



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

Ahmed Maalej, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava,
Stefano Berretti

► To cite this version:

Ahmed Maalej, Boulbaba Ben Amor, Mohamed Daoudi, Anuj Srivastava, Stefano Berretti. Shape analysis of local facial patches for 3D facial expression recognition. *Pattern Recognition*, 2011, 44 (8), pp.1581-1589. 10.1016/j.patcog.2011.02.012 . hal-00661725

HAL Id: hal-00661725

<https://hal.science/hal-00661725>

Submitted on 20 Jan 2012

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.

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 Lille1, 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.*

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 Multi-boosting 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.

Keywords: 3D facial expression classification, shape analysis, geodesic path, multi-boosting, SVM.

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.

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] and [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-

3D FE 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 Multi-boosting and SVM. Finally, discussion and conclusion are given in section 8.

2. Related work

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* vs. *dynamic*; (2) *global* vs. *local*; (3) *2D* vs. *3D*. Most of the approaches are based on feature 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 eleven facial landmarks, and using Neural Network architecture that analysis the calculated distances, they classified the BU-3DFE facial scans into 7 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

75 regularized multi-class AdaBoost classification algorithm, they reported an
76 average recognition rate of 95.1%, and they mentioned that the surprise ex-
77 pression was recognized with an accuracy of 99.2%.

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

86 3. Database Description

87 BU-3DFE is one of the very few publicly available databases of annotated
88 3D facial expressions, collected by Yin et al. [12] at Binghamton University.
89 It was designed for research on 3D human face and facial expression and to
90 develop a general understanding of the human behavior. Thus the BU-3DFE
91 database is beneficial for several fields and applications dealing with human
92 computer interaction, security, communication, psychology, etc. There are a
93 total of 100 subjects in the database, 56 females and 44 males. A neutral
94 scan was captured for each subject, then they were asked to perform six
95 expressions namely: Happiness (HA), Anger (AN), Fear (FE), Disgust (DI),
96 Sad (SA) and Surprise (SU). The expressions vary according to four levels
97 of intensity (low, middle, high and highest or 01-04). Thus, there are 25 3D
98 facial expression models per subject in the database. A set of 83 manually

99 annotated facial landmarks is associated to each model. These landmarks are
 100 used to define the regions of the face that undergo to specific deformations
 101 due to single muscles movements when conveying facial expression [7]. In
 102 Fig. 1, we illustrate examples of the six universal facial expressions 3D models
 including the highest intensity level.



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

103

104 4. 3D Facial Patches-based Representation

105 Most of the earlier work in 3D shape analysis use shape descriptors such as
 106 curvature, crest lines, shape index (e.g., ridge, saddle, rut, dome, etc.). These
 107 descriptors are defined based on the geometric and topological properties of
 108 the 3D object, and are used as features to simplify the representation and
 109 thus the comparison for 3D shape matching and recognition tasks. Despite
 110 their rigorous definition, such features are computed based on numerical
 111 approximation that involves second derivatives and can be sensitive to noisy
 112 data. In case of 3D facial range models, the facial surface labeling is a
 113 critical step to describe the facial behavior or expression, and a robust facial

114 surface representation is needed. In Samir et al. [13] the authors proposed
 115 to represent facial surfaces by an indexed collections of 3D closed curves
 116 on faces. These curves are level curves of a surface distance function (i.e.,
 117 geodesic distance) defined to be the length of the shortest path between a
 118 fixed reference point (taken to be the nose tip) and a point of the extracted
 119 curve along the facial surface. This being motivated by the robustness of the
 120 geodesic distance to facial expressions and rigid motions. Using this approach
 121 they were able to compare 3D shapes by comparing facial curves rather than
 122 comparing corresponding shape descriptors.

123 In our work we intend to further investigate on local shapes of the facial
 124 surface. We are especially interested in capturing deformations of local facial
 125 regions caused by facial expressions. Using a different solution, we compute
 126 curves using the Euclidean distance which is sensitive to deformations and
 127 thus can better capture differences related to variant expressions. To this
 128 end, we choose to consider N reference points (landmarks) $\{r_l\}_{1 \leq l \leq N}$ (Fig.2
 129 (a)) and associated sets of level curves $\{c_\lambda^l\}_{1 \leq \lambda \leq \lambda_0}$ (Fig.2 (b)). These curves
 130 are extracted over the patches centered at these points. Here λ stands for the
 131 value of the distance function between the reference point r_l and the point
 132 belonging to the curve c_λ^l , and λ_0 stands for the maximum value taken by
 133 λ . Accompanying each facial model there are 83 manually picked landmarks,
 134 these landmarks are practically similar to the MPEG-4 feature points and
 135 are selected based on the facial anatomy structure. Given these points the
 136 feature region on the face can be easily determined and extracted. We were
 137 interested in a subset of 68 landmarks laying within the face area, discarding
 138 those marked on the face border. Contrary to the MPEG-4 feature points

139 specification that annotates the cheeks center and bone, in BU-3DFE there
 140 were no landmarks associated with the cheek regions. Thus, we add two extra
 141 landmarks at both cheeks, obtained by extracting the middle point along the
 142 geodesic path between the mouth corner and the outside eye corner.

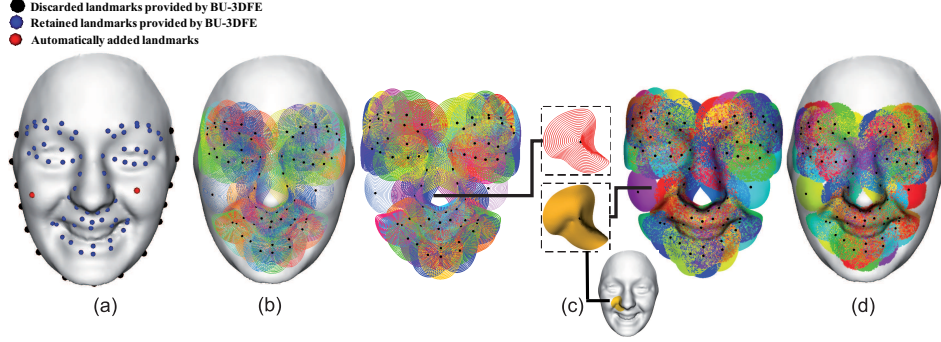


Figure 2: (a) 3D annotated facial shape model (70 landmarks); (b) 3D closed curves extracted around the landmarks; (c) 3D curve-based patches composed of 20 level curves with a size fixed by a radius $\lambda_0 = 20mm$; (d) Extracted patches on the face.

143 We propose to represent each facial scan by a number of patches centered
 144 on the considered points. Let r_l be the reference point and P_l a given patch
 145 centered on this point and localized on the facial surface denoted by S . Each
 146 patch will be represented by an indexed collection of level curves. To extract
 147 these curves, we use the Euclidean distance function $\|r_l - p\|$ to characterize
 148 the length between r_l and any point p on S . Indeed, unlike the geodesic
 149 distance, the Euclidean distance is sensitive to deformations, and besides,
 150 it permits to derive curve extraction in a fast and simple way. Using this
 151 function we defined the curves as level sets of:

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

152 Each c_λ^l is a closed curve, consisting of a collection of points situated at an

153 equal distance λ from r_l . The Fig. 2 resumes the scheme of patches extraction.

154 5. Framework for 3D Shape Analysis

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

166 5.1. 3D Curve Shape Analysis

167 We start by considering a closed curve β in \mathbb{R}^3 . While there are several
 168 ways to analyze shapes of closed curves, an elastic analysis of the parametrized
 169 curves is particularly appropriate in 3D curves analysis. This is because (1)
 170 such analysis uses a square-root velocity function representation which al-
 171 lows us to compare local facial shapes in presence of elastic deformations,
 172 (2) this method uses a square-root representation under which the elastic
 173 metric reduces to the standard L^2 metric and thus simplifies the analysis,
 174 (3) under this metric the Riemannian distance between curves is invariant
 175 to the re-parametrization. To analyze the shape of β , we shall represent it
 176 mathematically using a square-root representation of β as follows ; for an

177 interval $I = [0, 1]$, let $\beta : I \longrightarrow \mathbb{R}^3$ be a curve and define $q : I \longrightarrow \mathbb{R}^3$ to be
 178 its square-root velocity function (SRVF), given by:

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

179 Here t is a parameter $\in I$ and $\|\cdot\|$ is the Euclidean norm in \mathbb{R}^3 . We
 180 note that $q(t)$ is a special function that captures the shape of β and is par-
 181 ticularly convenient for shape analysis, as we describe next. The classical
 182 elastic metric for comparing shapes of curves becomes the \mathbb{L}^2 -metric under
 183 the SRVF representation [14]. This point is very important as it simpli-
 184 fies the calculus of elastic metric to the well-known calculus of functional
 185 analysis under the \mathbb{L}^2 -metric. Also, the squared \mathbb{L}^2 -norm of q , given by:
 186 $\|q\|^2 = \int_{\mathbb{S}^1} \langle q(t), q(t) \rangle dt = \int_{\mathbb{S}^1} \|\dot{\beta}(t)\| dt$, which is the length of β .
 187 In order to restrict our shape analysis to closed curves, we define the set:
 188 $\mathcal{C} = \{q : \mathbb{S}^1 \longrightarrow \mathbb{R}^3 \mid \int_{\mathbb{S}^1} q(t) \|q(t)\| dt = 0\} \subset \mathbb{L}^2(\mathbb{S}^1, \mathbb{R}^3)$. Notice that the
 189 elements of \mathcal{C} are allowed to have different lengths. Due to a non-linear (clo-
 190 sure) constraint on its elements, \mathcal{C} is a non-linear manifold. We can make it
 191 a Riemannian manifold by using the metric: for any $u, v \in T_q(\mathcal{C})$, we define:

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

192 So far we have described a set of closed curves and have endowed it with a
 193 Riemannian structure. Next we consider the issue of representing the *shapes*
 194 of these curves. It is easy to see that several elements of \mathcal{C} can represent
 195 curves with the same shape. For example, if we rotate a curve in \mathbb{R}^3 , we get a
 196 different SRVF but its shape remains unchanged. Another similar situation
 197 arises when a curve is re-parametrized; a re-parameterization changes the

198 SRVF of curve but not its shape. In order to handle this variability, we define
 199 orbits of the rotation group $SO(3)$ and the re-parameterization group Γ as
 200 the equivalence classes in \mathcal{C} . Here, Γ is the set of all orientation-preserving
 201 diffeomorphisms of \mathbb{S}^1 (to itself) and the elements of Γ are viewed as re-
 202 parameterization functions. For example, for a curve $\beta : \mathbb{S}^1 \rightarrow \mathbb{R}^3$ and a
 203 function $\gamma : \mathbb{S}^1 \rightarrow \mathbb{S}^1$, $\gamma \in \Gamma$, the curve $\beta \circ \gamma$ is a re-parameterization of β .
 204 The corresponding SRVF changes according to $q(t) \mapsto \sqrt{\dot{\gamma}(t)}q(\gamma(t))$. We set
 205 the elements of the orbit:

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

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

210 The main ingredient in comparing and analysing shapes of curves is the
 211 construction of a geodesic between any two elements of \mathcal{S} , under the Riemannian
 212 metric given in Eq.(3). Given any two curves β_1 and β_2 , represented
 213 by their SRVFs q_1 and q_2 , we want to compute a geodesic path between the
 214 orbits $[q_1]$ and $[q_2]$ in the shape space \mathcal{S} . This task is accomplished using
 215 a *path-straightening approach* which was introduced in [15]. The basic idea
 216 here is to connect the two points $[q_1]$ and $[q_2]$ by an arbitrary initial path α
 217 and to iteratively update this path using the negative gradient of an energy
 218 function $E[\alpha] = \frac{1}{2} \int_s \langle \dot{\alpha}(s), \dot{\alpha}(s) \rangle ds$. The interesting part is that the gradient
 219 of E has been derived analytically and can be used directly for updating α .
 220 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 $\|r_l - \cdot\|$ 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 $\|r_l - \cdot\| = \lambda$. 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:

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

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

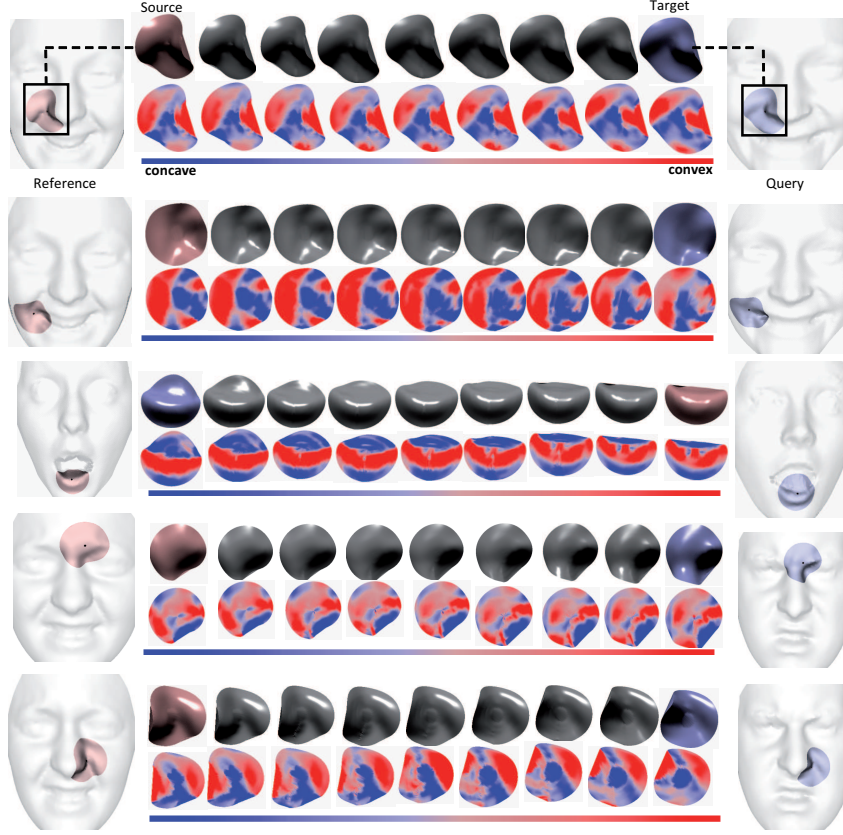


Figure 3: Examples of intra-class (same expression) geodesic paths with shape and mean curvature mapping between corresponding patches.

244 sharing the same expression, and termed *intra-class geodesics*. In the first
 245 column we illustrate the source, which represents scan models of the same
 246 subject, but under different expressions. The third column represents the
 247 targets as scan models of different subjects. As for the middle column, it
 248 shows the geodesic paths. In each row we have both the shape and the
 249 mean curvature mapping representations of the patches along the geodesic
 250 path from the source to the target. The mean curvature representation is
 251 added to identify concave/convex areas on the source and target patches and

252 equally-spaced steps of geodesics. This figure shows that certain patches,
 253 belonging to the same class of expression, are deformed in a similar way.
 254 In contrast, Fig. 4 shows geodesic paths between patches of different facial
 255 expressions. These geodesics are termed *inter-class geodesics*. Unlike the
 256 intra-class geodesics shown in Fig. 3, these patches deform in a different way.

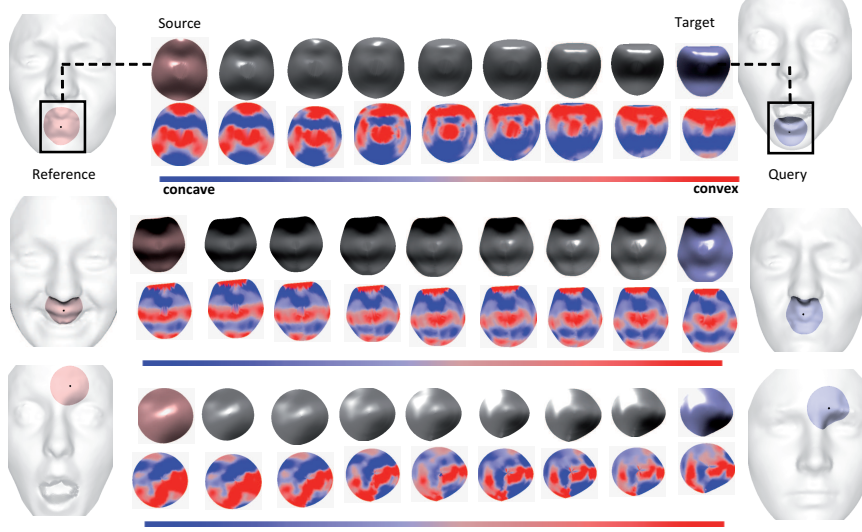


Figure 4: Examples of inter-class (different expressions) geodesic paths between source and target patches.

257 6. Feature Vector Generation for Classification

258 In order to classify expressions, we build a feature vector for each facial
 259 scan. Given a candidate facial scan of a person j , facial patches are extracted
 260 around facial landmarks. For a facial patch P_j^i , a set of level curves $\{c_\lambda\}_j^i$ are
 261 extracted centered on the i^{th} landmark. Similarly, a patch P_{ref}^i is extracted
 262 in correspondence to landmarks of a reference scans ref . The length of the
 263 geodesic path between each level curve and its corresponding curve on the

reference scan are 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 i^{th} 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 machine learning algorithms such as Multi-boosting and SVM, where Multi-boosting 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

288 intensity levels 3D expressive models (03-04). These data are assumed to be
 289 scaled to the true physical dimensions of the captured human faces. Follow-
 290 ing a similar setup as in [16], we randomly divided the 60 subjects into two
 291 sets, the training set containing 54 subjects (648 samples), and the test set
 292 containing 6 subjects (72 samples).

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

300 7.2. Discussion of the Results

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

310 We note that these rates are obtained by averaging the results of the
 311 10 independent and arbitrarily run experiments (10-fold cross validation)
 312 and their respective recognition rate obtained using the Multiboost-LDA

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%

313 classifier. We note that different selections of the reference scans do not
 314 affect significantly the recognition results and there is no large variations in
 315 recognition rates values. The reported results represent the average over the
 316 six runned experiments. The Multiboost-LDA classifier achieves the highest
 317 recognition rate and shows a better performance in terms of accuracy than
 318 the other classifiers. This is mainly due to the capability of the LDA-based
 319 classifier to transform the features into a more discriminative space and,
 320 consequently, result in a better linear separation between facial expression
 321 classes.

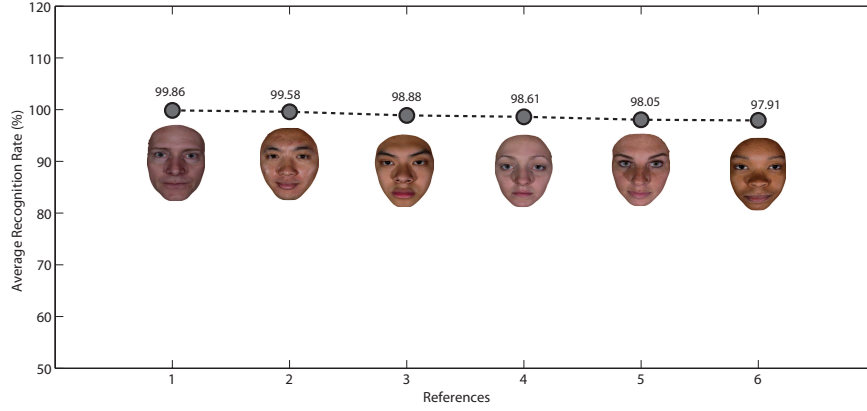


Figure 5: Different facial expression average recognition rates obtained using different reference subjects (using Multiboost-LDA).

322 The average confusion matrix relative to the the best performing classi-

fication using Multiboost-LDA is given in Table 2.

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

323

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

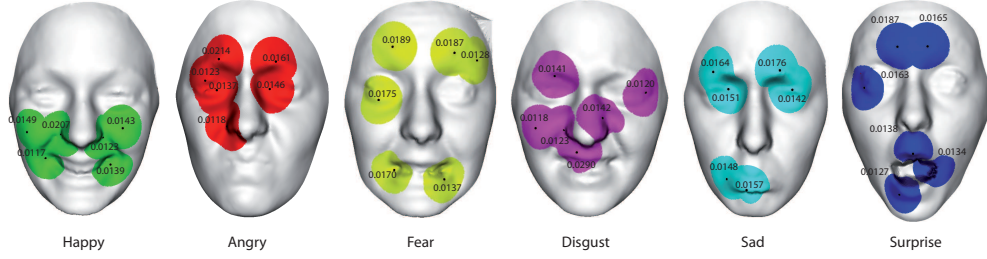


Figure 6: Selected patches at the early few iterations of Multiboost classifier for the six facial expressions (Angry, Disgust, Fear, Happy, Sadness, Surprise) with their associated weights.

329 Fig. 6 illustrates for each class of expression the most relevant patches.
 330 Notice that, for example, for the Happy expression the selected patches are
 331 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], and [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 captures 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] and [9] the authors used 10-fold cross-validation to validate their approach.

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

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

7.4. Non-frontal View Facial Expression Recognition

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 k^{th} 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

370 and is given by:

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

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

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

376 To evaluate the robustness of our approach in a context of non-frontal views,
 377 we derive a view-independent facial expression recognition. Error recognition
 378 rates are evaluated throughout different testing facial views using the four
 379 classifiers trained only on frontal-view facial scans. The Fig. 7 shows the
 380 average error rates of the four classification methods. The Multiboost-LDA
 381 shows the best performance for facial expression classification on the chosen
 382 database. From the figure, it can be observed that the average error rates
 383 increase with the rotation angle (values from 0 to 90 degrees of rotation are
 384 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 degree in y-direction.

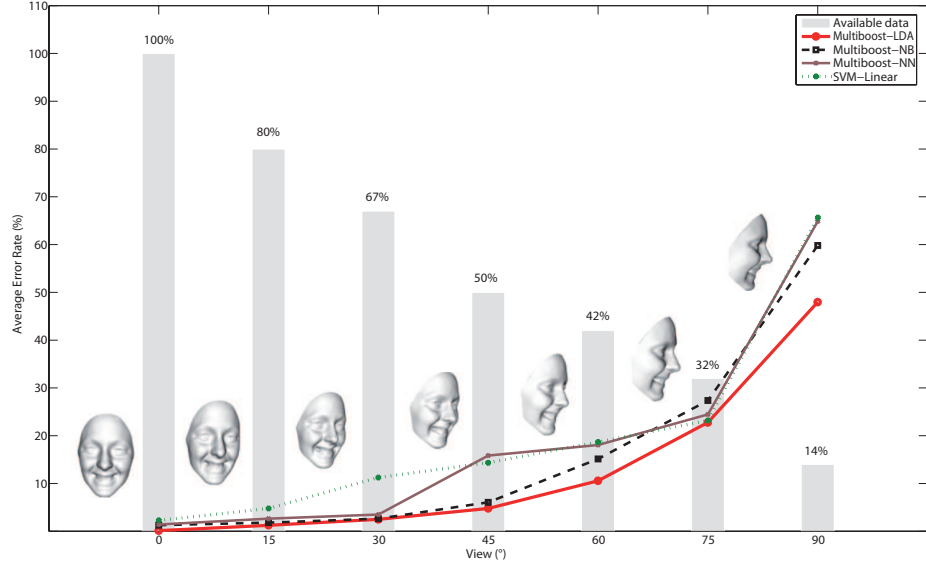


Figure 7: The average error rates of six expressions with different choice of views corresponding to the best reference and using different classifiers.

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

395 to detect automatically. In these regions, we added noise to the landmarks
 396 provided with the BU-3DFED. In particular, we added noise to the position
 397 of the landmarks by moving them randomly in a region with a radius of
 398 $10mm$, as illustrated Fig. 8 by blue circles. Then we performed expression
 399 recognition experiments with such noisy landmarks. The results are reported
 400 in Fig. 8. It can be noted that with the Multiboost-LDA algorithm the lower
 401 decrease in the recognition rate is observed, and even with a recognition rate
 402 equal to 85.64% the result still outperforms the one reported in Wang et al
 [8].

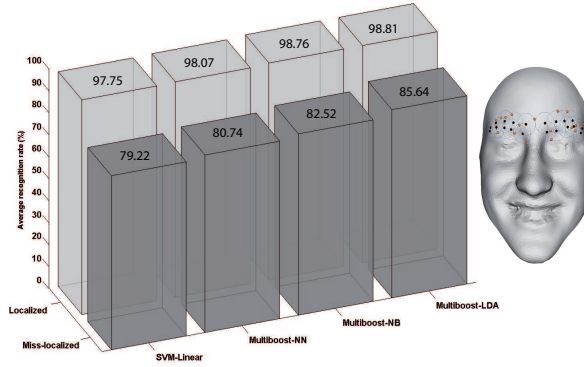


Figure 8: Recognition experiment performed adding noise to the eyebrow landmarks (random displacement).

403

404 8. Conclusions

405 In this paper we presented a novel approach for identity-independent fa-
 406 cial expression recognition from 3D facial shapes. Our idea was to describe
 407 the change in facial expression as a deformation in the vicinity of facial
 408 patches in 3D shape scan. An automatic extraction of local curve based

409 patches within the 3D facial surfaces was proposed. These patches were used
 410 as local shape descriptors for facial expression representation. A Riemannian
 411 framework was applied to compute the geodesic path between correspond-
 412 ing patches. Qualitative (inter and intra-geodesic paths) and quantitative
 413 (geodesic distances) measures of the geodesic path were explored to derive
 414 shape analysis. The geodesic distances between patches were labeled with
 415 respect to the six prototypical expressions and used as samples to train and
 416 test machine learning algorithms. Using Multiboost algorithm for multi-class
 417 classification, we achieved a 98.81% average recognition rate for six proto-
 418 typical facial expressions on the BU-3DFE database. We demonstrated the
 419 robustness of the proposed method to pose variations. In fact, the obtained
 420 recognition rate remain acceptable (over 93%) even half of the facial scan is
 421 missed.

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

426 **References**

- 427 [1] P. Ekman, T. S. Huang, T. J. Sejnowski, J. C. Hager, Final report to
 428 nsf of the planning workshop on facial expression understanding, Tech.
 429 rep., available from Human Interaction Lab, LPPI Box 0984, University
 430 of California, San Francisco, CA 94143 (1992).
- 431 [2] P. Ekman, W. V. Friesen, Constants Across Cultures in the Face and
 432 Emotion (1971).

- 433 [3] S. Joshi, E. Klassen, A. Srivastava, I. H. Jermyn, A novel representation
434 for riemannian analysis of elastic curves in \mathbb{R}^n , in: Proc. IEEE Computer
435 Vision and Pattern Recognition (CVPR), 2007.
- 436 [4] M. Pantic, L. Rothkrantz, Automatic analysis of facial expressions: The
437 state of the art, IEEE Transactions on Pattern Analysis and Machine
438 Intelligence 22 (12) (2000) 1424–1445.
- 439 [5] A. Samal, P. A. Iyengar, Automatic recognition and analysis of human
440 faces and facial expressions: a survey, Pattern Recognition 25 (1) (1992)
441 65–77.
- 442 [6] J. Whitehill, C. W. Omlin, Local versus global segmentation for facial
443 expression recognition, in: FGR '06: Proceedings of the 7th Interna-
444 tional Conference on Automatic Face and Gesture Recognition, 2006,
445 pp. 357–362.
- 446 [7] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for
447 the Measurement of Facial Movement, consulting Psychologists Press
448 (1978).
- 449 [8] J. Wang, L. Yin, X. Wei, Y. Sun, 3d facial expression recognition based
450 on primitive surface feature distribution, IEEE Conference on Computer
451 Vision and Pattern Recognition (CVPR) (2006) 1399–1406.
- 452 [9] H. Soyel, H. Demirel, Facial expression recognition using 3d facial feature
453 distances, International Conference on Image Analysis and Recognition
454 (ICIAR) (2007) 831–838.

- 455 [10] I. Mpiparis, S. Malassiotis, M. G. Strintzis, Bilinear models for 3d face
456 and facial expression recognition, *IEEE Transactions on Information*
457 *Forensics and Security* 3 (3) (2008) 498–511.
- 458 [11] H. Tang, T. Huang, 3d facial expression recognition based on automat-
459 ically selected features, In *First IEEE Workshop on CVPR for Human*
460 *Communicative Behavior Analysis (CVPR4HB)* (2008) 1–8.
- 461 [12] L. Yin, X. Wei, Y. Sun, J. Wang, M. J. Rosato, A 3d facial expression
462 database for facial behavior research, in: *FGR '06: Proceedings of the*
463 *7th International Conference on Automatic Face and Gesture Recogni-*
464 *tion*, 2006, pp. 211–216.
- 465 [13] C. Samir, A. Srivastava, M. Daoudi, E. Klassen, An intrinsic framework
466 for analysis of facial surfaces, *International Journal of Computer Vision*
467 82 (1) (2009) 80–95.
- 468 [14] A. Srivastava, E. Klassen, S. H. Joshi, I. H. Jermyn, Shape analysis of
469 elastic curves in euclidean spaces, *IEEE Transactions on Pattern Anal-*
470 *ysis and Machine Intelligence* accepted for publication.
- 471 [15] E. Klassen, A. Srivastava, Geodesics between 3d closed curves using
472 path-straightening, in: *ECCV* (1), 2006, pp. 95–106.
- 473 [16] B. Gong, Y. Wang, J. Liu, X. Tang, Automatic facial expression recogni-
474 tion on a single 3D face by exploring shape deformation, in: *Proceedings*
475 *of the ACM International Conference on Multimedia*, Beijing, China,
476 2009, pp. 569–572.

- 477 [17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H.
478 Witten, The weka data mining software: An update, SIGKDD Explor.
479 Newsl 11 (2009) 10–18.
- 480 [18] G. Batista, M. C. Monard, An analysis of four missing data treatment
481 methods for supervised learning, Applied Artificial Intelligence 17 (2003)
482 519–533.
- 483 [19] S. Gupta, M. K. Markey, A. C. Bovik, Anthropometric 3d face recogni-
484 tion, International Journal of Computer Vision.
- 485 [20] Y. Sun, X. Chen, M. Rosato, L. Yin, Tracking vertex flow and model
486 adaptation for three-dimensional spatiotemporal face analysis, IEEE
487 Transactions on Systems, Man, and Cybernetics–Part A 40 (3) (2010)
488 461–474.