemment utilisé dans la littérature. Dans nos expérimentations, nous avons testé des subdivisions 5x5 et 6x6 des CDCM, ainsi que 9 directions signées dans l'algorithme HOG. Le choix des subdivisions a été guidé par des résultats préliminaires, ainsi qu'une observation visuelle (figure 3.5). Il semblerait que les différents élements du visage soient mieux répartis et isolés par de telles subdivisions. Une grille 7x7 engendrait une dimensionnalité jugée trop importante pour l'étape de la concaténation. Au total, chacune des 7 cartes faciales (dont les paramètres ont été évoqués précédemment) a été représentée par un vecteur descripteur de taille 225 (cas des subdivisions 5x5) ou 324 (cas des subdivisions 6x6). La concaténation de tous ces différents vecteurs descripteurs résulte en un vecteur de dimension 3843, ce qui est encore acceptable pour l'implémentation publique des SVM multi-classes sans avoir recours à une méthode de réduction de la dimensionnalité ou à une étape de fusion des scores.

3.3.2 Résultats expérimentaux

Dans la partie 3.3.2.1, nous décrivons d'abord les résultats obtenus en s'inscrivant dans le protocole défini dans [Gong *et al.* 2009], et que nous avons déjà évoqué dans la partie 2.4.4). De même, nous utiliserons les abréviations suivantes : Colère (AN), Dégoût (DI), Peur (FE), Joie (HA), Tristesse (SA) et Surprise (SU). Notre méthode ne recquérant pas d'apprentissage préliminaire à la classification, nous pouvons nous

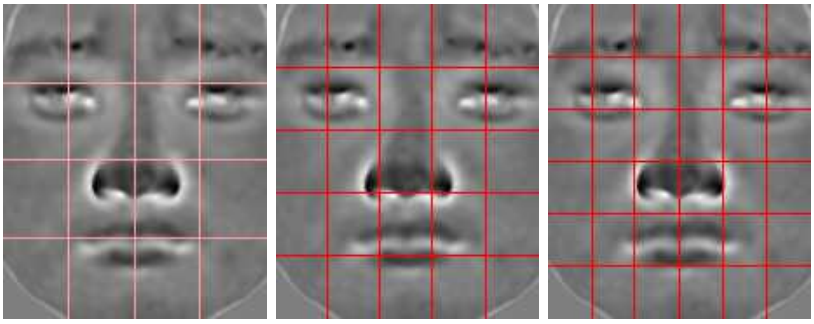


FIGURE 3.5 – Exemples de subdivision en grilles 4x4 (non utilisée dans nos expérimentations), 5x5 et 6x6 des visages normalisés par notre méthode.

permettre de tester la validation croisée directement sur les scans des 100 personnes constituant la base BU-3DFE. Les résultats de cette expérimentation sont présentés dans la partie 3.3.2.2.

3.3.2.1 Avec le protocole usuel.

Dans un premier temps, afin d'obtenir des résultats pleinement comparables avec l'état de l'art, nous avons tiré aléatoirement les 54 personnes sur lesquelles effectuer l'apprentissage et les 6 personnes sur lesquelles effectuer le test, à chaque itération de la validation croisée.

Nous présentons d'abord dans le tableau 3.2 le taux de reconnaissance moyen individuel pour chaque ensemble de cartes, combiné aux divisions des paramètres et aux rayons des CDCM. Dans ce tableau, S_{tous} correspond à la concaténation de toutes les cartes S_n, qui sont décrites dans la section 3.3.1.1. La dernière ligne donne le résultat pour la concaténation de toutes les combinaisons de rayons et de subdivisions employées pour le test de l'algorithme HOG. Nous faisons également apparaître le taux de reconnaissance obtenu avec HOG appliqué directement sur les images de profondeur.

Le tableau 3.3 montre la matrice de confusion moyenne obtenue par la concaténation des subdivisions 5x5 et 6x6 et de tous les rayons. Ce tableau correspond à la ligne S_{tous} présente dans le tableau 3.2.

Cette expérimentation montre que des résultats acceptables sont obtenus pour tous les rayons, tout en demeurrant complémentaires en vue d'une étape de fusion. Elle montre également que, dans le contexte de la reconnaissance d'expressions, nos cartes CDCM sont plus informatives que les images de profondeur originales. La présence d'un maximum consistant à S_4 pour les 2 subdivisions envisagées nous fait également penser qu'il existe une échelle typique à laquelle les expressions affectent particulièrement la surface 3D du visage.

TABLE 3.2 – Taux moyens de reconnaissance d'expressions en fonction des rayons choisis pour la génération des CDCM (voir partie 3.3.1.1) ainsi que les paramètres de découpage en régions de HOG. S_{tous} correspond à la concaténation de tous les rayons.

	Taux (%)	
DMCM	grille 5x5	grille 6x6
range	62.17	63.75
S_1	68.63	68.68
S_2	71.99	70.38
S_3	72.94	71.06
S_4	74.72	72.21
S_5	73.74	71.38
S_6	73.53	72.72
S_7	73.28	72.53
S_{tous}	75.78	75.99
S_{tous}	76.61	

TABLE 3.3 – Matrice de confusion moyenne obtenue avec S_{tous} et les grilles 5x5 et 6x6.

%	AN	DI	FE	HA	SA	SU
AN	**72**	7.3	4.5	0	17.9	0.25
DI	8.3	**74.9**	9.8	1.08	2.8	1.9
FE	4.17	10.5	**62.25**	11.17	5.5	3.67
HA	0.25	3	13.3	**86.42**	0.58	0.75
SA	14.33	2.5	5.17	0	**72**	1.33
SU	0.92	1.75	4.92	1.3	1.17	**92.08**
Moyenne	**76.6**					

TABLE 3.4 – Taux de reconnaissance moyen pour les méthodes proposées par [Berretti *et al.* 2010b], [Gong *et al.* 2009], [Wang *et al.* 2006a], [Soyel & Demirel 2007], [Tang & Huang 2008], le chapitre 2 et la méthode proposée ici.

%	Berretti	Gong	Wang	Soyel	Tang	Régions	Notre méthode
Reco	77.54	76.22	61.79	67.52	74.51	75.76	76.61

Dans le tableau 3.4, nous proposons également une comparaison avec les algorithmes de l'état de l'art s'étant pliés au même protocole expérimental. Nous montrons que notre méthode obtient donc des résultats comparables, tout en étant totalement automatique.

Si l'amélioration, bien que chiffrée, n'est pas très spectaculaire par rapport aux résultats présentés dans le chapitre précédent, la souplesse de la méthode dans son ensemble ainsi que les différents points d'amélioration possibles (description, fusion, classification), à l'aide de l'état de l'art de la reconnaissance d'expressions faciales en 2D, nous laissent croire à un intérêt réel de cette approche.

Notre méthode ne demandant pas d'apprentissage préalable à la description et à la classification des expressions, nous avons décidé de mener une étude exploitant cette fois la totalité des visages de la base de données BU-3DFE dans le scénario de validation croisée.

3.3.2.2 Expérimentation étendue

Dans cette partie, nous présentons l'expérimentation que nous avons conduite et qui inclut l'ensemble des 100 personnes de la base BU-3DFE dans le schéma de validation croisée. De nouveau, nous itérons en utilisant 90% des personnes présentes dans la base en apprentissage et 10% d'entre elles en test, en utilisant exclusivement les niveaux 3 et 4 d'intensité des 6 expressions prototypiques dans BU-3DFE. Cette fois, nous avons seulement itéré 200 fois chaque validation croisée, les résultats obtenus s'avérant plus stables. En effet, dans cette expérimentation, nous avons noté que l'écart-type entre le taux de reconnaissance des tirages effectués est deux fois moindre que dans l'expérimentation précédente. Le contenu et l'organisation des tableaux correspondants 3.5 et 3.6 est similaire à celui de la section précédente.

À notre surprise, le taux de reconnaissance est meilleur de manière consistante pour la totalité des expérimentations menées, bien que l'algorithme d'apprentissage utilise le même ratio entre apprentissage et test que dans l'expérimentation précédente. Nous interprétons ce phénomène comme étant une nouvelle mise en valeur des carences pointées dans [Gong *et al.* 2009]. Il laisse également espérer une amélioration des performances de l'algorithme avec des bases plus importantes, et donc

TABLE 3.5 – Taux de reconnaissance moyen dans notre expérimentation étendue pour différents paramètres de grille et rayons d'extraction des CDCM.

	Taux (%)	
DMCM	5x5 grid	6x6 grid
S_1	71.27	71.13
S_2	72.77	72.4
S_3	73.82	77.12
S_4	73.58	75.43
S_5	76.65	75.9
S_6	75.42	76.57
S_7	76.18	76.07
S_{tous}	76.68	78.1
S_{tous}	78.13	

TABLE 3.6 – Matrice de confusion moyenne obtenue avec S_{tous} et les grilles 5x5 et 6x6 dans notre expérimentation étendue.

%	AN	DI	FE	HA	SA	SU
AN	**74.1**	7.7	3.6	0	15.7	0
DI	8	**74.9**	12.3	1.7	3.6	1.3
FE	5.1	10.8	**64.6**	8.1	4.6	5.6
HA	0	3.1	10.7	**89.8**	0.6	1
SA	12.4	2.3	5.1	0	**74.5**	1.2
SU	0.4	1.2	3.7	0.4	1	**90.9**
Moyenne	**78.13**					

dans des champs d'application réelles.

3.4 Applications à la reconnaissance de visages en 3D

Bien qu'ayant proposé les CDCM comme le penchant de la méthode MS-eLBP pour un scénario de reconnaissance d'expressions du visage, nous avons souhaité expérimenter avec ces dernières dans un contexte de reconnaissance faciale. Un avantage potentiel des CDCM sur la méthode proposée par [Huang *et al.* 2011] est une meilleure intégration de la notion de rayon.

Par ailleurs, les CDCM sont potentiellement plus robustes aux variations de position. Bien que sensible à la pose lors de la projection d'un modèle en 3D vers une image de profondeur, le calcul des CDCM est, par construction, insensible aux variations de pose tant qu'on garantit l'absence de phénomènes d'auto-occultation. Celà signifie par exemple qu'à un même point du visage dans l'espace, entre une pose de profil et une pose de face, sera associée la même valeur dans l'espace des CDCM.

Ces points constituent selon nous des arguments en faveur d'une expérimentation dans un scénario de reconnaissance faciale exploitant les CDCM.

3.4.1 Méthode employée

Pour la reconnaissance de personnes, nous avons conservé une approche similaire à celle proposée dans [Huang *et al.* 2011], à savoir l'application directe de l'algorithme d'extraction de points d'intérêt SIFT, puis l'appariement pour chaque niveau S_n de CDCM défini par ses rayons intérieur r_i et extérieur r_o. Ici, le but est bien de tester la pertinence des CDCM que nous avons générées en comparaison avec le système proposé avec les MS-eLBP.

3.4.1.1 Génération des cartes.

La première partie de notre méthode, c'est-à-dire la génération des cartes, est similaire en tous points à celle que nous avons exposée en 3.3. Les échelles S_n que nous avons utilisées dans cette approche sont disponibles dans le tableau 3.7. Nous avons également utilisé les mêmes techniques de prétraitement et de projection des visages.

Ces derniers sont agencés selon une échelle d'octaves, à la manière des rayons des filtres gaussiens dans l'algorithme SIFT. Les rayons choisis semblent mettre en avant des aspects différents du visage humain.

TABLE 3.7 – Rayons utilisés dans notre approche de reconnaissance faciale (en mm).

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
r_i	0	3	$3\sqrt{2}$	6	$6\sqrt{2}$	12	$12\sqrt{2}$	24
r_o	3	$3\sqrt{2}$	6	$6\sqrt{2}$	12	$12\sqrt{2}$	24	$24\sqrt{2}$

	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}
r_i	0	3	$3\sqrt{2}$	6	$6\sqrt{2}$	12	$12\sqrt{2}$
r_o	$3\sqrt{2}$	6	$6\sqrt{2}$	12	$12\sqrt{2}$	24	$24\sqrt{2}$

3.4.1.2 Extraction, appariement de points d'intérêt et classification.

Nous appliquons ensuite directement la méthode d'extraction de points d'intérêt SIFT [Lowe 2004] sur chaque CDCM, sans autre normalisation. Les images générées à partir des visages ont ainsi des dimensions différentes, selon les morphologies. Cette différence est *a priori* d'un avantage, au sens de la différenciation intra-classe et extra-classe. Suite à cette étape, un visage est caractérisé par un ensemble de points d'intérêt SIFT distinctif pour chaque échelle de carte S_i.

À la suite de cette extraction, nous appairons chaque modèle *probe* à chaque modèle *gallery*, pour chaque échelle de carte S_i. De la même manière que dans les travaux de [Huang *et al.* 2011], nous appairons les points d'intérêt SIFT en mesurant leur distance *via* la fonction *arccosinus*. Nous apparions un point p_g^i du modèle *gallery* à un point p_p^j du modèle *probe* lorsque la distance entre p_g^i et p_p^j est inférieure à *ratio* fois la distance de p_g^i à tout autre point SIFT du modèle *probe*. En d'autres termes, nous n'appairons deux points SIFT p_g^i et p_p^j de deux cartes que si la distance les séparant est nettement inférieure à toute autre paire de points (p_g^i, p_p^k) où $k \neq i$. Expérimentalement, nous avons fixé le seuil *ratio* à 0.6. Il génère un nombre limité d'appariements, mais visuellement très cohérents sur des visages issus de la même personne.

La distance entre deux visages est ensuite directement déduite du nombre de points SIFT appairés entre deux visages. Un grand nombre d'appariement signifie une distance faible. Le visage *gallery* que nous associons à un modèle *probe* dans un scénario d'identification est le visage ayant eu le plus de correspondances de points SIFT. Dans un scénario de vérification, le seuil appliqué est ramené au nombre de points SIFT extraits du visage *gallery*. En effet, une divergence importante dans le nombre de points extraits, et donc de correspondances possibles, peut être observée d'une personne à une autre.

La fusion entre les divers niveaux de CDCM est simplement obtenue en addi-

TABLE 3.8 – Taux de reconnaissance de rang 1 ((RR1) de notre approche sur BU-3DFE dans un scénario *All vs Neutral* (AvN).

Cartes	RR1 AvN (%)
S_1	6.67
S_2	26.79
S_3	45.5
S_4	61.67
S_5	62.17
S_6	65.13
S_7	51.04
S_8	44.17
S_9	19.96
S_{10}	43.21
S_{11}	62.17
S_{12}	65.08
S_{13}	65.54
S_{14}	55.63
S_{15}	45.54
S_{tous}	84.17

tionnant le nombre de correspondances de chaque échelle.

3.4.2 Résultats expérimentaux.

Nous avons testé brièvement notre approche sur la base BU-3DFE, afin de montrer que la génération des CDCM pouvait être exploitée conjointement en reconnaissance de personnes et en reconnaissance d'expressions.

Le scénario envisagé ici est un *"All versus Neutral"*, où la *gallery* est constituée du modèle neutre de chaque personne (100 visages) et l'ensemble des modèles *probe* est l'ensemble modèles restants (ces visages sont donc tous expressifs et sont au nombre de 2400). Cette expérimentation est donc constituée de $2400 \times 100 = 240000$ comparaisons. Les résultats sont disponibles dans le tableau .

À titre de comparaison, les performances données pour l'approche de [Mpiperis *et al.* 2008] sur la base BU-3DFE sont environ de 86% de reconnaissance de rang 1 dans un scénario d'application relativement comparable.

Ces résultats montrent qu'il existe de nouveau une échelle et un ensemble de rayons optimaux pour l'exploitation des données du visage 3D. Bien choisies, les CDCM proposées dans ce chapitre sont exploitables non seulement en reconnaissance d'expressions, mais également en reconnaissance de personnes, y compris en présence d'expressions du visage. Le fait d'utiliser l'ensemble des échelles disponibles

produit d'excellents résultats, mais rallonge le temps de calcul de manière relativement importante. Toutefois, le scénario de fusion envisagé ici est simpliste, et laisse espérer une amélioration des performances, tant en termes d'efficacité calculatoire (élaguer au fur et à mesure en commençant par les rayons les plus favorables) qu'en résultats.

3.5 Résumé

Dans ce chapitre, nous avons proposé une approche holistique pour les problématiques d'analyse de visage en 3D, entièrement automatique. En reconnaissance d'expressions du visage en 3D, elle se compare aux autres méthodes holistiques, bien qu'elle ne se base pas sur une approche exploitant des modèles déformables. Elle se positionne donc comme une approche peu coûteuse et présente l'avantage d'une mise en œuvre et d'un apprentissage simples. Nous avons également pu montrer qu'elle était exploitable dans des scénarios de reconnaissance faciale, ce qui à terme nous permettrait d'envisager une reconnaissance conjointe de l'identité et des expressions, à l'instar de ce qu'ont proposé les auteurs de [Mpiperis *et al.* 2008] avec l'AFM.

Cette approche permet aussi d'exploiter les méthodes mises au point pour l'analyse de visages en 2D, avec l'espoir d'améliorer nettement les performances. Toutefois, si elle est efficace sur le plan calculatoire, un des principaux écueils rencontrés est sa sensibilité aux problèmes de pose. À ce propos, notons tout de même que, dans la plupart des cas, les méthodes de l'état de l'art requièrent également un recalage de la pose dans une très large proportion. Ce problème a été souvent traité avec plus ou moins de sophistication, comme dans [Spreeuwers 2011] ou [Drira *et al.* 2009].

Enfin, les auteurs de [Pottmann *et al.* 2007] ont montré qu'il était possible d'obtenir d'autres descripteurs de courbes *via* la technique de calcul intégral. Si de tels calculs sont plus coûteux que la technique que nous avons mise en œuvre dans ce chapitre, ils permettent d'envisager une description par cartes plus détaillée et générique.

Chapitre 4

Influence des dégradations sur les performances des algorithmes

4.1 Introduction

Il est généralement admis par la communauté, que les dégradations des modèles 3D ont un impact *a priori* négatif sur les performances des algorithmes, que ce soit en reconnaissance et identification de personnes ou en reconnaissance d'expressions. Par dégradations, nous entendons des altérations du modèle 3D, correspondant à la différence entre sa représentation numérique et le modèle qu'il est sensé émuler dans sa forme physique. Le nombre et la variété des méthodes de prétraitement évoquées (souvent succintement) dans la plupart des articles de notre état de l'art en forment une preuve implicite. Comme nous l'avons vu, des techniques de filtre moyen, gaussien ou median sur la carte de profondeur, ou des techniques de rééchantillonnage, de triangulation sont régulièrement employées. Plusieurs travaux proposent également l'emploi de techniques de suppression de trous de la surface et de pics. Certaines méthodes sont axées autour des tâches de prétraitement comme celles évoquées dans la partie 1.8. Dans ces études, le problème des dégradations de modèles 3D est donc admis et pris en compte. Néanmoins, il n'est pas évalué explicitement, ni approfondi.

D'autre part, des études centrées sur le problème des occultations (ou des données partielles) ont vu le jour ces dernières années, avec notamment la création d'une base de données telle que GavabDB [Moreno & Sánchez 2004] et de challenges associés. Toutefois, ces derniers travaux cités se limitent généralement à une classe précise de dégradations, certes essentielle, mais malheureusement parcellaires en regard de la diversité des dégradations auxquelles les algorithmes d'analyse faciale en 3D peuvent être confrontés.

Il semble raisonnable d'assumer que ce problème doit être inspecté et évalué plus avant dans l'hypothèse de l'application des algorithmes d'analyse faciale en 3D dans un cas réel.

4.2 Des dégradations canoniques

Dans un premier temps, il est nécessaire de déterminer quelles dégradations sont à prendre en considération, quelles sont leur origine, et la manière de les évaluer.

4.2.1 Origine des dégradations

Les dégradations peuvent découler de l'acquisition en 3D elle-même. Les scanners 3D au laser sont sensibles à la longueur d'onde des éclairages artificiels. Dans certains cas, une zone particulièrement réfléchissante sur le modèle scanné peut générer des trous ou des pics sur la surface 3D, que ce soit avec un scanner à technologie laser ou oculaire. Ce genre de défaut peut s'observer notamment sur la base FRGC v1, et les capteurs à laser intègrent en embarqué des solutions de filtrage (comme le filtrage médian chez Minolta). En tant que matériel électronique, les capteurs des appareils optiques sont sujets au bruit thermique (bruit de Nyquist-Johnson), au bruit Schottky, etc.

Il est également envisageable de considérer le modèle physique scanné comme source de bruit, ou plus précisément son comportement lors du scan. Il n'est en effet pas rare de constater dans les bases de données de visages en 3D des effets d'auto-occultations. Par exemple, le nez masque une narine ou un bout de la joue, ou encore un pan du visage n'est pas capté par le capteur en raison d'un angle défavorable. Autre phénomène parfois observable, le visage 3D déformé. Au cours de l'acquisition, le sujet a bougé, engendrant en une distorsion importante. Une étude rapide de ces cas dans la base FRGC a été conduite dans [Faltemier *et al.* 2008] ainsi que [Kakadiaris *et al.* 2007].

La compression, ainsi que l'échantillonnage, dans les situations où le volume des données est une contrainte, peuvent mener à des dégradations. La diminution de la résolution d'un modèle 3D introduit des distortions dans celui-ci, et l'étude de ces distorsions ainsi que leur réduction est un domaine de recherche à part entière [Alliez & Gotsman 2005]. La distance du sujet au scanner a également un impact sur la résolution du modèle 3D. Ces éléments sont à prendre en compte tout spécialement dans un scénario asymétrique de reconnaissance de personnes, où la base de données *gallery* peut etre constituée de modèles de grande qualité, mais o u les personnes à vérifier sont scannées dans des conditions peu voire non contrôlées, et/ou avec un matériel limité.

Enfin, les occultations sont un problème classique dans l'analyse de visages, *a fortiori* en 3D. Liées au port de vêtements, bijoux, lunettes, à d'autres éléments anatomiques (une main devant le visage, des cheveux) ou des éléments externes au sujet, elles ont également fait l'objet de bases de données dédiées. Bien que n'étant

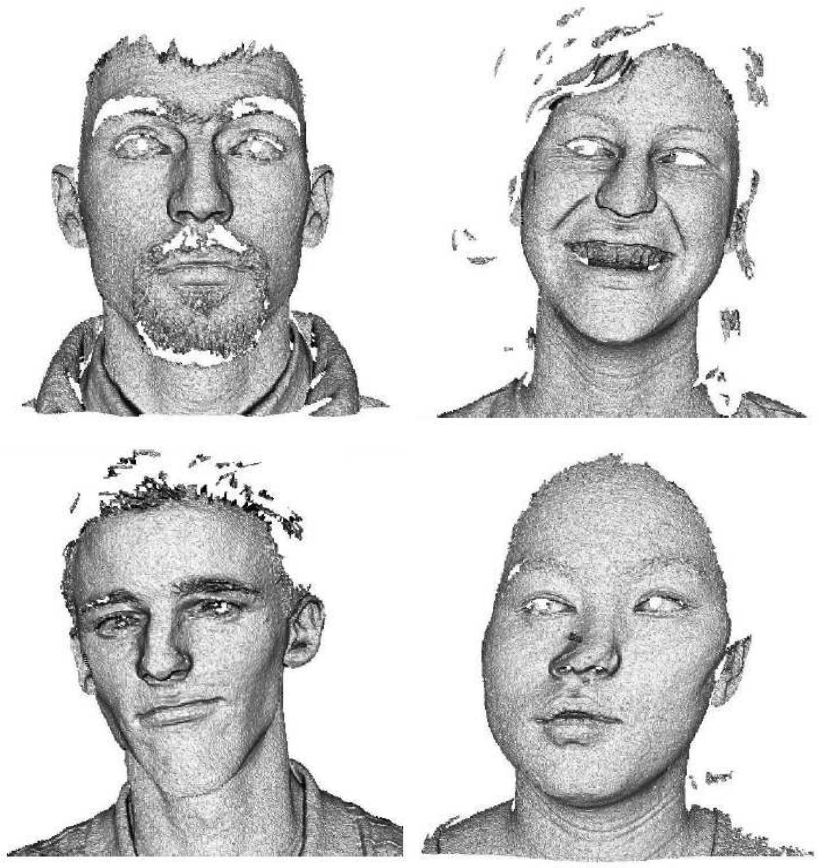


FIGURE 4.1 – Exemples de visages souffrant de défauts de qualité dans la base FRGC v1. Des trous et des artefacts sont observables à divers endroits du visage. Le modèle en bas à droite souffre d'une distorsion de mouvement.

pas des dégradations liées directement à l'acquisition, on peut les considérer comme telles puisqu'étant un frein à la qualité de l'acquisition de données.

Un exemple de visages dégradés directement extraits de la base de données FRGC v1 (voir partie 1.2.2.1) est visible en figure 4.1.

4.2.2 Méthodologie

Il apparait que les origines des dégradations sont multiples, que leur type est varié et que leur quantification est délicate. Mesurer leur impact sur les algorithmes d'analyse faciale semble problématique. En effet, une telle évaluation nécessite une reproductibilité, alors même qu'en cas réel, les origines des dégradations ne sont pas par essence controlées. Dans ce contexte, la démarche à adopter est de partir d'une base qu'on définira comme idéale, et de générer à partir de cette dernière des données dégradées, de manière relativement déterministe, et tout du moins reproductible.

La comparaison chiffrée des performances des algorithmes d'analyse faciale sur une base similaire, en l'absence et en présence de dégradations, nous permettra d'évaluer leur impact. Deux éléments restent à définir : la base de données sur laquelle nous allons travailler, considérée comme idéale ; et la nature, la forme et la magnitude des dégradations que nous allons lui appliquer.

4.2.2.1 Base de données choisie

La base que nous avons choisie pour cette étude est FRGC v2 (partie 1.2.2.1). A l'exception de certains modèles que nous prendrons le soin d'écarter, cette base est considérée comme relativement propre. Dans notre étude, il ne s'agit pas d'évaluer les performances des algorithmes en tant que tels, mais plutôt de comparer les performances des algorithmes en conditions dégradées, et surtout d'évaluer la pertinence d'une telle méthodologie. Nous nous restreindrons également à l'étude de l'impact des dégradations sur les algorithmes de reconnaissance et d'identification de personnes. L'étude de l'impact des dégradations sur les performances des algorithmes de reconnaissance d'expressions pourra faire l'objet d'une étude ultérieure suivant le même protocole.

Ainsi, nous avons choisi, pour des raisons liées aux temps de calcul, de baser notre étude sur un sous-ensemble de FRGC v2. Les visages *gallery* seront un sous-ensemble de FRGC v2 comprenant exclusivement des modèles de personnes avec une expression neutre. Les visages *probe* seront composés indifféremment de visages neutres ou expressifs, et seront au nombre de 1 par visage *gallery*. Les acquisitions de mauvaise qualité ont été rejetées. Au total, 410 visages *gallery* (pour autant de personnes différentes) et autant de visages *probe* ont été utilisés dans cette expérimentation.

Afin d'obtenir une base la plus propre possible, nous avons appliqué à chaque modèle de notre base de travail un filtre médian sur l'image de profondeur. Nous avons également recadré les visages en ne conservant que l'intersection entre le visage d'origine et une sphère centrée sur le bout du nez de rayon 100mm. Ces techniques étant déjà utilisées dans une grande partie de l'état de l'art, nous avons estimé qu'elles ne nuiraient pas aux performances des algorithmes testés, et qu'elles nous aideraient à obtenir une base la plus proche possible de l'idéal.

4.2.3 Dégradations considérées

L'étude des origines des dégradations nous a conduit a en définir trois types, que nous appellerons canoniques. À l'aide de ces dernières, nous pouvons décomposer une grande partie des dégradations évoquées précédemment et que nous rencontrons

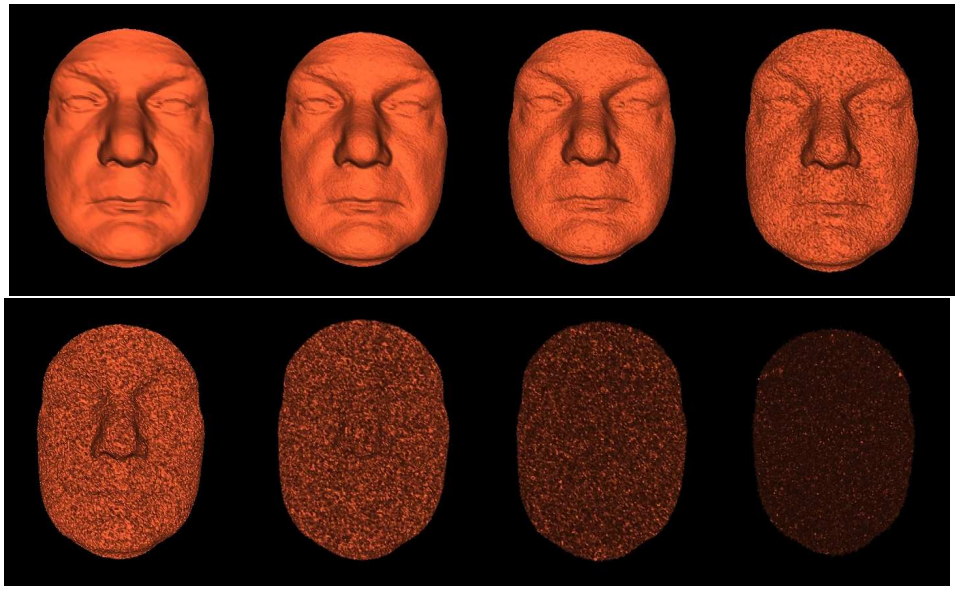


FIGURE 4.2 – Ajout de bruit sur les coordonnées en Z d'un visage. En bruit RMS et en mm, les valeurs sont, de gauche à droite et de haut en bas : 0 ; 0.1 ; 0.2 ; 0.4 ; 0.8 ; 1.6 ; 3.2 ; 6.4

effectivement dans des cas réels. Notre objectif sera, en plus de définir ces dégradations canoniques, d'être capables de les approximer génératívement. Dans la mesure où nous n'avons pas eu les moyens ni le temps de modéliser les dégradations rencontrées en cas réel, nous nous en tiendrons à des modélisations théoriques, synthétiques et simplistes, ce afin de valider cette approche dans un premier temps.

4.2.3.1 Le bruit

Le bruit apparait comme une première dégradation canonique. Résultant de l'imprécision des capteurs et de divers phénomènes physiques (dont le bruit thermique, le bruit schottky...), nous allons simplifier sa modélisation en le considérant comme Gaussien, bien qu'il soit souvent structuré en pratique. Pour le matérialiser, nous allons injecter une erreur, contrainte en valeur RMS, sur la coordonnée en Z des points du modèle 3D. Différentes valeurs d'erreur RMS permettent de quantifier plusieurs niveaux de dégradation.

4.2.3.2 La décimation

La décimation correspond à l'action de supprimer des points des données originales. Elle correspondrait physiquement à la réduction de l'échantillonnage, à l'augmentation de la distance du sujet avec le capteur ou à une compression. Pour la modéliser, nous sélectionnerons aléatoirement des points à retirer du maillage ini-

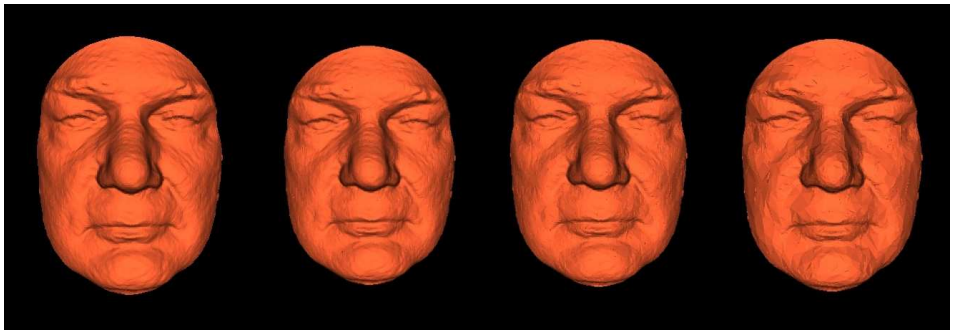


FIGURE 4.3 – Décimation d'un visage. En facteur de décimation, de gauche à droite : 1 (visage original sans affectation) ; 2 ; 4 ; 8

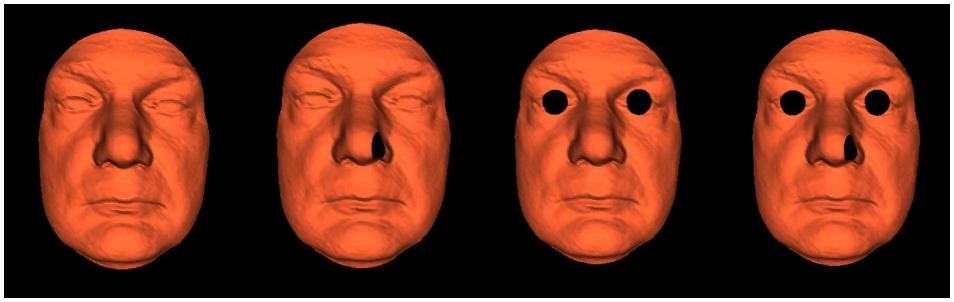


FIGURE 4.4 – Génération de lacunes sur un visage. En nombre de trous, de gauche à droite : 0 ; 1 ; 2 ; 3

tial. La proportion de points retirés par rapport au nombre de points du maillage initial nous servira de mesure d'intensité de la décimation. Une décimation de facteur 4 signifie que le maillage décimé comprend 4 fois moins de points que le maillage initial.

4.2.3.3 Les données lacunaires

Les lacunes correspondent, entre autres, aux problèmes d'auto-occultation et de trous constatés dans les cas où la surface scannée est trop réfléchissante. Les occultations, erreurs d'acquisition ponctuelles (pics, déformations locales) peuvent entrer dans cette catégorie, notamment si elles font l'objet d'un prétraitement visant à supprimer les éléments n'appartenant pas directement au visage du sujet dont on a fait l'acquisition. Nous simplifierons ces dégradations en les modélisant par des trous obtenus par l'intersection d'une sphère avec la surface faciale. Le nombre de sphères et leur diamètre servira à définir plusieurs niveaux dans ces données lacunaires.

FIGURE 4.5 – L'exemple d'un modèle dans les bases pour évaluer l'impact du bruit. De gauche à droite, O, $N2$, $N4$ et $N8$

FIGURE 4.6 – L'exemple d'un modèle dans les bases pour évaluer l'impact de la décimation. De gauche à droite, O, $D2$, $D4$ et $D8$

4.3 Résultats expérimentaux

4.3.1 Ensembles utilisés pour le test

La base *probe* originale (sans dégradations) sera notée O. Les bases dégradées que nous avons générées comportent les propriétés suivantes :

- Le bruit gaussien : les niveaux de bruit gaussien employés sont les suivants : 0,2 mm, 0,4 mm et 0,8 mm RMS. Les versions dégradées correspondantes de la base seront notées respectivement $N2$, $N4$ et $N8$.
- La décimation : les niveaux de décimation choisis sont les suivant : x2, x4 et x8. Les versions dégradées correspondantes de la base seront notées respectivement $D2$, $D4$ et $D8$.
- Les lacunes : nous avons supprimé du maillage son intersection avec respectivement 1, 2 ou 3 sphères de rayon 10 mm dont le centre est positionné aléatoirement sur la surface faciale. Pour des raisons de clarté, les versions dégradées correspondantes de la base seront notées respectivement $L2$, $L4$ et $L8$.

4.3.2 Quatre experts et leur fusion en concurrence.

Cette partie présente des résultats quant à l'analyse de l'influence des dégradations sur les performances des approches de reconnaissance de visages en 3D. Dans ces travaux, nous avons observé le comportement de quatre algorithmes différents,

FIGURE 4.7 – L'exemple d'un modèle dans les bases pour évaluer l'impact des données lacunaires. De gauche à droite, O, $L2$, $L4$ et $L8$

dits experts, de reconnaissance de visages, ainsi que de leur fusion. L'objet de ce chapitre n'étant pas centré sur les méthodes de fusion, nous ne les détaillerons pas outre mesure. Ces travaux ont fait l'objet d'une publication [Ben Soltana *et al.* 2012].

Dans cette étude, nous avons mis en concurrence quatre experts, que l'on nommera $E1$, $E2$, $E3$ et $E4$.

- $E1$ est une méthode utilisant les courbes géodésiques. Elle a été proposée par [Drira *et al.* 2010]. Son principe est d'assimiler le visage à un ensemble de courbes radiales, et d'étudier les déformations à appliquer à ces cartes pour passer d'un visage à un autre.
- $E2$ est l'expert MS-eLBP précédemment cité [Huang *et al.* 2011] et détaillé dans la partie 3.2.1.
- $E3$ est basé sur le recalage par TPS des visages sur le *template* de visage de la norme H.264 [Wiegand *et al.* 2003]. L'espace des paramètres de cette déformation est ensuite utilisé comme base pour le calcul de distance entre deux visages. Plus de détails sont disponibles dans [Erdogmus & Dugelay 2012].
- $E4$ correspond à l'algorithme *baseline* ICP [Zhang 1992] (partie 1.6.1).

Nous avons également envisagé divers scénarios de fusion. Pour rester succint, nous nous cantonnerons aux resultats obtenus par la fusion des résultats de $E1$, $E2$ et $E3$. La méthode de fusion employée consiste en une somme pondérée des scores des trois experts. La pondération du score d'un expert donné est proportionnelle au taux de reconnaissance de cet expert. Cette approche implique de diviser notre base d'étude en une base d'apprentissage pour les paramètres de pondération et une base de test pour évaluer les performances de la fusion de ces trois experts. Nous avons donc conduit une validation croisée de rapport 0.5 (bases d'apprentissage et test de taille similaire, soit la moitié de notre base d'origine chacune), itérée 50 fois. La performance finale que nous reportons est la moyenne des performances de chaque épreuve. Dans la suite de cette section, nous désignerons par F la fusion de $E1$, $E2$ et $E3$.

Le choix de ces experts est justifié par l'exploitation de différentes caractéris-

%	O	$N2$	$N4$	$N8$
$E1$	93.9	90.98	90.98	87.32
$E2$	93.41	89.27	90.73	72.93
$E3$	82.44	79.27	74.53	60.98
$E4$	71.28	71.01	69.34	58.88
F	99.77	97.01	96.68	91.13

TABLE 4.1 – Taux de Reconnaissance en Rang 1 des 4 experts et de leur fusion en présence de dégradations du type Bruit Gaussien.

%	O	$D2$	$D4$	$D8$
$E1$	93.9	82.44	80	66.59
$E2$	93.41	91.95	92.93	90.98
$E3$	82.44	80.98	76.59	67.8
$E4$	71.28	71.04	66.91	61.55
F	99.77	98.31	97.68	96.09

TABLE 4.2 – Taux de Reconnaissance en Rang 1 des 4 experts et de leur fusion en présence de dégradations du type Décimation.

tiques du visage, afin d'étudier leurs intéractions avec les différents types de dégradations. $E3$ est une méthode holistique, $E2$ est une méthode locale et $E1$ est hybride.

Les trois tableaux 4.3.2, 4.3.2 et 4.3.2 montrent le comportement de ces quatre experts

Ces données montrent d'une part l'appauvrissement, prévisible, des résultats suite aux différents types de dégradations. Chacun de ces types de dégradations a bien un impact sur les performances des algorithmes de reconnaissance faciale.

Par ailleurs, ils mettent en évidence un comportement différent des algorithmes vis à vis des dégradations envisagées. Ce résultat était déjà partiellement connu grâce à l'existence de bases de données telles que GavabDb [Moreno & Sánchez 2004],

%	O	$L2$	$L4$	$L8$
$E1$	93.9	83.41	81.22	80
$E2$	93.41	92.93	90.24	90.24
$E3$	82.44	81.22	80.98	81.46
$E4$	71.28	72.01	71.04	70.08
F	99.77	97.39	96.56	96.34

TABLE 4.3 – Taux de Reconnaissance en Rang 1 des 4 experts et de leur fusion en présence de dégradations du type Données Lacunaires.

où l'accent est mis sur les données lacunaires *via* des changements de pose. La présente étude met également en avant une influence diverse des dégradations sur les performances des algorithmes de reconnaissance faciale.

Plus précisément, l'expert $E1$ montre une meilleure robustesse au bruit gaussien que l'expert $E2$, et *a fortiori* l'expert $E3$ (tableau 4.3.2). ICP ($E4$) présente une bonne robustesse dans les premiers niveaux de dégradation, pour ensuite se détériorer fortement à $N8$. Une supposition que nous pouvons faire pour expliquer la meilleure résistance au bruit de $E1$ est le fait que les courbes radiales sont dans un premier temps interpolées avant d'être comparées. Les autres algorithmes répercutent quant à eux l'erreur dans la suite des calculs.

A contrario, l'algorithme $E2$ présente la meilleure robustesse vis à vis de la décimation (tableau 4.3.2). Dans ce scénario, $E3$ ainsi que $E4$ résistent mieux que $E1$. Dans l'implémentation de $E2$, la surface faciale est ramenée à une représentation d'image de profondeur en 2D sans trous grâce à une interpolation bilinéaire avant l'application de SIFT. Celà compense presque intégralement les effets de la décimation. $E1$, dans son implémentation, définit d'abord ses coupes radiales par l'ensemble de points du modèle 3D les plus proches du plan sécant. La décimation introduit ainsi une erreur qui, au contraire du bruit, n'a pas de raisons d'être corrigée par l'interpolation qui suit cette étape. Concernant $E2$, le recalage par TPS est effectué sur la base d'une grille appliquée au visage entre points d'intérêt. La décimation engendre ainsi une erreur au niveau de la précision de cette grille, qui se répercute faiblement sur l'ensemble de l'algorithme.

Enfin, l'étude sur les donées lacunaires (tableau 4.3.2) montre, comme celà était prévisible étant donné le type de techniques utilisées, une moins bonne résistance de l'algorithme $E1$ aux données lacunaires. La raison est que la présence d'un trou sur le tracé d'une courbe fausse largement le calcul subséquent. Il montre par ailleurs l'excellente robustesse des autres experts, y compris ICP, dont les performances ne sont quasiment pas affectées par la présence de trous sur la surface faciale. L'algorithme $E2$ est simplement privé d'un sous-ensemble des points utilisés pour effectuer l'appariement, ce qui n'affecte que peu l'algorithme ; $E3$ interpole les données manquantes grâce au modèle moyen auquel il se rattache. Ces résultats sont toutefois à tempérer par le fait que, dans notre étude, la pose était normalisée antérieurement à la dégradation. Une des difficultés majeures pour les algorithmes de reconnaissance faciale dans les situations où des données sont manquantes réside dans la difficulté de normaliser la pose.

Certains de ces résultats étaient relativement prévisibles. Il semble par exemple assez logique que les approches locales opposent une forte résistance aux données lacunaires. Il semblerait également que les méthodes holistiques ($E3$, mais aussi

ICP) soient assez robustes à la décimation et au bruit. La robustesse de ces derniers aux lacunes est selon nous plus heureuse, reposant principalement sur la méthode de paramétrisation de la surface en 3D. Certains résultats de cette étude trouvent par contre plutôt leur explication potentielle dans la manière dont les algorithmes sont implémentés, et plus particulièrement au niveau du prétraitement du modèle 3D. On peut imaginer une meilleure performance de $E1$ si l'extraction des courbes radiales était effectuée sur les images de profondeur générées dans le cadre de l'approche $E2$, plutôt qu'à partir des points du modèle 3D.

Il semble donc délicat de conclure suite à cette étude quant à une hypothétique robustesse intrinsèque des approches holistiques, locales ou hybrides vis à vis des différent types de dégradations canoniques. Certaines propriétés semblent liées d'avantage à l'implémentation qu'à la famille de méthode.

Enfin, nous pouvons noter dans ces résultats les excellentes performances de la fusion F des algorithmes $E1$, $E2$ et $E3$ dans chacun des cas de figure. Dans chacun des cas de figure, F est moins sensible aux dégradations que le meilleur des experts le composant. De même, ses performances sont systématiquement supérieures à celles du meilleur expert le composant. Ces résultats, relativement anecdotiques pour notre étude, démontrent l'intérêt et l'apport des approches de fusion sur les problématiques de reconnaissance faciale en 3D, et sans doute plus généralement d'analyse du visage en 3D. En plus de permettre d'excellents résultats, elles autorisent semble-t-il une sensibilité moindre aux difficultés rencontrées par chaque expert individuellement.

4.4 Résumé

Dans ce chapitre, nous avons développé une étude de l'impact des dégradations des modèles 3D sur les algorithmes d'analyse de visages en 3D. Notre approche a consisté à isoler plusieurs types de dégradations, dites canoniques, et de les appliquer à une base supposément exempte de défauts. L'application de diverses méthodes mises en concurrence permet ainsi de quantifier leur sensibilité à des formes de dégradations bien définies.

L'expérimentation que nous avons menée sur un ensemble de méthodes de reconnaissance faciale en 3D montre ainsi, entre différentes approches, des disparités dans leur comportement vis à vis de divers types de dégradations. Cette expérimentation valide la pertinence de notre approche, qui est de chercher à en isoler différentes composantes de manière indépendante. Leur génération, bien qu'utilisant des modèles simplistes par rapport aux situations rencontrées en cas réel, permet de les quantifier facilement, et par voie de conséquence autorise une étude approfondie des performances des algorithmes en leur présence. Nous avons mené notre expérimenta-

tion sur des méthodes de reconnaissance faciale, mais la méthodologie mise en place dans ce chapitre peut être facilement appliquée aux méthodes de reconnaissance d'expressions en 3D, voire à d'autres applications utilisant des modèles en 3D.

Par ailleurs, les résultats exposés dans la partie 4.3.2 montrent que, si certaines baisses de performances étaient prévisibles en raison de la nature des algorithmes testés, d'autres l'étaient à nos yeux moins, comme certains résultats relatifs à la décimation. Ces baisses de performances peuvent être liées à l'implémentation des algorithmes plutôt qu'à leur fondement mathématique et algorithmique. L'étude que nous avons menée rend possible la mise en évidence de telles carences. Enfin, les résultats exposés dans ce chapitre mettent en exergue l'apport des approches de fusion, non seulement dans les cas idéaux d'utilisation des méthodes employées lors de cette étude, mais surtout dans les cas où les données exploitées ne sont pas optimales.

Chapitre 5

Conclusion et perspectives

5.1 Contributions

Dans ce travail de recherche, nous nous sommes intéressés au problème de l'analyse de visages en 3D. Plus précisément, nous avons envisagé les problèmes de reconnaissance de personnes et d'expressions faciales par le prisme d'une approche régions ainsi que d'une technique par cartes de représentations. Nous nous sommes également intéressés à l'impact des dégradations sur les performances des algorithmes d'analyse de visage en 3D.

Les contributions sont les suivantes.

5.1.1 Approche régions

Ekman a étudié chez l'humain l'expression des émotions sur le visage. Ses travaux sont fondamentaux pour la communauté scientifique de l'analyse de visages en 3D, puisqu'ils sont à la base du protocole d'évaluation standard de reconnaissance d'expressions. À ce titre, les bases d'évaluation les plus employées comportent des échantillons des six expressions prototypiques décrites par Ekman. Il a également décomposé les expressions du visage sous forme d'*Action Units* (AU), qui consistent en l'activation de muscles ou de groupes de muscles faciaux. Dans nos travaux, nous avons souhaité exploiter ces résultats à travers une méthode basée sur la segmentation du visage 3D en régions. En s'inspirant des travaux d'Ekman concernant les principales AU impliquées dans la réalisation des expressions faciales prototypiques, nous avons proposé une segmentation adaptée, selon les cas, à la reconnaissance faciale ou à la reconnaissance d'expressions faciales. Pour l'une comme pour l'autre, cette segmentation a été autorisée par la localisation automatique de points d'intérêt anatomiques. Associée à une méthode de recalage rigide (Interative Closest Point, ICP) ainsi qu'à une technique de fusion simple, elle nous a permis d'observer une amélioration des résultats par rapport à l'application directe d'ICP, en tant que méthode holistique.

Une des contributions de nos travaux est la mise au point d'une paramétrisation du visage invariante en pose. Elle est basée sur l'emploi de distances géodésiques aux 3 points d'intérêt que sont les coins intérieurs des yeux et le bout du nez. Ces

points peuvent être localisés automatiquement et avec précision à l'aide de mesures de courbure et de seuillages successifs. Cette paramétrisation nous a permis de mener à bien une première segmentation du visage en régions en présence d'expressions, et de proposer une alternative à l'application classique d'ICP basée sur une paramétrisation euclidienne.

Dans le chapitre 2, nous avons également mis en avant le fait qu'ICP compense les erreurs liées aux techniques de localisation automatique de points d'intérêt. ICP nous permet, à travers le recalage rigide de régions, d'exploiter le voisinage des points d'intérêt plutôt que leur position relative.

5.1.2 Représentation de la surface 3D par Cartes de Différence de Courbure Moyenne

En reconnaissance d'expressions du visage en 3D, la plupart des méthodes se basent soit sur une approche holistique à modèle déformable, soit sur l'utilisation de points d'intérêt anatomiques ainsi que leur voisinage. La mise au point et la mise en œuvre de la première catégorie est complexe, tandis que la deuxième approche fait aujourd'hui encore largement appel à l'intervention humaine pour la localisation des points d'intérêt. Nous avons donc souhaité explorer une autre voie avec une technique de représentation par cartes. En effectuant un calcul intégral directement sur les images de profondeur, nous générons un ensemble de cartes représentatives de la topologie de la surface 3D, appelées Cartes de Différence de Courbure Moyenne (CDCM). Ces cartes en niveau de gris intègrent la notion d'échelle et permettent de mettre en évidence la courbure de la surface à différents niveaux.

Dans nos expérimentations, nous avons montré que les CDCM peuvent être exploitées à la fois en reconnaissance d'expressions et en reconnaissance faciale. En reconnaissance faciale, nous les avons associées à la méthode d'extraction et d'appariement de points caractéristiques saillants *Scale Invariant Feature Transform* (SIFT). Pour la reconnaissance d'expressions, nous avons exploité les CDCM grâce aux Histogrammes de Gradient Orienté (HOG) ainsi que les *Support Vector machine* (SVM) multi-classes.

Durant nos expérimentations sur la reconnaissance d'expressions faciales en 3D basée sur l'extraction de cartes de représentation CDCM, nous nous sommes également aperçus que la normalisation des images de visage jouait un rôle important quant aux performances de notre algorithme. En particulier, le choix du recadrage ainsi que celui du format de l'image (incluant le ratio hauteur sur largeur) affectent les scores de classification.

5.1.3 Un protocole pour l'étude de l'impact des dégradations des modèles 3D

Enfin, nous avons proposé une méthodologie destinée à mesurer l'impact des dégradations sur les modèles de visages en 3D relatif aux performances des algorithmes d'analyse faciale. À notre connaissance, leur impact est inconnu, ou du moins il n'est pas quantifié. Pourtant, et particulièrement en situation d'application sur le terrain, de nombreux défauts peuvent apparaître sur la qualité des scans en 3D. Ils peuvent avoir pour origine le capteur physique, les conditions d'acquisition, la transmission et la prise en charge des données 3D. Notre idée est de générer un ensemble de données dégradées, artificiellement, de manière quantifiable, à partir d'une base considérée comme exempte de défauts. L'étude des performances relatives des algorithmes en présence ou non des dégradations donne une indication chiffrée de leur robustesse. Dans nos travaux, nous avons considéré un ensemble de dégradations canoniques telles que le bruit gaussien, la décimation et les données lacunaires. L'expérimentation menée nous a conforté dans cette approche, et a validé notre méthodologie. Elle nous a permis de montrer que diverses méthodes réagissaient de manière variée à différents types de dégradations. Nous avons également pu vérifier que la fusion de multiples experts favorisait la résistance aux dégradations rencontrées.

5.2 Perspectives et travaux à venir

Dans les paragraphes suivants, nous proposons quelques perspectives quant à la poursuite des travaux présentés dans ce manuscrit.

5.2.1 Investigations potentielles quant aux approches régions

Lorsque nous avons implémenté notre approche régions pour la reconnaissance faciale, nous avons constaté que l'importance accordée aux régions par l'étape de fusion était, quasi inversement proportionnelle à leur superficie. La segmentation du visage en plusieurs régions peut présenter un risque : en augmentant le nombre de subdivision, on dissout potentiellement l'information. Il ne faut pas se contenter de sommer les résultats obtenus individuellement pour chaque région, mais il faut également analyser leurs relations spatiales. C'est un des éléments de l'approche de [Mian *et al.* 2008], qui analyse les graphes formés par les points appairés grâce à la similarité de leurs voisinages. Cette critique peut d'ailleurs être formulée à l'encontre de la méthode que nous avons mise au point en reconnaissance d'expressions (partie 2.4). Son principe est de quantifier les déformations au voisinage de points d'intérêt anatomiques localisés *via* la technique SFAM [Zhao *et al.* 2009a]. Cependant, elle ne

tient pas compte pour la classification des expressions des contraintes entre points d'intérêt, alors que ces dernières permettent aux auteurs de [Zhao *et al.* 2009a] de réaliser leur classification. Il nous semblerait judicieux, pour des travaux ultérieurs, de combiner les relations spaciales entre régions et les scores d'appariement individuels entre ces mêmes régions.

Dans le cas de la reconnaissance faciale comme dans celui de la reconnaissance d'expressions, un autre axe naturel de recherche serait d'expérimenter d'autres méthodes d'appariement entre surfaces, comme celles décrites dans la partie 1.6.

5.2.2 Investigations potentielles quant aux Cartes de Différence de Courbure Moyenne

Dans le chapitre 3, nous nous sommes inspirés des recherches de [Pottmann *et al.* 2007] pour définir nos Cartes de Différence de Courbure Moyenne (CDCM). Dans leurs travaux, les auteurs proposent également une méthode exploitant l'Analyse en Composantes Principales pour extraire des données de courbures supplémentaires, dont les courbures K_1 et K_2, et par déduction la courbure Gaussienne et le *Shape-Index*. Bien que nettement plus coûteuses sur le plan calculatoire, ces méthodes permettraient potentiellement de générer un nombre supérieur de cartes, mettant en avant d'autres aspects de la topologie de la surface faciale.

Un autre avantage des CDCM est la compacité de l'information extraite. Nous envisageons de mener des travaux sur leur application à la base de données BU-4DFE, c'est-à-dire à la reconnaissance d'expressions sur des séquences de visages en 3D dynamique.

5.2.3 Investigations potentielles quant aux dégradations sur les modèles en 3D

Naturellement, les études que nous avons menées regardant l'impact des dégradations sur les performances des algorithmes d'analyse de visages en 3D prendraient plus de sens en étant associées à une situation d'application réelle. Une fois le schéma d'application précisé, un affinement des modèles et des types de dégradations pourrait être une bonne perspective d'amélioration.

Liste des tableaux

Table des figures

Bibliographie

[Achermann & Bunke 2000] B. Achermann and H. Bunke. *Classifying range images of human faces with Hausdorff distance.* In Pattern Recognition, 2000. Proceedings. 15th International Conference on, volume 2, pages 809–813. IEEE, 2000. 36, 50

[Achermann *et al.* 1997] B. Achermann, X. Jiang and H. Bunke. *Face recognition using range images.* In Virtual Systems and MultiMedia, 1997. VSMM'97. Proceedings., International Conference on, pages 129–136. IEEE, 1997. 34, 50

[Alliez & Gotsman 2005] P. Alliez and C. Gotsman. *Recent advances in compression of 3D meshes.* Advances in Multiresolution for Geometric Modelling, pages 3–26, 2005. 92

[Amor *et al.* 2006] B.B. Amor, M. Ardabilian and L. Chen. *New experiments on icp-based 3d face recognition and authentication.* In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, volume 3, pages 1195–1199. IEEE, 2006. 44, 45, 50, 110

[Ao *et al.* 2009] M. Ao, D. Yi, Z. Lei and S.Z. Li. *Face recognition at a distance : System issues.* Handbook of Remote Biometrics, pages 155–167, 2009. 7

[Ben Amor 2006] Boulbaba Ben Amor. *Contributions à la modélisation et à la reconnaissance faciales 3D.* PhD thesis, 2006. 11, 61

[Ben Soltana *et al.* 2010] W. Ben Soltana, M. Ardabilian, L. Chen and C.B. Amar. *Adaptive Feature and Score Level Fusion Strategy Using Genetic Algorithms.* In Proc. Intern. Conf. Pattern Recognition (ICPR), volume 6, 2010. 47, 50

[Ben Soltana *et al.* 2012] W. Ben Soltana, M. Ardabilian, P. Lemaire, D. Huang, P. Szeptycki, L. Chen, N. Erdogmus, L. Daniel, J. Dugelay, B. Ben Amor*et al.* *3D face recognition : A robust multi-matcher approach to data degradations.* In Biometrics (ICB), 2012 5th IAPR International Conference on, pages 103–110. IEEE, 2012. 98

[Berretti *et al.* 2010a] S. Berretti, A.D. Bimbo, P. Pala, B.B. Amor and M. Daoudi. *A set of selected sift features for 3d facial expression recognition.* In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 4125–4128. IEEE, 2010. 29, 51

[Berretti *et al.* 2010b] S. Berretti, A.D. Bimbo, P. Pala, B.B. Amor and M. Daoudi. *A Set of Selected SIFT Features for 3D Facial Expression Recognition.* In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 4125 –4128, aug. 2010. 70, 71, 85, 107

[Blanz & Vetter 2003] V. Blanz and T. Vetter. *Face recognition based on fitting a 3D morphable model.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25, no. 9, pages 1063–1074, 2003. 16

[Browser 2012] Big Browser. *BIG BROTHER - Une caméra pour identifier des clients à partir de photos sur Facebook*, 2012. 8

[Bruce & Young 1986] V. Bruce and A. Young. *Understanding face recognition.* Br J Psychol, vol. 77, pages 303–327, 1986. 22

[Castellani *et al.* 2008] U. Castellani, M. Cristani, X. Lu, V. Murino and AK Jain. *HMM-based geometric signatures for compact 3D face representation and matching.* In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, pages 1–6. IEEE, 2008. 28, 50

[Chakraborthy *et al.* 2011] R. Chakraborthy, H. Rengamani, P. Kumaraguru and R. Rao. *The UID Project : Lessons Learned from the Wast and Challenges Identified for India.* Cyber Security, Cyber Crime and Cyber Forensics : Application and Perspective, Copyright, 2011. 8

[Dahmane & Meunier 2011] M. Dahmane and J. Meunier. *Emotion recognition using dynamic grid-based hog features.* In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pages 884–888. IEEE, 2011. 73, 78, 81

[Daniyal *et al.* 2009] F. Daniyal, P. Nair and A. Cavallaro. *Compact signatures for 3D face recognition under varying expressions.* In Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on, pages 302–307. IEEE, 2009. 28, 50

[Drira *et al.* 2009] H. Drira, B. Ben Amor, A. Srivastava and M. Daoudi. *A riemannian analysis of 3D nose shapes for partial human biometrics.* In Computer Vision, 2009 IEEE 12th International Conference on, pages 2050–2057. IEEE, 2009. 41, 50, 90

[Drira *et al.* 2010] H. Drira, B.A. Boulbaba, D. Mohamed, S. Anuj*et al. Pose and expression-invariant 3d face recognition using elastic radial curves.* In Proceeding of British machine vision conference, pages 1–11, 2010. 41, 42, 44, 50, 98, 110

[Ekman & Friesen 1971] Paul Ekman and Wallace V. Friesen. *Constants across cultures in the face and emotion.* Journal of Personality and Social Psychology, vol. 17, no. 2, pages 124–129, 1971. 23, 63

[Ekman *et al.* 2002] P. Ekman, W.V. Friesen and J.C. Hager. *The Facial Action Coding System.* In Research Nexus eBook, 2002. 23, 56

[Erdogmus & Dugelay 2012] N. Erdogmus and J.L. Dugelay. *On discriminative properties of TPS warping parameters for 3D face recognition.* In Informatics, Electronics & Vision (ICIEV), 2012 International Conference on, pages 225–230. IEEE, 2012. 98

[Faltemier *et al.* 2008] T.C. Faltemier, K.W. Bowyer and P.J. Flynn. *A region ensemble for 3-D face recognition.* Information Forensics and Security, IEEE Transactions on, vol. 3, no. 1, pages 62–73, 2008. 44, 50, 53, 61, 92

[Fang *et al.* 2011] T. Fang, X. Zhao, O. Ocegueda, SK Shah and IA Kakadiaris. *3D facial expression recognition : A perspective on promises and challenges.* In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pages 603–610. IEEE, 2011. 38, 51, 73

[Flores 2005] Victoria Contreras Flores. *Artnatomy/Artnatomia website*, 2005. 67

[Franc & Hlavac 2002] V. Franc and V. Hlavac. *Multi-class support vector machine.* In Pattern Recognition, 2002. Proceedings. 16th International Conference on, volume 2, pages 236 – 239 vol.2, 2002. 68, 69, 73, 79, 82

[Gökberk *et al.* 2005] B. Gökberk, A. Salah and L. Akarun. *Rank-based decision fusion for 3D shape-based face recognition.* In Audio-and Video-Based Biometric Person Authentication, pages 119–130. Springer, 2005. 46, 50

[Gong *et al.* 2009] Boqing Gong, Yueming Wang, Jianzhuang Liu and Xiaoou Tang. *Automatic facial expression recognition on a single 3D face by exploring shape deformation.* In Proceedings of the 17th ACM international conference on Multimedia, MM '09, pages 569–572, New York, NY, USA, 2009. ACM. 34, 51, 69, 70, 73, 82, 85, 107

[Grgic & Delac 2012] M. Grgic and K. Delac. *Face Recognition Homepage - http ://www.face-rec.org/databases/*, 2012. 21

[Gupta *et al.* 2007] S. Gupta, JK Aggarwal, M.K. Markey and A.C. Bovik. *3D face recognition founded on the structural diversity of human faces.* In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pages 1–7. IEEE, 2007. 28, 50, 109

[Gupta *et al.* 2010] Shalini Gupta, Mia K. Markey and Alan C. Bovik. *Anthropometric 3D Face Recognition.* Int. J. Comput. Vision, vol. 90, pages 331–349, December 2010. 27, 47

[Heseltine *et al.* 2004] T. Heseltine, N. Pears and J. Austin. *Three-dimensional face recognition : An eigensurface approach.* In Image Processing, 2004. ICIP'04. 2004 International Conference on, volume 2, pages 1421–1424. IEEE, 2004. 34, 50

[Hesher *et al.* 2003] C. Hesher, A. Srivastava and G. Erlebacher. *A novel technique for face recognition using range imaging.* In Signal processing and its applications, 2003. Proceedings. Seventh international symposium on, volume 2, pages 201–204. IEEE, 2003. 34, 50

[Hjaltason & Samet 2003] G.R. Hjaltason and H. Samet. *Properties of embedding methods for similarity searching in metric spaces.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25, no. 5, pages 530 – 549, may 2003. 64

[Huang *et al.* 2011] D. Huang, M. Ardabilian, Y. Wang and L. Chen. *A novel geometric facial representation based on multi-scale extended local binary patterns.* In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pages 1–7. IEEE, 2011. 30, 44, 50, 73, 74, 75, 76, 78, 87, 88, 98

[Jack *et al.* 2012] R.E. Jack, O.G.B. Garrod, H. Yu, R. Caldara and P.G. Schyns. *Facial expressions of emotion are not culturally universal.* Proceedings of the National Academy of Sciences of the United States of America, 2012. 25

[Jaimes & Sebe 2007] Alejandro Jaimes and Nicu Sebe. *Multimodal human-computer interaction : A survey.* Computer Vision and Image Understanding, pages 116–134, 2007. 8

[Kakadiaris *et al.* 2007] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis. *Three-dimensional face recognition in the presence of facial expressions : An annotated deformable model approach.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 4, pages 640–649, 2007. 38, 40, 44, 50, 53, 92, 110

[Koenderink & van Doorn 1992] J.J. Koenderink and A.J. van Doorn. *Surface shape and curvature scales.* Image and vision computing, vol. 10, no. 8, pages 557–564, 1992. 59

[Lee & Shim 2004] Y. Lee and J. Shim. *Curvature based human face recognition using depth weighted hausdorff distance.* In Image Processing, 2004. ICIP'04. 2004 International Conference on, volume 3, pages 1429–1432. IEEE, 2004. 37, 50

[Li & Zhang 2007] X. Li and H. Zhang. *Adapting geometric attributes for expression-invariant 3D face recognition.* In Shape Modeling and Applications, 2007. SMI'07. IEEE International Conference on, pages 21–32. IEEE, 2007. 46, 50

[Lowe 2004] D.G. Lowe. *Distinctive image features from scale-invariant keypoints.* International journal of computer vision, vol. 60, no. 2, pages 91–110, 2004. 16, 30, 73, 74, 88

[Lu & Jain 2005] X. Lu and A.K. Jain. *Deformation analysis for 3D face matching.* In Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on, volume 1, pages 99–104. IEEE, 2005. 37, 39, 50, 110

[Lu *et al.* 2004] X. Lu, D. Colbry and A.K. Jain. *Three-dimensional model based face recognition.* In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, volume 1, pages 362–366. IEEE, 2004. 33, 50, 110

[Maalej *et al.* 2011a] A. Maalej, B.B. Amor, M. Daoudi, A. Srivastava and S. Berretti. *Shape analysis of local facial patches for 3D facial expression recognition.* Pattern Recognition, vol. 44, no. 8, pages 1581–1589, 2011. 41, 43, 44, 51, 110

[Maalej *et al.* 2011b] Ahmed Maalej, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava and Stefano Berretti. *Shape analysis of local facial patches for 3D facial expression recognition.* Pattern Recognition, vol. 44, no. 8, pages 1581–1589, 2011. 67, 68, 70, 71

[Maes *et al.* 2010] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens and D. Vandermeulen. *Feature detection on 3D face surfaces for pose normalisation and recognition.* In Biometrics : Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pages 1–6. IEEE, 2010. 32, 50

[Matsumoto *et al.* 2009] David Matsumoto, Andres Olide, Joanna Schug, Bob Willingham and Mike Callan. *Cross-Cultural Judgments of Spontaneous Facial Expressions of Emotion.* Journal of Nonverbal Behavior, vol. 33, pages 213–238, 2009. 10.1007/s10919-009-0071-4. 25

[Medioni & Waupotitsch 2003] G. Medioni and R. Waupotitsch. *Face modeling and recognition in 3-D.* In Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on, pages 232–233. IEEE, 2003. 33, 50

[Mian *et al.* 2007] A.S. Mian, M. Bennamoun and R. Owens. *An efficient multimodal 2D-3D hybrid approach to automatic face recognition.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 11, pages 1927–1943, 2007. 46, 50, 53

[Mian *et al.* 2008] A.S. Mian, M. Bennamoun and R. Owens. *Keypoint detection and local feature matching for textured 3D face recognition.* International Journal of Computer Vision, vol. 79, no. 1, pages 1–12, 2008. 30, 31, 50, 105, 109

[Mikolajczyk & Schmid 2005] K. Mikolajczyk and C. Schmid. *A performance evaluation of local descriptors.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 10, pages 1615–1630, 2005. 16

[Moreno & Sánchez 2004] A. B. Moreno and A. Sánchez. *GavabDB : a 3D Face Database.* In Workshop on Biometrics on the Internet, pages 77–85, Vigo, March 2004. 21, 91, 99

[Mpiperis *et al.* 2007a] I. Mpiperis, S. Malasiotis and M.G. Strintzis. *3D face recognition by point signatures and iso-contours.* Proc. of SPPRA, 2007. 43, 50

[Mpiperis *et al.* 2007b] I. Mpiperis, S. Malassiotis and M.G. Strintzis. *3-D face recognition with the geodesic polar representation.* Information Forensics and Security, IEEE Transactions on, vol. 2, no. 3, pages 537–547, 2007. 34, 35, 50, 110

[Mpiperis *et al.* 2008] I. Mpiperis, S. Malassiotis and M.G. Strintzis. *Bilinear Models for 3-D Face and Facial Expression Recognition.* Information Forensics and Security, IEEE Transactions on, vol. 3, no. 3, pages 498 –511, sept. 2008. 38, 50, 68, 70, 73, 89, 90

[Ojala *et al.* 2002] T. Ojala, M. Pietikainen and T. Maenpaa. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, pages 971–987, 2002. 74

[Otsuka *et al.* 2007] Kazuhiro Otsuka, Hiroshi Sawada and Junji Yamato. *Automatic inference of cross-modal nonverbal interactions in multiparty conversations : "who responds to whom, when, and how ?" from gaze, head gestures, and utterances.* In Proceedings of the 9th international conference on Multimodal interfaces, ICMI '07, pages 255–262, New York, NY, USA, 2007. ACM. 8

[Pan *et al.* 2003] G. Pan, Z. Wu and Y. Pan. *Automatic 3D face verification from range data.* In Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on, volume 3, pages III–193. IEEE, 2003. 36, 50

[Phillips *et al.* 2005] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek. *Overview of the Face Recognition Grand Challenge*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pages 947–954, 2005. 17

[Pottmann *et al.* 2007] H. Pottmann, J. Wallner, Y.L. Yang, Y.K. Lai and S.M. Hu. *Principal curvatures from the integral invariant viewpoint.* Computer Aided Geometric Design, vol. 24, no. 8, pages 428–442, 2007. 77, 90, 106

[Queirolo *et al.* 2010] C.C. Queirolo, L. Silva, O.R.P. Bellon and M.P. Segundo. *3D face recognition using simulated annealing and the surface interpenetration measure.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 2, pages 206–219, 2010. 37, 50

[Russ *et al.* 2004] T.D. Russ, M.W. Koch and C.Q. Little. *3D facial recognition : a quantitative analysis.* In Security Technology, 2004. 38th Annual 2004 International Carnahan Conference on, pages 338–344. IEEE, 2004. 37, 50

[Samir *et al.* 2006] C. Samir, A. Srivastava and M. Daoudi. *Three-dimensional face recognition using shapes of facial curves.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, no. 11, pages 1858–1863, 2006. 41, 50

[Savran *et al.* 2008] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur and Lale Akarun. *Biometrics and Identity Management.* In Ben Schouten, Niels Christian Juul, Andrzej Drygajlo and Massimo Tistarelli, editeurs, Biometrics and Identity Management, chapitre Bosphorus Database for 3D Face Analysis, pages 47–56. Springer-Verlag, Berlin, Heidelberg, 2008. 18

[Smuts 1926] Jan Christiaan Smuts. Holism and evolution,. New York,The Macmillan company,, 1926. http ://www.biodiversitylibrary.org/bibliography/4568. 32

[Soltana *et al.*] W.B. Soltana, D. Huang, M. Ardabilian, L. Chen and C.B. Amar. *Comparison of 2D/3D Features and Their Adaptive Score Level Fusion for 3D Face Recognition.* 61, 62

[Soyel & Demirel 2007] Hamit Soyel and Hasan Demirel. *Facial Expression Recognition Using 3D Facial Feature Distances.* In ICIAR'07, pages 831–838, 2007. 29, 51, 70, 85, 107

[Spreeuwers 2011] L. Spreeuwers. *Fast and accurate 3d face recognition.* International journal of computer vision, vol. 93, no. 3, pages 389–414, 2011. 44, 50, 90

[Szeptycki *et al.* 2009] Przemyslaw Szeptycki, Mohsen Ardabilian and Liming Chen. *A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking.* In Proceedings of the 3rd IEEE international conference on Biometrics : Theory, applications and systems, BTAS'09, pages 32–37, Piscataway, NJ, USA, 2009. IEEE Press. 27, 47, 48, 57, 111

[Szeptycki *et al.* 2010] Przemyslaw Szeptycki, Mohsen Ardabilian, Liming Chen, Wei Zeng, David Gu and Dimitris Samaras. *Conformal mapping-based 3D*

face recognition. In 3DPVT 2010 - Fifth International Symposium on 3D Data Processing, Visualization and Transmission, pages 1–8, may 2010. 68

[Tang & Huang 2008] Hao Tang and T.S. Huang. *3D facial expression recognition based on automatically selected features.* In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on, pages 1 –8, june 2008. 29, 51, 70, 85, 107

[Tola *et al.* 2010] E. Tola, V. Lepetit and P. Fua. *Daisy : An efficient dense descriptor applied to wide-baseline stereo.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 5, pages 815–830, 2010. 16

[Wang *et al.* 2006a] Jun Wang, Lijun Yin, Xiaozhou Wei and Yi Sun. *3D facial expression recognition based on primitive surface feature distribution.* In in Proc. Conf. Computer Vision and Pattern Recognition, pages 1399–1406, 2006. 70, 85, 107

[Wang *et al.* 2006b] Y. Wang, G. Pan, Z. Wu and Y. Wang. *Exploring facial expression effects in 3D face recognition using partial ICP.* Computer Vision–ACCV 2006, pages 581–590, 2006. 33, 50

[Wang *et al.* 2010] Y. Wang, J. Liu and X. Tang. *Robust 3D face recognition by local shape difference boosting.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 10, pages 1858–1870, 2010. 34, 50, 51

[Wiegand *et al.* 2003] T. Wiegand, G.J. Sullivan, G. Bjontegaard and A. Luthra. *Overview of the H. 264/AVC video coding standard.* Circuits and Systems for Video Technology, IEEE Transactions on, vol. 13, no. 7, pages 560–576, 2003. 98

[Wu *et al.* 2004] Zhaohui Wu, Yueming Wang and Gang Pan. *3D face recognition using local shape map.* In Image Processing, 2004. ICIP '04. 2004 International Conference on, volume 3, pages 2003 – 2006 Vol. 3, oct. 2004. 28, 50

[Yin *et al.* 2006] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang and Matthew J. Rosato. *A 3D Facial Expression Database For Facial Behavior Research.* In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, FGR '06, pages 211–216, Washington, DC, USA, 2006. IEEE Computer Society. 20

[Yin *et al.* 2008] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm and Michael Reale. *A high-resolution 3D dynamic facial expression database.* In FG'08, pages 1–6, 2008. 21

[Zhang 1992] Zhengyou Zhang. *Iterative Point Matching for Registration of Free-form Curves*, 1992. 32, 33, 60, 98

[Zhao *et al.* 2009a] X. Zhao, E. Dellandréa and L. Chen. *A 3d statistical facial feature model and its application on locating facial landmarks.* In Advanced Concepts for Intelligent Vision Systems, pages 686–697. Springer, 2009. 48, 49, 51, 105, 106

[Zhao *et al.* 2009b] Xi Zhao, Emmanuel Dellandréa and Liming Chen. *A 3D Statistical Facial Feature Model and Its Application on Locating Facial Landmarks.*

In Jacques Blanc-Talon, Wilfried Philips, Dan Popescu and Paul Scheunders, editeurs, Advanced Concepts for Intelligent Vision Systems, volume 5807 of *Lecture Notes in Computer Science*, pages 686–697. Springer Berlin / Heidelberg, 2009. 27, 64

[Zhao *et al.* 2010] X. Zhao, D. Huang, E. Dellandréa and L. Chen. *Automatic 3D facial expression recognition based on a Bayesian Belief Net and a Statistical Facial Feature Model.* In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 3724–3727. IEEE, 2010. 29

[Zhao *et al.* 2011] X. Zhao, E. Dellandréa, L. Chen and I.A. Kakadiaris. *Accurate landmarking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model.* Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on, vol. 41, no. 5, pages 1417–1428, 2011. 49, 111