Pattern Recognition 44 (2011) 1581–1589

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Shape analysis of local facial patches for 3D facial expression recognition

Ahmed Maalej [a,b], Boulbaba Ben Amor [a,b,*], Mohamed Daoudi [a,b], Anuj Srivastava [c], Stefano Berretti [d]

[a] LIFL (UMR CNRS 8022), University of Lille 1, France
[b] Institut TELECOM; TELECOM Lille 1, France
[c] Department of Statistics, Florida State University, USA
[d] Dipartimento di Sistemi e Informatica, University of Firenze, Italy

ARTICLE INFO

Article history:
Received 6 July 2010
Received in revised form
10 February 2011
Accepted 14 February 2011
Available online 19 February 2011

Keywords:
3D facial expression classification
Shape analysis
Geodesic path
Multiboosting
SVM

ABSTRACT

In this paper we address the problem of 3D facial expression recognition. We propose a local geometric shape analysis of facial surfaces coupled with machine learning techniques for expression classification. A computation of the length of the geodesic path between corresponding patches, using a Riemannian framework, in a shape space provides a quantitative information about their similarities. These measures are then used as inputs to several classification methods. The experimental results demonstrate the effectiveness of the proposed approach. Using multiboosting and support vector machines (SVM) classifiers, we achieved 98.81% and 97.75% recognition average rates, respectively, for recognition of the six prototypical facial expressions on BU-3DFE database. A comparative study using the same experimental setting shows that the suggested approach outperforms previous work.

© 2011 Elsevier Ltd. All rights reserved.

结合　　面部几何形状分析

1. Introduction

In recent years, 3D facial expression recognition has received growing attention. It has become an active research topic in computer vision and pattern recognition community, impacting important applications in fields related to human–machine interaction (e.g., interactive computer games) and psychological research. Increasing attention has been given to 3D acquisition systems due to the natural fascination induced by 3D objects visualization and rendering. In addition 3D data have advantages over the 2D data, in that 3D facial data have high resolution and convey valuable information that overcomes the problem of pose/lighting variations and the detail concealment of low resolution acquisition. 获取低解析度问题的隐藏细节。

In this paper we present a novel approach for 3D identity-independent facial expression recognition based on a local shape analysis. Unlike the identity recognition task that has been the subject of many papers, only few works have addressed 3D facial expression recognition. This could be explained through the challenge imposed by the demanding security and surveillance requirements. Besides, there has long been a shortage of publicly available 3D facial expression databases that serve the researchers exploring 3D information to understand human behaviors and emotions. The main task is to classify the facial expression of a given 3D model, into one of the six prototypical expressions, namely *Happiness*, *Anger*, *Fear*, *Disgust*, *Sadness* and *Surprise*. It is stated that these expressions are universal among human ethnicity as described in [1,2].

The remainder of this paper is organized as follows. First, a brief overview of related work is presented in Section 2. In Section 3 we describe the BU-3DFE database designed to explore 3D information and improve facial expression recognition. In Section 4, we summarize the shape analysis framework applied earlier for 3D curves matching by Joshi et al. [3], and discuss its use to perform 3D patches analysis. This framework is further expounded in Section 5, so as to define methods for shapes analysis and matching. In Section 6 a description of the feature vector and used classifiers is given. In Section 7, experiments and results of our approach are reported, and the average recognition rate over 97% is achieved using machine-learning algorithms for the recognition of facial expressions such as multiboosting and SVM. Finally, discussion and conclusion are given in Section 8.

身份识别

2. Related work

安全需要及监测要求

Facial expression recognition has been extensively studied over the past decades especially in 2D domain (e.g., images and videos) resulting in a valuable enhancement. Existing approaches that address facial expression recognition can be divided into three categories: (1) *static* versus *dynamic*; (2) *global* versus *local*; (3) *2D* versus *3D*. Most of the approaches are based on feature

* Corresponding author at: LIFL (UMR CNRS 8022), University of Lille 1, France.
E-mail address: boulbaba.benamor@telecom-lille1.eu (B.B. Amor).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.02.012

extraction/detection as a mean to represent and understand facial expressions. Pantic and Rothkrantz [4] and Samal and Iyengar [5] presented a survey where they explored and compared different approaches that were proposed, since the mid 1970s,for facial expression analysis from either static facial images or image sequences. Whitehill and Omlin [6] investigated on the Local versus Global segmentation for facial expression recognition. In particular, their study is based on the classification of action units (AUs), defined in the well-known Facial Action Coding System (FACS) manual by Ekman and Friesen [7], and designating the elementary muscle movements involved in the bio-mechanical of facial expressions. They reported, in their study on face images, that the local expression analysis showed no consistent improvement in recognition accuracy compared to the global analysis. As for 3D facial expression recognition, the first work related to this issue was presented by Wang et al. [8]. They proposed a novel geometric feature based facial expression descriptor, derived from an estimation of primitive surface feature distribution.

A labeling scheme was associated with their extracted features, and they constructed samples that have been used to train and test several classifiers. They reported that the highest average recognition rate they obtained was 83%. They evaluated their approach not only on frontal view facial expressions of the BU-3DFE database, but they also tested its robustness to non-frontal views. A second work was reported by Soyel and Demirel [9] on the same database. They extracted six characteristic distances between 11 facial landmarks, using Neural Network architecture that analysis the calculated distances, they classified the BU-3DFE facial scans into seven facial expressions including neutral expression. The average recognition rate they achieved was 91.3%. Mpiperis et al. [10] proposed a joint 3D face and facial expression recognition using bilinear model. They fitted both formulations, using symmetric and asymmetric bilinear models to encode both identity and expression. They reported an average recognition rate of 90.5%. They also reported that the facial expressions of disgust and surprise were well identified with an accuracy of 100%. Tang and Huang [11] proposed an automatic feature selection computed from the normalized Euclidean distances between two picked landmarks from 83 possible ones. Using regularized multi-class Adaboost classification algorithm, they reported an average recognition rate of 95.1%, and they mentioned that the surprise expression was recognized with an accuracy of 99.2%.

In this paper, we further investigate the problem of 3D identity-independent facial expression recognition. The main contributions of our approach are the following: (1) We propose a new process for representing and extracting patches on the facial surface scan that cover multiple regions of the face and (2) we apply a framework to derive 3D shape analysis to quantify similarity measure between corresponding patches on different 3D facial scans. Thus, we combine a local geometric-based shape analysis approach of 3D faces and several machine-learning techniques to perform such classification.

导出一个3D形状分析去量化不同面部表情的相应块的相似性测量

3. Database description

BU-3DFE is one of the very few publicly available databases of annotated 3D facial expressions, collected by Yin et al. [12] at Binghamton University. It was designed for research on 3D human face and facial expression and to develop a general understanding of the human behavior. Thus the BU-3DFE database is beneficial for several fields and applications dealing with human computer interaction, security, communication, psychology, etc. There are a total of 100 subjects in the database, 56 females and 44 males. A neutral scan was captured for each subject, then they were asked to perform six expressions namely: Happiness (HA), Anger (AN), Fear (FE), Disgust (DI), Sad (SA) and Surprise (SU). The expressions vary according to four levels of intensity (low, middle, high and highest or 01–04). Thus, there are 25 3D facial expression models per subject in the database. A set of 83 manually annotated facial landmarks is associated to each model. These landmarks are used to define the regions of the face that undergo specific deformations due to single muscles movements when conveying facial expression [7]. In Fig. 1, we illustrate examples of the six universal facial expressions 3D models including the highest intensity level.

4. 3D facial patches-based representation

形状描述符

Most of the earlier work in 3D shape analysis use shape descriptors such as curvature, crest lines, shape index (e.g., ridge, saddle, rut, dome, etc.). These descriptors are defined based on the geometric and topological properties of the 3D object, and are used as features to simplify the representation and thus the

曲率、峰线、形状索引（峰线、马鞍状、凹凸以及圆屋顶状）

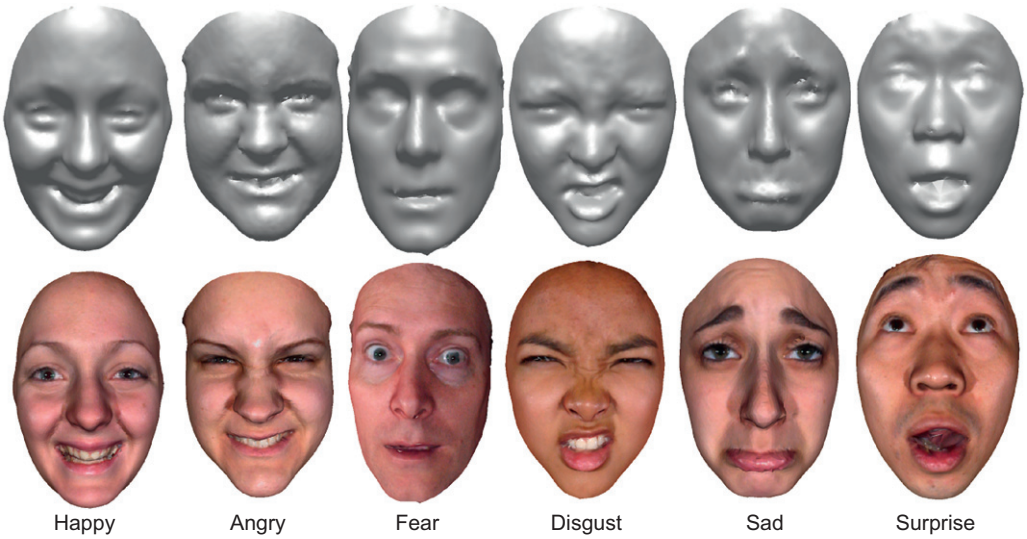

Fig. 1. Examples of 3D facial expression models (first row 3D shape models, second row 3D textured models) of the BU-3DFE database.

尽管他们严格的定义

comparison for 3D shape matching and recognition tasks. Despite their rigorous definition, such features are computed based on numerical approximation that involves second derivatives and can be sensitive to noisy data. In case of 3D facial range models, the facial surface labeling is a critical step to describe the facial behavior or expression, and a robust facial surface representation is needed. In Samir et al. [13] the authors proposed to represent facial surfaces by an indexed collections of 3D closed curves on faces. These curves are level curves of a surface distance function (i.e., geodesic distance) defined to be the length of the shortest path between a fixed reference point (taken to be the nose tip) and a point of the extracted curve along the facial surface. This being motivated by the robustness of the geodesic distance to facial expressions and rigid motions. Using this approach they were able to compare 3D shapes by comparing facial curves rather than comparing corresponding shape descriptors.

数值逼近

二阶导数

二次导数

In our work we intend to further investigate on local shapes of the facial surface. We are especially interested in capturing deformations of local facial regions caused by facial expressions. Using a different solution, we compute curves using the Euclidean distance which is sensitive to deformations and thus can better capture differences related to variant expressions. To this end, we choose to consider N reference points (landmarks) $\{r_l\}_{1 \le l \le N}$ (Fig. 2(a)) and associated sets of level curves $\{c_\lambda^l\}_{1 \le \lambda \le \lambda_0}$ (Fig. 2(b)). These curves are extracted over the patches centered at these points. Here λ stands for the value of the distance function between the reference point r_l and the point belonging to the curve c_λ^l, and λ_0 stands for the maximum value taken by λ. Accompanying each facial model there are 83 manually picked landmarks, these landmarks are practically similar to the MPEG-4 feature points and are selected based on the facial anatomy structure. Given these points the feature region on the face can be easily determined and extracted. We were interested in a subset of 68 landmarks laying within the face area, discarding those marked on the face border. Contrary to the MPEG-4 feature points specification that annotates the cheeks center and bone, in BU-3DFE there were no landmarks associated with the cheek regions. Thus, we add two extra landmarks at both cheeks, obtained by extracting the middle point along the geodesic path between the mouth corner and the outside eye corner.

We propose to represent each facial scan by a number of patches centered on the considered points. Let r_l be the reference point and P_l a given patch centered on this point and localized on the facial surface denoted by S. Each patch will be represented by an indexed collection of level curves. To extract these curves, we use the Euclidean distance function $\|r_l - p\|$ to characterize the length between r_l and any point p on S. Using this function we defined the curves as level sets of

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

Each c_λ^l is a closed curve, consisting of a collection of points situated at an equal distance λ from r_l. Fig. 2 resumes the scheme of patches extraction.

3d闭合曲线

5. Framework for 3D shape analysis

Once the patches are extracted, we aim at studying their shape and design a similarity measure between corresponding ones on different scans under different expressions. This is motivated by the common belief that people smile, or convey any other expression, the same way, or more appropriately certain regions taking part in a specific expression undergo practically the same dynamical deformation process. We expect that certain corresponding patches associated with the same given expression will be deformed in a similar way, while those associated with two different expressions will deform differently. The following sections describe the shape analysis of closed curves in $\mathbb{R}^3$, initially introduced by Joshi et al. [3], and its extension to analyze shape of local patches on facial surfaces.

5.1. 3D curve shape analysis

We start by considering a closed curve β in $\mathbb{R}^3$. While there are several ways to analyze shapes of closed curves, an elastic analysis of the parametrized curves is particularly appropriate in 3D curves analysis. This is because (1) such analysis uses a square-root velocity function representation which allows us to compare local facial shapes in presence of elastic deformations, (2) this method uses a square-root representation under which the elastic metric reduces to the standard $\mathbb{L}^2$ metric and thus simplifies the analysis, (3) under this metric the Riemannian distance between curves is invariant to the re-parametrization. To analyze the shape of β, we shall represent it mathematically using a square-root representation of β as follows; for an interval $I=[0,1]$, let $\beta : I \longrightarrow \mathbb{R}^3$ be a curve and define $q : I \longrightarrow \mathbb{R}^3$ to be its

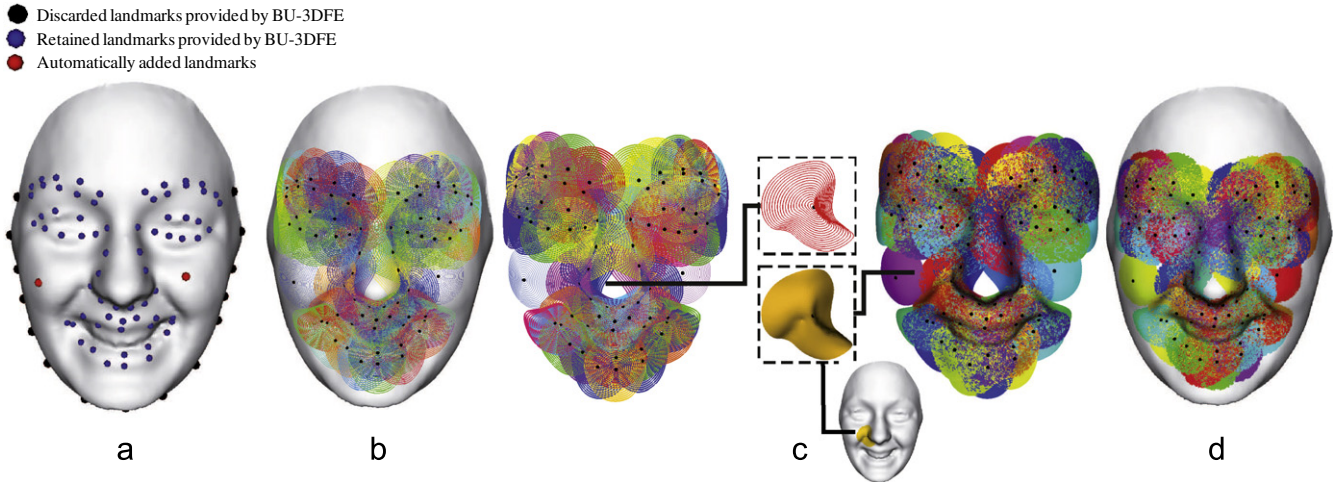

Fig. 2. (a) 3D annotated facial shape model (70 landmarks); (b) 3D closed curves extracted around the landmarks; (c) 3D curve-based patches composed of 20 level curves with a size fixed by a radius λ=20 mm; (d) extracted patches on the face.

square-root velocity function (SRVF), given by:

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

Here t is a parameter $\in I$ and $\|.\|$ is the Euclidean norm in $\mathbb{R}^3$. We note that $q(t)$ is a special function that captures the shape of β and is particularly convenient for shape analysis, as we describe next. The classical elastic metric for comparing shapes of curves becomes the $\mathbb{L}^2$-metric under the SRVF representation [14]. This point is very important as it simplifies the calculus of elastic metric to the well-known calculus of functional analysis under the $\mathbb{L}^2$-metric. Also, the squared $\mathbb{L}^2$-norm of q, given by: $\|q\|^2 = \int_{\mathbb{S}^1} < q(t), q(t) > dt = \int_{\mathbb{S}^1} \|\dot{\beta}(t)\|\, dt$, which is the length of β. In order to restrict our shape analysis to closed curves, we define the set: $\mathcal{C} = \{q : \mathbb{S}^1 \longrightarrow \mathbb{R}^3 | \int_{\mathbb{S}^1} q(t)\|q(t)\| dt = 0\} \subset \mathbb{L}^2(\mathbb{S}^1, \mathbb{R}^3)$. Notice that the elements of $\mathcal{C}$ are allowed to have different lengths. Due to a non-linear (closure) constraint on its elements, $\mathcal{C}$ is a non-linear manifold. We can make it a Riemannian manifold by using the metric: for any $u, v \in T_q(\mathcal{C})$, we define:

$$< u, v > = \int_{\mathbb{S}^1} < u(t), v(t) > dt. \quad (3)$$

So far we have described a set of closed curves and have endowed it with a Riemannian structure. Next we consider the issue of representing the *shapes* of these curves. It is easy to see that several elements of $\mathcal{C}$ can represent curves with the same shape. For example, if we rotate a curve in $\mathbb{R}^3$, we get a different SRVF but its shape remains unchanged. Another similar situation arises when a curve is re-parametrized; a re-parameterization changes the SRVF of curve but not its shape. In order to handle this variability, we define orbits of the rotation group $SO(3)$ and the re-parameterization group Γ as the equivalence classes in $\mathcal{C}$. Here, Γ is the set of all orientation-preserving diffeomorphisms of $\mathbb{S}^1$ (to itself) and the elements of Γ are viewed as re-parameterization functions. For example, for a curve $\beta : \mathbb{S}^1 \to \mathbb{R}^3$ and a function $\gamma : \mathbb{S}^1 \to \mathbb{S}^1$, $\gamma \in \Gamma$, the curve $\beta \circ \gamma$ is a re-parameterization of β. The corresponding SRVF changes according to $q(t) \mapsto \sqrt{\dot{\gamma}(t)}\, q(\gamma(t))$. We set the elements of the orbit:

$$[q] = \left\{ \sqrt{\dot{\gamma}(t)} O q(\gamma(t)) | O \in SO(3), \gamma \in \Gamma \right\}, \quad (4)$$

to be equivalent from the perspective of shape analysis. The set of such equivalence classes, denoted by $\mathcal{S} \doteq \mathcal{C}/(SO(3) \times \Gamma)$ is called the *shape space* of closed curves in $\mathbb{R}^3$. $\mathcal{S}$ inherits a Riemannian metric from the larger space $\mathcal{C}$ due to the quotient structure.

The main ingredient in comparing and analysing shapes of curves is the construction of a geodesic between any two elements of $\mathcal{S}$, under the Riemannian metric given in Eq. (3). Given any two curves β_1 and β_2, represented by their SRVFs q_1 and q_2, we want to compute a geodesic path between the orbits $[q_1]$ and $[q_2]$ in the shape space $\mathcal{S}$. This task is accomplished using a *path-straightening approach* which was introduced in [15]. The basic idea here is to connect the two points $[q_1]$ and $[q_2]$ by an arbitrary initial path α and to iteratively update this path using the negative gradient of an energy function $E[\alpha] = \frac{1}{2}\int_s < \dot{\alpha}(s), \dot{\alpha}(s) > ds$. The interesting part is that the gradient of E has been derived analytically and can be used directly for updating α. As shown in [15], the critical points of E are actually geodesic paths in $\mathcal{S}$. Thus, this gradient-based update leads to a critical point of E which, in turn, is a geodesic path between the given points. In the remainder of the paper, we will use the notation $d_{\mathcal{S}}(\beta_1, \beta_2)$ to denote the length of the geodesic in the *shape space* $\mathcal{S}$ between the orbits q_1 and q_2, to reduce the notation.

5.2. 3D patches shape analysis

Now, we extend ideas developed in the previous section from analyzing shapes of curves to the shapes of patches. As mentioned earlier, we are going to represent a number of l patches of a facial surface S with an indexed collection of the level curves of the $\|r_l - .\|$ function (Euclidean distance from the reference point r_l). That is, $P_l \leftrightarrow \{c_\lambda^l, \lambda \in [0, \lambda_0]\}$, where c_λ^l is the level set associated with $\|r_l - .\| = \lambda$. Through this relation, each patch has been represented as an element of the set $\mathcal{S}^{[0,\lambda_0]}$. In our framework, the shapes of any two patches are compared by comparing their corresponding level curves. Given any two patches P_1 and P_2, and their level curves $\{c_\lambda^1, \lambda \in [0, \lambda_0]\}$ and $\{c_\lambda^2, \lambda \in [0, \lambda_0]\}$, respectively, our idea is to compare the patches curves c_λ^1 and c_λ^2, and to accumulate these differences over all λ. More formally, we define a distance $d_{\mathcal{S}^{[0,\lambda_0]}}$ given by:

$$d_{\mathcal{S}^{[0,\lambda_0]}}(P_1, P_2) = \int_0^L d_{\mathcal{S}}(c_\lambda^1, c_\lambda^2)\, d\lambda. \quad (5)$$

In addition to the distance $d_{\mathcal{S}^{[0,\lambda_0]}}(P_1, P_2)$, which is useful in biometry and other classification experiments, we also have a geodesic path in $\mathcal{S}^{[0,\lambda_0]}$ between the two points represented by P_1 and P_2. This geodesic corresponds to the optimal elastic deformations of facial curves and, thus, facial surfaces from one to another. Fig. 3 shows some examples of geodesic paths that are computed between corresponding patches associated with shape models sharing the same expression, and termed *intra-class geodesics*. In the first column we illustrate the source, which represents scan models of the same subject, but under different expressions. The third column represents the targets as scan models of different subjects. As for the middle column, it shows the geodesic paths. In each row we have both the shape and the mean curvature mapping representations of the patches along the geodesic path from the source to the target. The mean curvature representation is added to identify concave/convex areas on the source and target patches and equally spaced steps of geodesics. This figure shows that certain patches, belonging to the same class of expression, are deformed in a similar way. In contrast, Fig. 4 shows geodesic paths between patches of different facial expressions. These geodesics are termed *inter-class geodesics*. Unlike the intra-class geodesics shown in Fig. 3, these patches deform in a different way.

6. Feature vector generation for classification

In order to classify expressions, we build a feature vector for each facial scan. Given a candidate facial scan of a person j, facial patches are extracted around facial landmarks. For a facial patch P_j^i, a set of level curves $\{c_\lambda\}_j^i$ are extracted centered on the ith landmark. Similarly, a patch P_{ref}^i is extracted in correspondence to landmarks of a reference scans *ref*. The length of the geodesic path between each level curve and its corresponding curve on the reference scan is computed using a Riemannian framework for shape analysis of 3D curves (see Sections 5.1 and 5.2). The shortest path between two patches at landmark i, one in a candidate scan and the other in the reference scan, is defined as the sum of the distances between all pairs of corresponding curves in the two patches as indicated in Eq. (5). The feature vector is then formed by the distances computed on all the patches and its dimension is equal to the number of used landmarks $N = 70$ (i.e., 68 landmarks are used out of the 83 provided by BU-3DFED and the two additional cheek points). The ith element of this vector represents the length of the geodesic path that separates the relative patch to the corresponding one on the reference face scan. All feature vectors computed on the overall dataset will be labeled and used as input data to

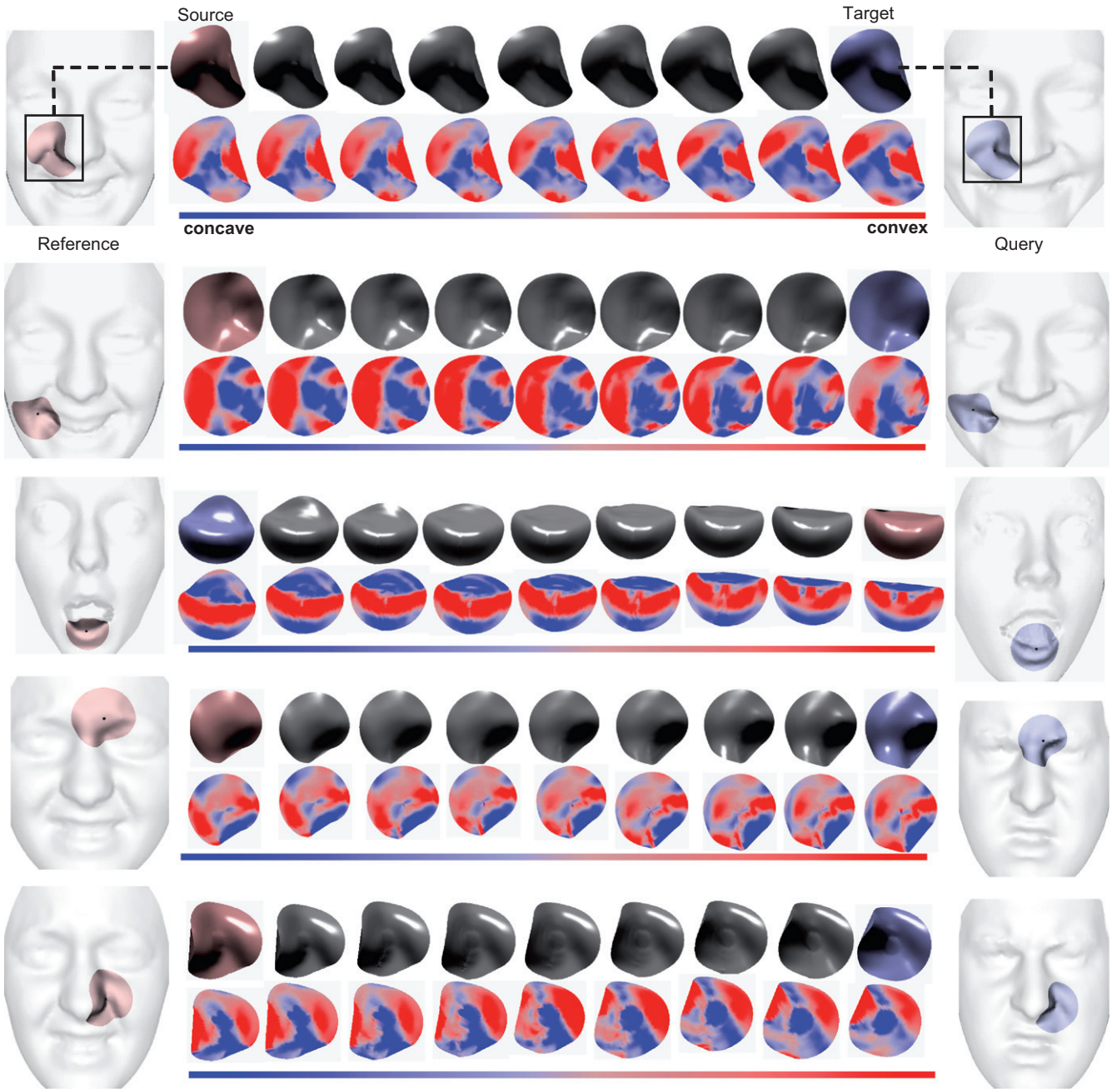

Fig. 3. Examples of intra-class (same expression) geodesic paths with shape and mean curvature mapping between corresponding patches.

machine-learning algorithms such as multiboosting and SVM, where multiboosting is an extension of the successful Adaboost technique for forming decision committees.

7. Recognition experiments

To investigate facial expression recognition, we have applied our proposed approach on a dataset that is appropriate for this task. In this section, we describe the experiments, obtained results and comparisons with related work.

7.1. Experimental setting

For the goal of performing identity-independent facial expression recognition, the experiments were conducted on the BU-3DFE static database. A dataset captured from 60 subjects were used, half (30) of them were female and the other half (30) were male, corresponding to the high and highest intensity levels 3D expressive models (03–04). These data are assumed to be scaled to the true physical dimensions of the captured human faces. Following a similar setup as in [16], we randomly divided the 60 subjects into two sets, the training set containing 54 subjects (648 samples), and the test set containing six subjects (72 samples).

To drive the classification experiments, we arbitrarily choose a set of six reference subjects with its six basic facial expressions. We point out that the selected reference scans do not appear neither in the training nor in the testing set. These references, shown in Fig. 5, with their relative expressive scans corresponding to the highest intensity level, are taken to play the role of representative models for each of the six classes of expressions. For each reference subject, we derive a facial expression recognition experience.

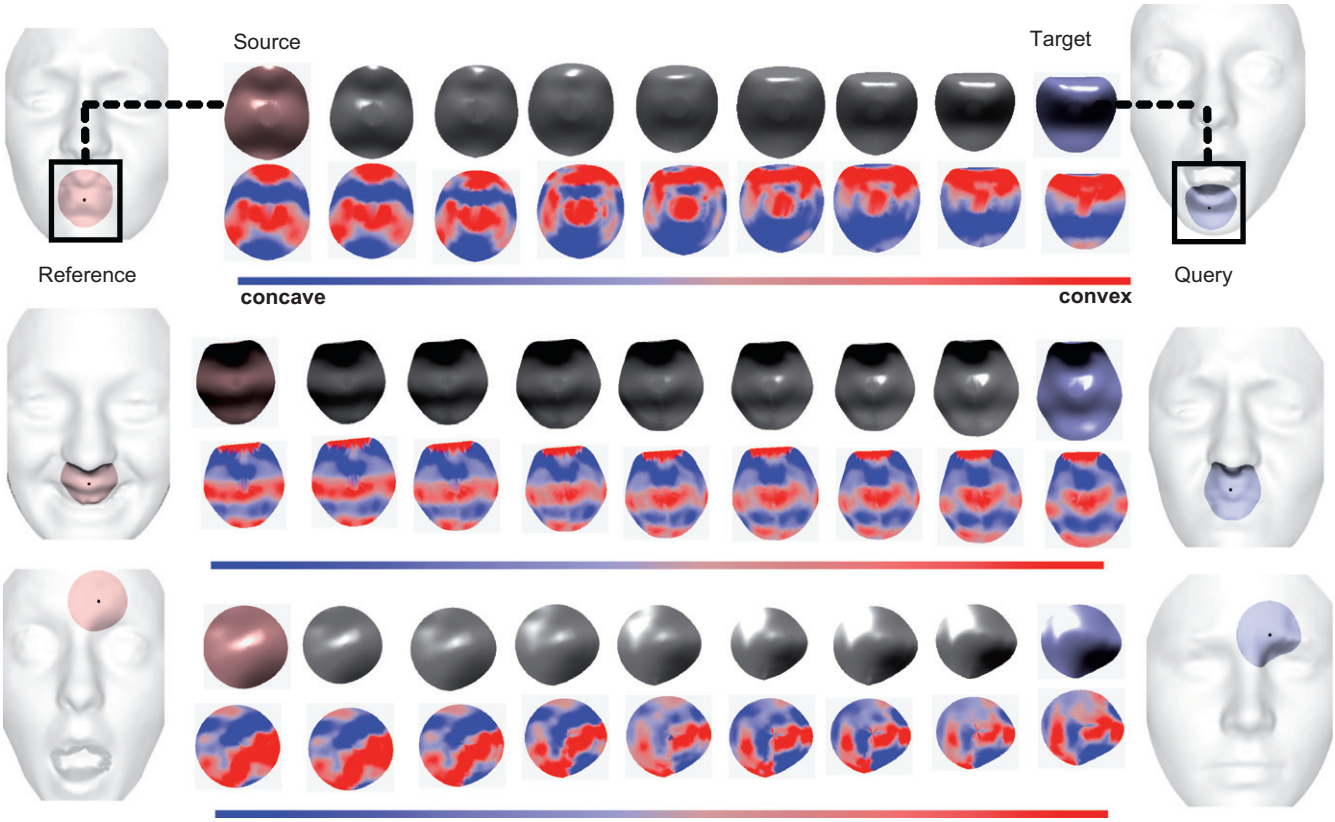

Fig. 4. Examples of inter-class (different expressions) geodesic paths between source and target patches.

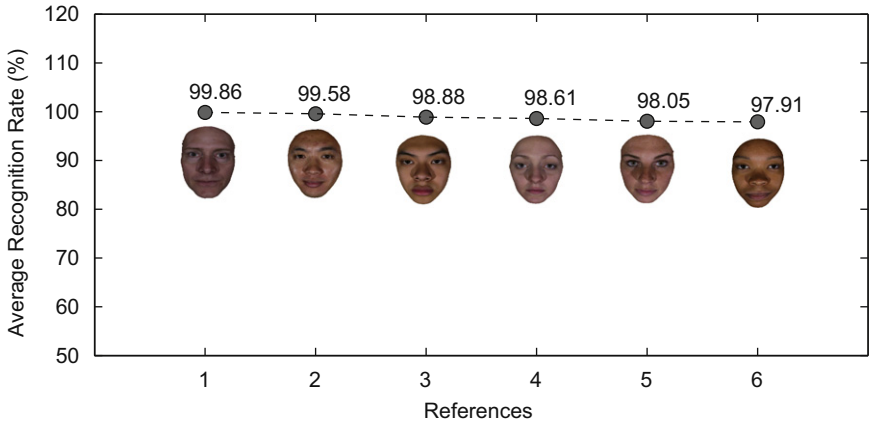

Fig. 5. Different facial expression average recognition rates obtained using different reference subjects (using multiboost-LDA).

7.2. Discussion of the results

Several facial expression recognition experiments were conducted with changing at each time the reference. Fig. 5 illustrates the selected references (neutral scan). Using the *Waikato Environment for Knowledge Analysis (Weka)* [17], we applied the multiboost algorithm with three weak classifiers, namely, Linear Discriminant Analysis (LDA), Naive Bayes (NB), and Nearest Neighbor (NN), to the extracted features, and we achieved average recognition rates of 98.81%, 98.76% and 98.07%, respectively. We applied the SVM linear classifier as well, and we achieved an average recognition rate of 97.75%. We summarize the resulting recognition rates in Table 1.

We note that these rates are obtained by averaging the results of the 10 independent and arbitrarily run experiments (10-fold cross-validation) and their respective recognition rate obtained using the multiboost-LDA classifier. We note that different selections of the reference scans do not affect significantly the recognition results and there is no large variations in recognition rates values. The reported results represent the average over the six runned experiments. The multiboost-LDA classifier achieves the highest recognition rate and shows a better performance in

Table 1
Classification results using local shape analysis and several classifiers.

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

terms of accuracy than the other classifiers. This is mainly due to the capability of the LDA-based classifier to transform the features into a more discriminative space and, consequently, result in a better linear separation between facial expression classes.

The average confusion matrix relative to the best performing classification using multiboost-LDA is given in Table 2.

In order to better understand and explain the results mentioned above, we apply the multiboost algorithm on feature vectors built from distances between patches for each class of expression. In this case, we consider these features as weak classifiers. Then, we look at the early iterations of the multiboost algorithm and the selected patches in each iteration.

Fig. 6 illustrates for each class of expression the most relevant patches. Notice that, for example, for the Happy expression the selected patches are localized in the lower part of the face, around the mouth and the chin. As for the Surprise expression, we can see that most relevant patches are localized around the eyebrows and the mouth region. It can be seen that patches selected for each expression lie on facial muscles that contribute to this expression.

7.3. Comparison with related work

In Table 3 results of our approach are compared against those reported in [11,9,8], on the same experimental setting (54 versus 6 subject partitions) of the BU-3DFE database. The differences between approaches should be noted: Tang et al. [11] performed automatic feature selection using normalized Euclidean distances between 83 landmarks, Soyel et al. [9] calculated six distances using a distribution of 11 landmarks, while Wang et al. [8] derived curvature estimation by locally approximating the 3D surface with a smooth polynomial function. In comparison, our approach capture the 3D shape information of local facial patches to derive shape analysis. For assessing how the results of their statistical analysis will generalize to an independent dataset, in [8] a 20-fold cross-validation technique was used, while in [11,9] the authors used 10-fold cross-validation to validate their approach.

Table 2
Average confusion matrix given by multiboost-LDA classifier.

%	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**

7.4. Non-frontal view facial expression recognition

额面相

In real world situations, frontal view facial scans may not be always available. Thus, non-frontal view facial expression recognition is a challenging issue that needs to be treated. We were interested in evaluating our approach on facial scan under large pose variations. By rotating the 3D shape models in the y-direction, we generate facial scans under six different non-frontal views corresponding to 15°, 30°, 45°, 60°, 75° and 90° rotation. We assume that shape information is unavailable for the occluded facial regions due to the face pose. For each view, we perform facial patches extraction around the visible landmarks in the given scan. In cases where a landmark is occluded, or where the landmark is visible, but the region nearby is partially occluded, we treat it as a missing data problem for all faces sharing this view. In these cases, we are not able to compute the geodesic path between corresponding patches. The corresponding entries in the distance matrix are blank and we fill them using an imputation technique [18]. In our experiments we employed the mean imputation method, which consists of replacing the missing values by the means of values already calculated in frontal view scenario obtained from the training set. Let $d_{ijk} = d_{\mathcal{S}^{[0,\lambda_0]}}(P_i^k, P_j^k)$ be the geodesic distance between the kth patch belonging to subjects i and j ($i \neq j$). In case of frontal view (fv), the set of instances X_i^{fv} relative to the subject i need to be labeled and is given by:

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

where N is the number of attributes. In case of non-frontal view (nfv), if an attribute k is missing, we replace the kth column vector in the distance matrix X_i^{nfv} by the mean of geodesic distances computed in the frontal view case, with respect to the kth attribute and given by: $m_k^{fv} = \sum_{j=1}^{J} d_{ijk}/J$, where J is the total

Table 3
Comparison of this work with respect to previous work [11,9,8].

Cross-validation	This work	Tang et al. [11]	Soyel et al. [9]	Wang et al. [8]
10-fold	**98.81%**	95.1%	91.3%	–
20-fold	**92.75%**	–	–	83.6%

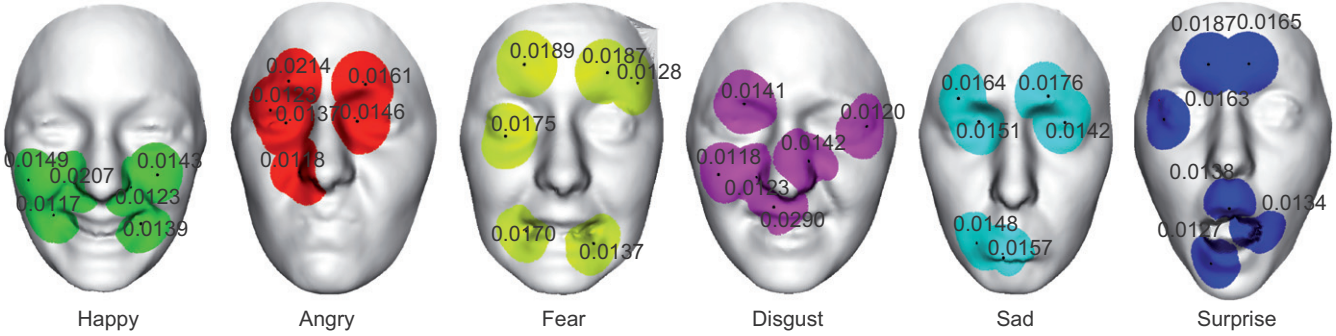

Fig. 6. Selected patches at the early few iterations of multiboost classifier for the six facial expressions (Angry, Disgust, Fear, Happy, Sadness, Surprise) with their associated weights.

number of instances.

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

To evaluate the robustness of our approach in a context of non-frontal views, we derive a view-independent facial expression recognition. Error recognition rates are evaluated throughout different testing facial views using the four classifiers trained only on frontal view facial scans. Fig. 7 shows the average error rates of the four classification methods. The multiboost-LDA shows the best performance for facial expression classification on the chosen database. From the figure, it can be observed that the average error rates increase with the rotation angle (values from 0° to 90° of rotation are considered), and the multiboost-LDA is the best performing methods also in the case of pose variations. As shown in this figure, recognition accuracy remains acceptable, even only 50% of data (half face) are available when we rotate the 3D face by 45° in *y*-direction.

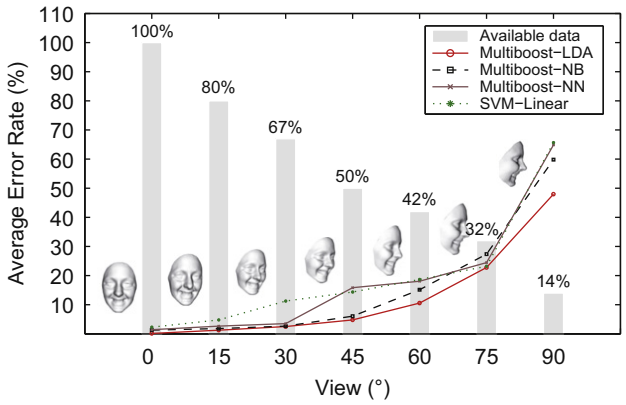

Fig. 7. The average error rates of six expressions with different choices of views corresponding to the best reference and using different classifiers.

7.5. Sensitivity to landmarks mis-localization

It is known that the automatic 3D facial feature points detection is a challenging problem. The most difficult task remains the localization of points around the eyebrow regions, which appear to play an important role in the expression of emotions. The effect of the mis-localization of the landmarks has been addressed in a specific experiment. We considered the eyebrow regions in that the points in these regions are expected to be the most difficult to detect automatically. In these regions, we added noise to the landmarks provided with the BU-3DFED. In particular, we added noise to the position of the landmarks by moving them randomly in a region with a radius of 10 mm, as illustrated Fig. 8 by blue circles. Then we performed expression recognition experiments with such noisy landmarks. The results are reported in Fig. 8. It can be noted that with the multiboost-LDA algorithm the lower decrease in the recognition rate is observed, and even with a recognition rate equal to 85.64% the result still outperforms the one reported in Wang et al [8].

8. Conclusions

In this paper we presented a novel approach for identity-independent facial expression recognition from 3D facial shapes. Our idea was to describe the change in facial expression as a deformation in the vicinity of facial patches in 3D shape scan. An automatic extraction of local curve-based patches within the 3D facial surfaces was proposed. These patches were used as local shape descriptors for facial expression representation. A Riemannian framework was applied to compute the geodesic path between corresponding patches. Qualitative (inter- and intrageodesic paths) and quantitative (geodesic distances) measures of the geodesic path were explored to derive shape analysis. The geodesic distances between patches were labeled with respect to the six prototypical expressions and used as samples to train and test machine-learning algorithms. Using multiboost algorithm for multi-class classification, we achieved a 98.81% average recognition rate for six prototypical facial expressions on the BU-3DFE database. We demonstrated the robustness of the proposed method to pose variations. In fact, the obtained recognition rate remains acceptable (over 93%) even half of the facial scan is missed.

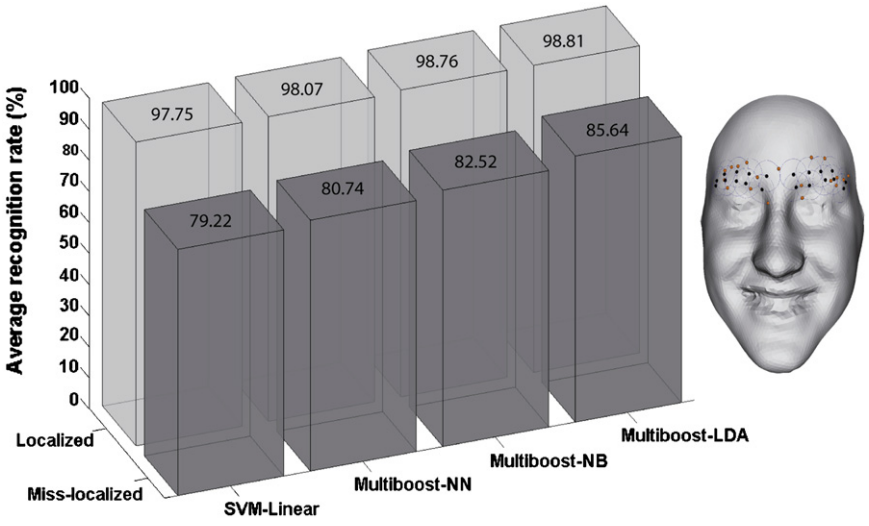

Fig. 8. Recognition experiment performed adding noise to the eyebrow landmarks (random displacement). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The major limitation of our approach is that the 68 landmarks we used to define the facial patches were manually labeled. For our future work we are interested in detecting and tracking facial feature points, as proposed in [19,20], for automatic 3D facial expression recognition.

References

[1] P. Ekman, T.S. Huang, T.J. Sejnowski, J.C. Hager, Final report to nsf of the planning workshop on facial expression understanding, Technical Report, 1992.

[2] P. Ekman, W.V. Friesen, Constants Across Cultures in the Face and Emotion, 1971.

[3] S. Joshi, E. Klassen, A. Srivastava, I.H. Jermyn, A novel representation for Riemannian analysis of elastic curves in $\mathbb{R}^n$, in: Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), 2007.

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

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

[6] J. Whitehill, 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, 2006, pp. 357–362.

[7] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, 1978.

[8] J. Wang, L. Yin, X. Wei, Y. Sun, 3D facial expression recognition based on primitive surface feature distribution, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006, pp. 1399–1406.

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

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

[11] H. Tang, T. Huang, 3D facial expression recognition based on automatically selected features, in: First IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB), 2008, pp. 1–8.

[12] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato, A 3d facial expression database for facial behavior research, in: FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 211–216.

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

[14] A. Srivastava, E. Klassen, S.H. Joshi, I.H. Jermyn, Shape analysis of elastic curves in Euclidean spaces, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, IEEE Computer Society, Los Alamitos, CA, USA, 2010 〈http://doi.ieeecomputersociety.org/10.1109/TPAMI.2010.184〉.

[15] E. Klassen, A. Srivastava, Geodesics between 3d closed curves using path-straightening, in: ECCV, vol. 1, 2006, pp. 95–106.

[16] B. Gong, Y. Wang, J. Liu, X. Tang, Automatic facial expression recognition on a single 3D face by exploring shape deformation, in: Proceedings of the ACM International Conference on Multimedia, Beijing, China, 2009, pp. 569–572.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weak data mining software: an update, SIGKDD Explorations Newsletters 11 (2009) 10–18.

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

[19] S. Gupta, M.K. Markey, A.C. Bovik, Anthropometric 3D face recognition, International Journal of Computer Vision 90 (3) (2010) 331–349.

[20] Y. Sun, X. Chen, M. Rosato, L. Yin, Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis, IEEE Transactions on Systems, Man, and Cybernetics – Part A 40 (3) (2010) 461–474.

Ahmed Maalej is currently a Ph.D. candidate within the Fundamental Computer Science laboratory of Lille (LIFL UMR 8022), France. He obtained the M.S. degree in Telecommunications from the Higher School of Communications of Tunis (SUP'COM), Tunisia, in 2008, and the electrical engineering degree from the National Engineering School of Monastir (ENIM), Tunisia, in 2005. His main research interests focus on 3D facial expression recognition.

Boulbaba Ben Amor received the M.Sc. degree in 2003 and the Ph.D. degree in Computer Science in 2006, both from Ecole Centrale de Lyon, France. Currently, he is an associate-professor in Institut Telecom/Telecom Lille 1. He is also a member of the Computer Science Laboratory in University Lille 1 (LIFL UMR CNRS 8022). His research interests are mainly focused on statistical three-dimensional face analysis and recognition and facial expression recognition using 3D. He is co-author of several papers in refereed journals and proceedings of international conferences. He has been involved in French and International projects and has served as program committee member and reviewer for international journals and conferences.

Mohamed Daoudi is a Full Professor of Computer Science in the Institut TELECOM ; TELECOM Lille 1. He received the Ph.D. degree in Computer Engineering from the University Lille 1, France, in 1993 and Habilitation à Diriger des Recherches (HDR) from the University of Littoral, France, in 2000. He is the founder and the head of the MIIRE research group of LIFL (UMR CNRS 8022). His research interests include pattern recognition, image processing, invariant representation of images and shapes, three-dimensional analysis and retrieval and more recently 3D face recognition. He has published more than 80 papers in refereed journals and proceedings of international conferences. He is the author the book 3D Processing: Compression, Indexing and Watermarking (Wiley, 2008). He has served as a Program Committee member for the International Conference on Pattern Recognition (ICPR) in 2004 and the International Conference on Multimedia and Expo (ICME) in 2004 and 2005. He is a frequent reviewer for IEEE Transactions on Pattern Analysis and Machine Intelligence and for Pattern Recognition Letters. His research has been funded by ANR, RNRT and European Commission grants. He is Senior Member of IEEE.

Anuj Srivastava is a Professor of Statistics at Florida State University in Tallahassee, FL. He obtained his MS and Ph.D. degrees in Electrical Engineering from Washington University in St. Louis in 1993 and 1996, respectively. After spending the year 1996–1997 at Brown University as a visiting researcher, he joined FSU as an Assistant Professor in 1997. He has received the Developing Scholar and the Graduate Faculty Mentor Awards at FSU. His research is focused on pattern theoretic approaches to problems in image analysis, computer vision, and signal processing. He has developed computational tools for performing statistical inferences on certain non-linear manifolds, in particular the shape spaces of curves and surfaces. He has published over 120 journal and conference articles in these areas.

Stefano Berretti received the Laurea Degree in Electronics Engineering, the Postlaurea Degree in Multimedia Content Design, and the Ph.D. degree in Information and Telecommunications engineering in 1997, 2000, and 2001, respectively, from the University of Florence, Italy, where, since 2002, he has been an assistant professor and teaches ''Operating Systems'' and ''Fundamentals of Computer Programming'' in the School of Computer Engineering, and since 2001, he has also been teaching ''Database Systems'' in the postdoctoral school in ''Multimedia Content Design.'' His scientific interests are pattern recognition, content-based image retrieval, 3D object partitioning and retrieval, and 3D face recognition.