Analyse locale de la forme 3D pour la reconnaissance d'expressions faciales

Ahmed Maalej, Boulbaba Ben Amor, Mohamed Daoudi

▶ **To cite this version:**

Ahmed Maalej, Boulbaba Ben Amor, Mohamed Daoudi. Analyse locale de la forme 3D pour la reconnaissance d'expressions faciales. ORASIS - Congrès des jeunes chercheurs en vision par ordinateur, INRIA Grenoble Rhône-Alpes, Jun 2011, Praz-sur-Arly, France. inria-00595490

HAL Id: inria-00595490

https://inria.hal.science/inria-00595490

Submitted on 24 May 2011

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

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

Analyse locale de la forme 3D pour la reconnaissance d'expressions faciales
Local 3D shape analysis for facial expression recognition

A. Maalej[1] B. Ben Amor[1,2] M. Daoudi[1,2]

[1] LIFL (UMR CNRS 8022), University of Lille1, France
[2] Institut TELECOM ; TELECOM Lille 1, France

Telecom Lille 1, Cité scientifique, Villeneuve d'Ascq, France
ahmed.maalej@telecom-lille1.eu

Résumé

Dans cet article, nous proposons une nouvelle approche pour la reconnaissance d'expressions faciales 3D invariante par rapport à l'identité. Cette approche est basée sur l'analyse de formes de « patches »locaux extraits à partir de modèles de visages 3D. Un cadre Riemannien est utilisé pour le calcul de distances géodésiques entre les patches correspondants appartenant a des visages différents sous différentes expressions. Des mesures quantitatives de similarité sont alors obtenues et sont utilisées comme des paramètres d'entrée pour des algorithmes de classification multiclasses. En utilisant des techniques de Multiboosting et de Machines à Vecteurs de Support (SVM), les taux de reconnaissance des six expressions de base obtenus sur la base BU-3DFE sont respectivement 98.81% et 97.75%.

Mots Clef

Reconnaissance d'expressions faciales 3D, analyse de forme, chemin géodésique, classification multiclasse.

Abstract

In this paper we propose a novel approach for indentity-independent 3D facial expression recognition. Our approach is based on shape analysis of local patches extracted from 3D facial shape model. A Riemannian framework is applied to compute geodesic distances between correspondent patches belonging to different faces of the BU-3DFE database and conveying different expressions. Quantitative measures of similarity are obtained and then used as inputs to several classification methods. Using Multiboosting and Support Vector Machines (SVM) classifiers, we achieved average recognition rates respectively equal to 98.81% and 97.75%.

Keywords

3D Facial expression recognition, shape analysis, geodesic path, multi-class classiffication.

1 Introduction

Les travaux de recherche en sciences cognitives ont démontré que les expressions faciales jouent un rôle prépondérant en termes de coordination et de conversation humaine. L'émission en permanence des signes émanant des différentes expressions faciales renseigne sur l'état émotionnel, l'intention, la personnalité et la psychologie d'une personne en question. Il s'agit d'un moyen de communication important et complexe qui détermine la nature de la relation que peut entreprendre une personne avec l'autrui. L'analyse automatique des expressions faciales constitue un outil important pour la recherche dans les domaines des sciences de l'étude du comportement et de la psychologie, ainsi que dans les domaines des interfaces homme-machine et de l'intelligence ambiante. Les progrès effectués dans des outils et techniques d'acquisition bidimensionnels (images 2D et vidéo) ont permis l'évolution des méthodes de détection, de suivi et de la reconnaissance de visage, et ont apporté une contribution significative à la reconnaissance d'expressions faciales sur des données 2D. Néanmoins, les performances des ces méthodes restent limitées et les techniques utilisées restent contraignantes dues aux variations de pose et d'illumination qui affectent les données brutes. Désormais, avec l'avènement de la 3D et le développement de caméras 3D, on dispose de données plus complètes et invariantes par rapport à la pose et à l'illumination. Actuellement les travaux de recherche se penchent sur l'analyse des images 3D de visages afin de proposer des approches de reconnaissance d'expressions plus performantes et invariantes aux variations dues à la pose et à l'illumination.

Les approches existantes qui s'adressent au problème de la reconnaissance d'expressions faciales peuvent être classées selon trois catégories : (1) *statiques* / *dynamiques* ; (2) *globales* / *locales* ; (3) *2D* / *3D*. La plus part de ces approches se basent sur l'extraction de descripteurs comme moyen pour décrire et représenter les expressions faciales. Des études sur les méthodes et approches utilisées ainsi que les comparaisons entre eux sont présentées dans diffé-

rents travaux comme ceux de Pantic et Rothkrantz [8] et de Samal et Iyengar [9]. Au milieu des années 70, l'analyse et la reconnaissance des expressions faciales se voyaient traiter des données statiques (images) ainsi que des données dynamiques (séquences vidéos). Whitehill et Omlin [14] ont proposé des approches de segmentation locales ainsi que globales. Leur étude est particulièrement basée sur la classification des unités d'action (AUs), ces AUs ont été définies dans le manuel de système de codage facial d'action (FACS) développé par Ekman et Friesen [3], qui présente une nomenclature et une décomposition les différentes expressions de visage en termes de 46 AUs. En ce qui concerne la reconnaissance des expressions faciales 3D, les premiers travaux sur cette problématique sont ceux de Wang et al. [4]. Ils ont proposé une nouvelle approche basée sur des caractéristiques et propriétés géométriques pour le calcul de descripteurs. Il s'agit d'une estimation d'une distribution en utilisant des primitives surfaciques. Un système d'étiquetage associé aux descripteurs obtenus a permis de former un ensemble d'attributs qui sont, ensuite, utilisés dans des processus d'apprentissage et de test de différent classifieurs. Le taux de reconnaissance obtenu est alors égale à 83%. Ils ont évalué leur approche non seulement sur des scans de visages en pose frontale de la base BU-3DFE (Binghamton University 3D facial expression database), mais aussi ils ont testé la robustesse de leur approche par rapport à la variation de pose. Un second travail a été proposé par Soyel et Demirel [11], leur méthode repose sur le calcul de six mesures de distances euclidienne caractéristiques entre onze points d'intérêt, choisis selon une étude d'optimisation basée sur la symétrie du visage. Ces distances sont ensuite utilisées comme des données d'entrée à une architecture de réseau de neurones dans le but de classer les expressions des visages selon sept classes (les six expressions de base et le neutre). Le taux de reconnaissance atteint est de 91.3%. Mpiperis et al. [7] ont proposé une approche de reconnaissance d'identité et d'expression simultanée en utilisant des modèles bilinéaires, le taux de reconnaissance moyen signalé est 90.5%. Tang et Huang [13] ont proposé une approche atomatique pour la sélection de descripteurs pertinents parmi la combinaison de toutes les distances euclidiennes normalisées calculées enre un ensemble de 83 points d'intérêt sur le visage. En utilisant l'algorithme Adaboost pour une multi-classification, ils ont raporté un taux de reconnaissance de 95.1%.

Les contributions majeurs de notre approche sont : (1) une nouvelle représentation de régions locales de visages 3D, (2) une application d'un cadre Riemannien en vue de l'analyse de forme via le calcul de mesures quantitative visant à évaluer la similarité entre des patches correspondants sur différents visages.

2 Approche proposée

Dans cet article nous proposons d'exploiter l'information tridimensionnelle du visage, et de fournir des outils mathématiques issus de la géométrie Riemannienne pour la reconnaissance des expressions faciales. En se basant sur les études en psychologie entreprises par Ekman [3], les expressions sont classées sous six catégories : la joie, la tristesse, la peur, le dégoût, la colère et la surprise. La figure 1 illustre des exemples de scans de visage 3D de la base BU-3DFE affichant les expressions de base.

En vue d'entreprendre une analyse d'expressions faciales

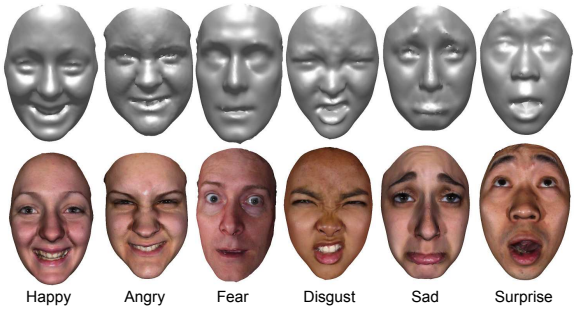

FIGURE 1 – Illustration des six différentes expressions faciales catégorisées dans la base BU-3DFE, première ligne : modèles de forme , seconde ligne : modèles texturés.

dans une image 3D, il convient d'en extraire des informations potentiellement pertinentes. Il existe différentes méthodes pour l'extraction de ces informations. Il y a les méthodes globales qui modélisent le visage dans son intégralité. Il y a aussi les méthodes reposant sur des traits locales des régions faciales, qui sont susceptibles, plus qu d'autres, de se déformer suite aux mouvements des muscles faciaux. Nous nous intéressons particulièrement aux méthodes de modélisation locale, et notre approche consiste à extraire des patches relatifs à des régions locales du visage.

2.1 Représentation par des patches locaux

Pour représenter l'ensemble des régions d'un visage donné, on procède par l'extraction de patches, ce qui revient à faire l'extraction d'un ensemble de courbes de niveaux indexés. L'ensemble des courbes extraites autour d'un point anatomique repéré sur la surface faciale constitue le patch. Pour ce faire, étant donné un point d'intérêt r_l annoté sur une surface faciale 3D qu'on dénote par S, on extrait un patch P_l centré en ce point. L'extraction des courbes formant P_l est faite moyennant la distance euclidienne qui sépare r_l de tout point p appartenant à la surface faciale S et définit par :

$$\|r_l - .\| : c_\lambda^l = \{p \in S \mid \|r_l - p\| = \lambda\} \subset S,\ \lambda \in [0, \lambda_0] \quad (1)$$

Chaque courbe c_λ^l est formée par un ensemble de points situés à distance λ égale du point r_l. Ces courbes tridimensionnelles permettent de nous révéler l'information de forme 3D qui fera l'objet de nos analyses. La figure 2 illustre l'étape d'extraction des patches basés sur des courbes de niveaux c_λ^l.

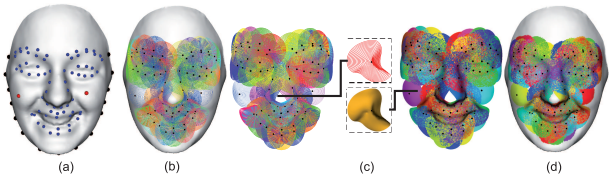

FIGURE 2 – (a) Modèle de visage 3D annoté par 70 points de référence ; (b) extraction des courbes de niveaux autour des points de référence ; (c) construction des patches basée sur la représentation par des courbes fermées ; (d) représentation des patches extraits sur la surface faciale.

2.2 Analyse de forme de courbes 3D

La mise au point d'algorithmes d'analyse de surfaces 3D nécessite une modélisation mathématique rigoureuse des objets d'étude (ici les patches de visage 3D). Le concept et les éléments introduits par la géométrie Riemannienne proposent des solutions optimales pour la quantification de la forme et la comparaison entre des surfaces 3D. Nous nous sommes basés sur ces outils pour développer une analyse de formes des patches formés à partir des courbes faciales. Une analyse de forme implique l'utilisation d'une métrique pour quantifier et calculer la déformation optimale entre deux objets 3D. Cette métrique se doit d'être invariante à certaines transformations (la translation, la rotation, l'échelle). Dans le cadre de nos travaux, nous disposons d'une structure mathématique et statistique rigoureusement développée pour l'étude de la géométrie différentielle sur les variétés. L'utilisation de cet outil nous permet de calculer la distance qui sépare deux courbes 3D dans l'espace des courbes. Il s'agit de la distance géodésique qui est définit comme étant la longueur du chemin optimal séparant deux courbes. Ainsi pour déterminer la géodésique entre les patches dans l'espace des surfaces on calcule la moyenne des distances géodésiques entre les courbes relatives. Une fois l'extraction des patches est faite, on procède par une analyse de forme et un calcul de mesure quantitative du score de similarité entre les courbes correspondantes sur différents scans et différentes expressions. Cette approche est motivée par le constat présenté par Ekman [3] qui montre que les personnes affichent sur leur visage les mêmes mimiques lorsqu'ils expriment l'une des émotions de base. Ce sont les mêmes régions et muscles du visage qui sont sollicités et qui subissent pratiquement le même processus de déformation dynamique. On peut alors en déduire que certains patches correspondants associés à une même expression se déforment d'une manière similaire, alors que les patches correspondants associés à des expressions différentes se déforment différemment. Dans la section suivante on présente une description de l'analyse de forme basée sur la représentation des objets par des courbes. Cette approche, initialement introduite par Joshi et al. [5], a montré ses performances dans le domaine de reconnaissance d'identité de visage 3D Samir et al. [10] et Drira et al. [2], ainsi que dans une application de reconnaissance d'expression faciale binaire Maalej et al. [6]. Pour l'analyse de la forme d'une courbe β dans $\mathbb{R}^3$, une paramétrisation de cette courbe est donnée par la fonction de *square-root velocity* (SRVF) définit comme suit ; soient un interval $I = [0, 1]$, $\beta : I \longrightarrow \mathbb{R}^3$ une courbe et $q : I \longrightarrow \mathbb{R}^3$ la fonction (SRVF) est données par :

$$q(t) \doteq \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}} . \qquad (2)$$

Où t est un paramètre dans $\in I$ et $\|.\|$ est la norme euclidienne dans $\mathbb{R}^3$. On note que $q(t)$ est une fonction spécifique qui caractérise la forme de β. la racine de la norme $\mathbb{L}^2$ de q est donnée par : $\|q\|^2 = \int_I < q(t), q(t) > dt = \int_I \|\dot{\beta}(t)\| dt$, qui est la la longueur de β. La norme $\mathbb{L}^2$ est commode pour l'analyse des courbes. Ainsi que la métrique élastique classique pour comparer la forme de courbes devient la métrique $\mathbb{L}^2$ dans la représentation de SRVF. Ce point est important car il permet de réduire le calcul de la métrique élastique au calcul de la fonction analytique sous la métrique $\mathbb{L}^2$. Dans le but d'une analyse spécifique de l'analyse de forme courbe fermée, on définit l'ensemble : $\mathcal{C} = \{q : I \longrightarrow \mathbb{R}^3 | \int_I q(t)\|q(t)\| dt = 0\} \subset \mathbb{L}^2(I, \mathbb{R}^3)$. Pour comparer deux courbes β_1 and β_2, nous utilisons $d_{\mathcal{C}}(q_1, q_2)$ pour dénoter la longueur du chemin chemin géodésique dans $\mathcal{C}$ entre β_1 et β_2, ou q_i $(i \in 1, 2)$ est la fonction *SRVF* de β_i $(i \in 1, 2)$. Pour plus de détails sur le cadre utilisé dans notre approche d'analyse de forme le lecteur peut se référer aux travaux de Srivastava et al. [12].

2.3 Analyse de forme de patches 3D

L'analyse de forme de courbes peut être étendue à une analyse de forme de patches. En se basant sur la représentation d'une surface faciale S par un nombre l de patches $P_k \leftrightarrow \{c^k_\lambda, \lambda \in [0, \lambda_0]\}$ avec $k \in [0..l]$ et c^k_λ est la courbe indexée associée à la fonction distance $\|r_l - .\| = \lambda$. Grâce à cette formulation, chaque patch est représenté comme un élément de l'espace $\mathcal{C}^{[0,\lambda_0]}$. Moyennant le cadre Riemannien, la forme de deux patches est évaluée à travers la comparaison des courbes correspondantes. Pour deux patches P_1 et P_2, et leurs courbes relatives et respectives $\{c^1_\lambda, \lambda \in [0, \lambda_0]\}$ et $\{c^2_\lambda, \lambda \in [0, \lambda_0]\}$, l'idée est de comparer les courbes c^1_λ et c^2_λ et de cumuler les distances qui séparent les éléments correspondants de ces ensembles. Formellement on dénote $d_{\mathcal{C}^{[0,\lambda_0]}}$ la distance définit par

$$d_{\mathcal{C}^{[0,\lambda_0]}}(P_1, P_2) = \int_0^L d_{\mathcal{C}}(c^1_\lambda, c^2_\lambda) d\lambda . \qquad (3)$$

Les figures 3 et 4 illustrent des exemples de chemins géodésiques reconstruits entre les patches correspondants et relatifs aux régions locales étudiées. La figure 3 met en évidence la similarité en termes de surface 3D entre les patches issus de visages partageant la même expression de part et d'autre du chemin. Cette similarité et plus visiblement apparente grâce à une cartographie de la cour-

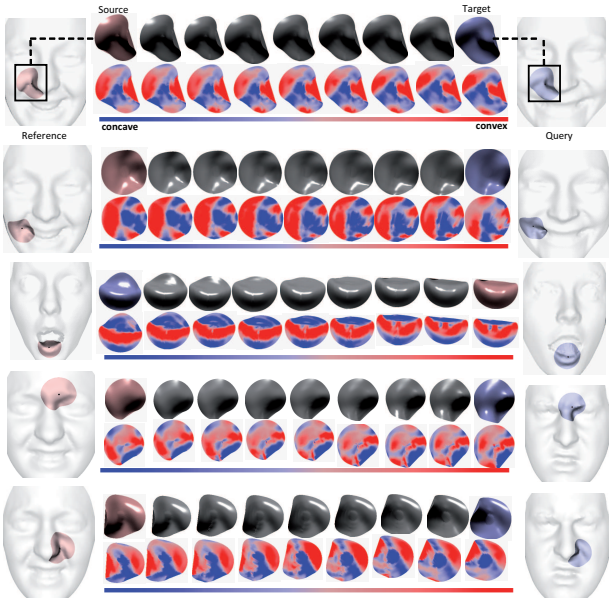

FIGURE 3 – Illustration de chemins géodésiques intra-classes entre les patches correspondants.

bure moyenne sur les surfaces qui forment le chemin géodésique. Cependant, on peut observer dans la figure 4 la différence entre les patches issus de visages éprouvant des expressions faciales différentes.

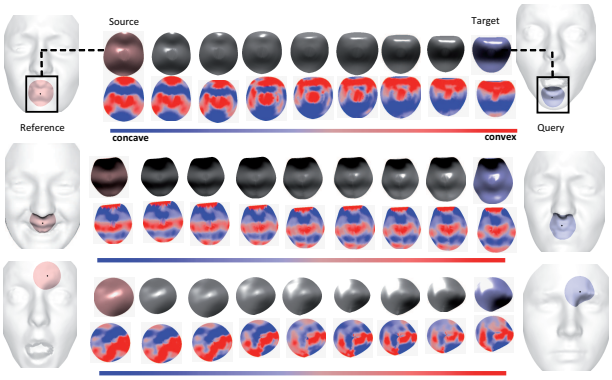

FIGURE 4 – Illustration de chemins géodésiques inter-classes entre les patches correspondants.

3 Expérimentations et résultats

Etant donné un visage candidat d'une personne j, des patches faciaux sont extraits comme décrit dans la sous-section 2.1. Le patch P_j^i est représenté par un ensemble de courbes indéxées $\{c_\lambda\}_j^i$ centrées sur un point d'intérêt i. D'une façon similaire, un patch P_{ref}^i relatif a un visage de référence *ref* est extrait au point i. La longueur du chemin géodésique entre le patch P_{ref}^i et son correspondant P_j^i est calculée à partir des distances géodésique des paires de courbes correspondantes et relatives aux deux patches et selon l'équation (3) introduite dans la sous-section (2.3).

Un vecteur caractéristique est alors formé pour chaque scan et dont la dimension est égale au nombre $N = 70$ de points d'intérêt considérés dans notre étude (68 points fournies avec la base statique BU-3DFE et 2 points ajoutés automatiquement sur les joues, qui sont les points milieux des distances géodésiques entre les coins extérieurs des yeux et les coins de la bouche). L'ensemble des vecteurs caractéristiques calculés sont étiquetés selon le nombre de classes d'expressions qui est égale à 6. Ces données sont ensuite utilisées en entrée aux algorithmes d'apprentissage automatiques tels que Multiboost avec les différents noyaux (analyse discriminante linéaire (LDA), Voisin le plus proche ou nearest neighbor (NN) et naive bayes (NB)) et SVM avec son noyau linéaire.
Dans un but de reconnaissance de l'expression d'une personne indépendamment de son identité, nos expériences ont été conduites sur une structure d'exploration de données (setup) de la base BU-3DFE. Cette structure est formée par des données acquises pour 60 sujets, répartit en 50% d'hommes et 50% de femmes. Ces données correspondent à des scans affichants des expressions d'intensité forte (03) et trés forte (04). Elles sont, ensuite, aléatoirement divisées en deux partie ; une première parite retenue pour un processus d'apprentissage et formée par 54 sujet (648 scans), et une deuxième partie formée par 6 sujets (72 scan) qui sert pour un processus de test. On note que les modèles de référence choisis ne sont pas utilisés ni dans la phase d'apprentissage ni dans celle du test. Les visages de référence sont illustrés dans la figure 5.
Plusieurs expériences ont été menées en changeant a chaque expérience le sujet de référence parmi les six choisis. En utilisant les classificateurs Multiboost-LDA, multiboost-NB, Multiboost-NN on obtient les taux moyen de reconnaissance : 98.81%, 98.76% et 98.07% respectivement. Avec le SVM linéaire on obtient un taux de reconnaissance de 97.75%. La table 1 résume les résultats obtenus moyennant les différents classifieurs, le meilleur taux est celui donné par le Multiboost-LDA. On note que ces résultats sont obtenus en appliquant la méthode de validation croisée d'ordre $K = 10$; c'est à dire que les données sont divisées en K groupes tirés aléatoirement qui vont successivement servir pour mener K expériences, où en chaque expérience $(K - 1)$ groupes sont utilisés pour l'apprentissage et le reste sert pour le test, le taux de reconnaissance est alors obtenu en calculant la moyenne des taux donné par chaque expérience.

Classifier	Multiboost-LDA	Multiboost-NB	Multiboost-NN	SVM-Linear
Recognition rate	**98.81%**	98.76%	98.07%	97.75%

TABLE 1 – Taux moyens de reconnaissance d'expressions faciales obtenus par les différents classifieurs utilisés.

Dans la table 2 on retrouve la matrice de confusion moyenne obtenu par l'algorithme de classification Multiboost, et donnant les meilleurs taux de reconnaissance pour chacune des six expressions de base. Les moindres taux

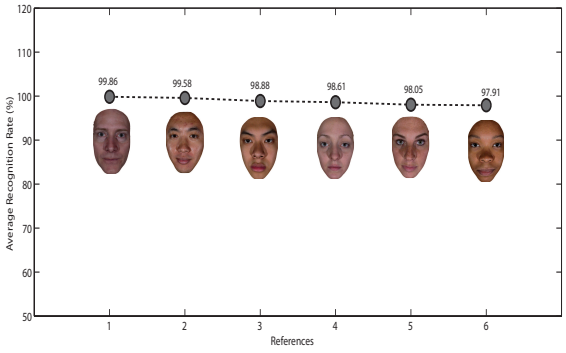

FIGURE 5 – Taux de reconnaissance d'expressions faciales pour chacun des visage de référence et en utilisant la méthode de classification Multiboost-LDA.

de recoannaissance touche les deux expressions de colère (AN) et de surprise (SU). Due à une certaine similrité entre l'expression AN et celle de dégoût (DI), en terme de régions et muscles faciaux sollicités suite à ces expressions, l'erreur de classification relative à l'expression AN, reconu comme étant l'expression DI, s'élève à un taux de 1.11%. Alors que pour l'expression (SU), l'information synthétique que couvre la zone intérieur de la bouche ouverte altère la forme des courbes constituants les patches dans cette région, ce qui donne, par conséquence, une mesure de similaritée alternée entre les patches correspondants relatifs a cette région.

%	AN	DI	FE	HA	SA	SU
AN	**97.92**	1.11	0.14	0.14	0.69	0.0
DI	0.56	**99.16**	0.14	0.0	0.14	0.0
FE	0.14	0.14	**99.72**	0.0	0.0	0.0
HA	0.56	0.14	0.0	**98.60**	0.56	0.14
SA	0.28	0.14	0.0	0.0	**99.30**	0.28
SU	0.14	0.56	0.0	0.0	1.11	**98.19**

TABLE 2 – La matrice de confusion obtenue par l'algorithme de classification Multiboost-LDA.

La table 3 présente un résultat comparatif des taux de reconnaissances des expressions faciales entre notre approche et des travaux existants de l'état de l'art. Nous pouvons noter que notre approche donne les meilleurs taux de reconnaissance en comparaison avec des travaux qui utilise pratiquement les même structure d'exploration de données, tels que les travaux de Tang et al. [13], Soyel et al. [11] qui ont validé leur approche moyennant une technique de validation croisé d'ordre $K = 10$, et Wang et al. [4] où les résultats sont obtenus par une validation croisée d'ordre $K = 20$.

Cross-validation	*Notre approche*	*Tang et al. [13]*	*Soyel et al. [11]*	*Wang et al. [4]*
10-fold	**98.81%**	95.1%	91.3%	-
20-fold	**92.75%**	-	-	83.6%

TABLE 3 – Comparaison des résultats de notre approche par rapport au travaux de l'état de l'art [13], [11] and [4].

Le principe de l'algorithme Multiboost est similaire à celui d'Adaboost et permet d'attribuer des poids aux classifieurs faibles (ici les patches), en fonction des valeurs de ces poids nous pouvons, alors, en déduire les régions du visage qui sont discriminantes et jouent un rôle plus important que le restant des patches pour la classification d'une expression donnée.

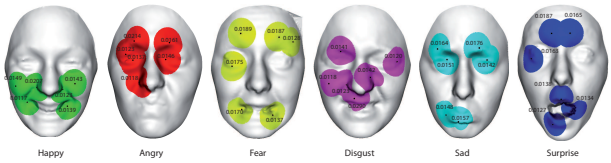

FIGURE 6 – Illustration des patches pertinents, pour chacune des six expressions de base, et auxquels sont affectés les poids les plus importants données par les premières itérations de l'algorithme de classification Multiboost.

3.1 Robustesse par rapport à la pose non frontale

Dans des situations pratiques, la pose frontale des visages n'est pas toujours disponible. Par conséquent, la reconnaissance d'expressions faciale de visages non frontaux présente encore un défi. Nous nous intéressons à l'évaluation de la robustesse de notre approche de reconnaissance d'expressions dans des conditions de variations de pose. En appliquant une rotation du modèle de visage 3D, on peut générer des scans sous différents angles de vues. Dans notre étude les angles choisis correspondent aux valeurs suivantes : $15\,°$, $30\,°$, $45\,°$, $60\,°$, $75\,°$ et $90\,°$. On suppose que l'information de forme est indisponible pour les parties cachées du visage due à une auto-occultation. Pour chaque angle de vue traité, on garde, de l'ensemble des patches représentants le visage, que ceux qui sont apparents et visibles. Dans le cas où un landmark est occulté, ou il est visible mais une partie du patch tout autour est occultée, ces patches sont considérés comme données manquantes. Dans ce cas de figure la matrice calculée, et dont les éléments sont les distances géodésique entre patches correspondants, se trouvera avec des éléments manquants. Pour traiter ce problème de données manquantes, nous utilisons une technique dite d'« imputation »basée sur le calcul de moyenne [1]. Dans nos expériences, la méthode d'imputation via le calcul de moyennes est utilisée, et elle consiste à remplacer ces données manqantes par des moyennes qui ont été déjà calculées sur des valeurs obtenues dans un processus d'apprentissage dans le cadre du scénario de la pose frontale.
Soit $d_{ijk} = d_{\mathcal{C}^{[0,\lambda_0]}}(P_i^k, P_j^k)$ la distance géodésique entre le k ième patch appartenant au sujets i et j $(i \neq j)$. Dans le cadre de la pose frontale, l'ensemble des instances X_i^{fv} relatif au sujet i est étiqueté, et est donné par :

$$\mathbf{X_i^{fv}} = \begin{pmatrix} d_{i11} & \dots & d_{i1k} & \dots & d_{i1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{ij1} & \dots & d_{ijk} & \vdots & d_{ijN} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{iJ1} & \dots & d_{iJk} & \dots & d_{iJN} \end{pmatrix}$$

Où N est le nombre des attributs. Dans le cas de la pose non-frontale (nfv), si un attribut k est manquant, le k^{ime} vecteur colonne de la matrice X_i^{nfv} est remplacé par la moyenne des distances géodésiques calculées dans le cas de pose frontale, la matrice est alors donnée par :

$$\mathbf{X_i^{nfv}} = \begin{pmatrix} d_{i11} & \dots & m_k^{fv} & \dots & d_{i1N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{ij1} & \dots & m_k^{fv} & \vdots & d_{ijN} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{iJ1} & \dots & m_k^{fv} & \dots & d_{iJN} \end{pmatrix}$$

Afin d'évaluer la robustesse de notre approche dans un contexte de pose non-frontale, on procède par une reconnaissance de l'expression indépendamment de l'angle de vue. Des taux d'erreur sont alors calculés à travers des expériences sur des visages de test et en appliquant des classifieurs entrainé sur des scans frontaux. La figure 7 montre des taux d'erreur moyens donnés par chaque méthode de classification utilisée. Le meilleur résultat noté est celui donné par la méthode Mutliboost-LDA. On peut noter que les taux d'erreur de reconnaissance des expressions croient faiblement pour des angles de vues variant de 0 jusqu'à 45 degré, alors que au delà, les taux d'erreur augmentent significativement avec l'augmentation de l'angle de rotation.

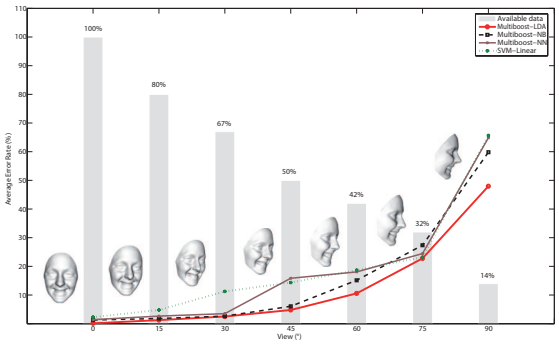

FIGURE 7 – Taux moyen d'erreur de reconnaissance des six expressions en fonction de la pose et utilisant les différents classifieurs.

4 Conclusion

Dans cet article nous présentons une nouvelle approche pour la reconnaissance d'expressions faciales 3D invariante par rapport à l'identité. L'idée proposée est la caractérisation des déformations locales que subit différentes régions de la surface faciale. Cette caractérisation est basée sur la représentation des différentes régions sur le visage par un ensemble de courbes fermées indexées ou patches. Un cadre Riemannien est utilisé pour une analyse de forme et le calcule des distances géodésiques entre les patches correspondants appartenant aux différents visages et sous différentes expressions. Ces distances sont, ensuite, étiquetées pour former un ensemble d'attributs en vue de construire les données d'entrée aux différents classifieurs. L'utilisation de Multiboost nous donne le meilleur taux de reconnaissance moyen qui est de l'ordre de 98.81%. La principale limitation de notre approche reste dans l'utilisation de points d'intérêt qui ont été annotés manuellement et fournis avecla base BU-3DFE statique.
Nous nous intéressons, actuellement, à la détection automatique d'un certains nombre de points d'intérêts, ce qui permettra de définir les principales régions anatomiques d'un visage 3D, et d'assurer ainsi une automatisation notre approche de reconnaissance d'expressions faciales.

Références

[1] G. Batista and M. C. Monard. An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence*, 17 :519–533, 2003.

[2] Hassen Drira, Boulbaba Ben Amor, Mohamed Daoudi, and Anuj Srivastava. Elastic radial curves to model 3d facial deformations. In *Proceedings of the ACM workshop on 3D object retrieval*, 3DOR '10, pages 75–80, New York, NY, USA, 2010. ACM.

[3] P. Ekman and W. Friesen. *Facial Action Coding System : A Technique for the Measurement of Facial Movement*, 1978. Consulting Psychologists Press.

[4] X. Wei J. Wang, L. Yin and Yi. Sun. 3d facial expression recognition based on primitive surface feature distribution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1399–1406, 2006.

[5] S. Joshi, E. Klassen, A. Srivastava, and I. H. Jermyn. A novel representation for riemannian analysis of elastic curves in $\mathbb{R}^n$. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, USA, June 2007.

[6] Ahmed Maalej, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava, and Stefano Berretti. Local 3d shape analysis for facial expression recognition. In *Proc. of International Conference on Pattern Recognition ICPR*, Istanbul, Turkey, 2010.

[7] I. Mpiperis, S. Malassiotis, and M. G. Strintzis. Bilinear models for 3d face and facial expression recognition. *IEEE Transactions on Information Forensics and Security*, 3(3) :498–511, 2008.

[8] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions : The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12) :1424–1445, December 2000.

[9] A. Samal and Prasana A. Iyengar. Automatic recognition and analysis of human faces and facial expressions : a survey. *Pattern Recogn.*, 25(1) :65–77, 1992.

[10] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen. An intrinsic framework for analysis of facial surfaces. *International Journal of Computer Vision*, 82(1) :80–95, 2009.

[11] H. Soyel and H. Demirel. Facial expression recognition using 3d facial feature distances. *International Conference on Image Analysis and Recognition (ICIAR)*, pages 831–838, 2007.

[12] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Analysis and Machine Intelligence* . accepted for publication.

[13] H. Tang and T.S. Huang. 3d facial expression recognition based on automatically selected features. *In First IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB)*, pages 1–8, 2008.

[14] J. Whitehill and C. W. Omlin. Local versus global segmentation for facial expression recognition. In *FGR '06 : Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pages 357–362, Washington, DC, USA, 2006. IEEE Computer Society.