

4D Facial Expression Recognition by Learning Geometric Deformations

Boulbaba Ben Amor, Hassen Drira, Stefano Berretti, Mohamed Daoudi, Anuj Srivastava

▶ **To cite this version:**

Boulbaba Ben Amor, Hassen Drira, Stefano Berretti, Mohamed Daoudi, Anuj Srivastava. 4D Facial Expression Recognition by Learning Geometric Deformations. IEEE Transactions on Cybernetics, 2014, 44 (12), pp.2443-2457. hal-00949002

HAL Id: hal-00949002
https://hal.science/hal-00949002

Submitted on 5 Mar 2014

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.

4D Facial Expression Recognition by Learning Geometric Deformations

Boulbaba Ben Amor*, *Member, IEEE,* Hassen Drira*, Stefano Berretti, *Member, IEEE,*
Mohamed Daoudi, *Senior, IEEE,* and Anuj Srivastava, *Senior, IEEE*

***Abstract*—In this paper, we present an automatic approach for facial expression recognition from 3D video sequences. In the proposed solution, the 3D faces are represented by collections of radial curves and a Riemannian shape analysis is applied to effectively quantify the deformations induced by the facial expressions, in a given subsequence of 3D frames. This is obtained from the *Dense Scalar Field*, which denotes the shooting directions of the geodesic paths constructed between pairs of corresponding radial curves of two faces. As the resulting Dense Scalar Fields show a high dimensionality, LDA transformation is applied to the dense feature space. Two methods are then used for classification: (i) 3D motion extraction with temporal HMM modeling; and (ii) Mean deformation capturing with Random Forest. While a dynamic HMM on the features is trained in the first approach, the second one computes mean deformations under a window and applies multi-class Random Forest. Both of the proposed classification schemes on the scalar fields showed comparable results and outperformed earlier studies on facial expression recognition from 3D video sequences.**

***Index Terms*—Expression recognition, 4D data, Riemannian geometry, Temporal analysis, HMM, Random Forest.**

I. Introduction

OVER the last few years, automatic recognition of facial expressions emerged as a field of active research, with applications in several different areas, such as human-machine interaction, psychology, computer graphics, transport security (by detecting driver fatigue, for example), and so on.

The importance of facial expressions was first realized and investigated by psychologists, among others. In a seminal work by Mehrabian et al. [1] the relative importance of verbal and nonverbal messages in communicating feelings and attitude is described. In particular, they provided evidence that face-to-face communication is governed by the 7%-38%-55% rule, that balances the relevance of verbal, vocal and visual elements, respectively, in communications. Despite this rigid quantification has since been refuted in later studies, it still provides an indication that the words and tone of the voice form only a part of human communication. The non-verbal elements related to the body language (e.g., gestures, posture, facial expressions) also play an important role. Starting from a different point of view, Ekman [2] conducted the first systematic studies on facial expressions in the late 70s. Through his experiments, it is demonstrated that there is an universal set of facial expressions representing *anger*, *disgust*, *fear*, *happiness*, *sadness* and *surprise*, plus the *neutral* one, that are universally recognized and remain consistent across different ethnicity and cultures. The presence of these prototypical facial expressions is now widely accepted for scientific analysis. Ekman also showed that facial expressions can be coded through the movement of face points as described by a set of *action units* [3].

These results, in turn, inspired many researchers to analyze facial expressions in video data, by tracking facial features and measuring the amount of facial movements in video frames [4]. This body of work demonstrates a collective knowledge that facial expressions are highly dynamical processes, and looking at sequences of face instances can help to improve the recognition performance. We further emphasize that, rather than being just a static or dynamic 2D image analysis, it is more natural to analyze expressions as spatio-temporal deformations of 3D faces, caused by the actions of facial muscles. In this approach, the facial expressions can be studied comprehensively by analyzing temporal dynamics of 3D face scans (3D plus time is often regarded as 4D data). From this perspective the relative immunity of 3D scans to lighting conditions and pose variations give support for the use of 3D and 4D data. Motivated by these considerations, there has been a progressive shift from 2D to 3D in performing facial shape analysis for recognition [5–9], and expression recognition [10], [11]. In particular, this latter research subject is gaining momentum thanks to the recent availability of public 3D datasets, like the *Binghamton University* 3D Facial Expression database (BU-3DFE) [12], and the *Bosphorus* 3D Face Database [13]. At the same time, advances in 3D imaging technology have permitted collections of large datasets that include temporal sequences of 3D scans (i.e., 4D datasets), such as the *Binghamton University* 4D Facial Expression database (BU-4DFE) [14], the 4D dataset constructed at *University of Central Lancashire* (Hi4D-ADSIP) [15], [16], and the dynamic 3D FACS dataset (D3DFACS) for facial expression research [17], which also includes fully coded FACS. This trend has been strengthened further by the introduction of inexpensive acquisition devices, such as the consumer 3D cameras like Kinect or Asus that provide fast albeit low-resolution streams of 3D data to a large number of users, thus opening new opportunities and challenges in 3D face recognition and facial expression recognition [18], [19].

Motivated by these facts, we focus in this paper on the problem of expression recognition from dynamic sequences of

*B. Ben Amor and H. Drira contributed equally to this work.

B. Ben Amor, H. Drira, and M. Daoudi are with Institut Mines-Télécom/Télécom Lille; Laboratoire d'Informatique Fondamentale de Lille (LIFL UMR 8022), France.

S. Berretti is with Department of Information Engineering, University of Florence, Florence, Italy.

A. Srivastava is with Department of Statistics, Florida State University, Tallahassee, FL 32306, USA.

3D facial scans. We propose a new framework for temporal analysis of 3D faces that combines scalar field modeling of face deformations with effective classifiers. To motivate our solution and to relate it to the state of the art, next we provide an overview of existing methods for 4D facial expression recognition (see also the recent work in [20] for a comprehensive survey on this subject), then we give a general overview of our approach.

A. Related Work

The use of 4D data for face analysis is still at the beginning, with just a few works performing face recognition from sequences of 3D face scans [19], [21], [22], and some works focussing on facial expression recognition.

In particular, the first approach addressing the problem of facial expression recognition from dynamic sequences of 3D scans was proposed by Sun et al. [23], [24]. Their approach basically relies on the use of a generic deformable 3D model whose changes are tracked both in space and time in order to extract a spatio-temporal description of the face. In the temporal analysis, a vertex flow tracking technique is applied to adapt the 3D deformable model to each frame of a 3D face sequence. Correspondences between vertices across the 3D dynamic facial sequences provide a set of motion trajectories (vertex flow) of 3D face scans. As a result, each depth scan in the sequence can be represented by a spatio-temporal feature vector that describes both shape and motion information and provides a robust facial surface representation. Once spatio-temporal features are extracted, a two-dimensional Hidden Markov Model (HMM) is used for classification. In particular, a spatial HMM and a temporal HMM were used to model the spatial and temporal relationships between the extracted features. Exhaustive analysis was performed on the BU-4DFE database. The main limit of this solution resides in the use of the 83 manually annotated landmarks of the BU-4DFE that are not released for public use.

The approach proposed by Sandbach et al. [25] exploits the dynamics of 3D facial movements to analyze expressions. This is obtained by first capturing motion between frames using Free-Form Deformations and extracting motion features using a quad-tree decomposition of several motion fields. GentleBoost classifiers are used in order to simultaneously select the best features to use and perform the training using two classifiers for each expression: one for the onset temporal segment, and the other for the offset segment. Then, HMMs are used for temporal modeling of the full expression sequence, which is represented as the composition of four temporal segments, namely, neutral, onset, apex, offset. These model a sequence with an initial neutral segment followed by the activation of the expression, the maximum intensity of the expression, deactivation of the expression and closing of the sequence again with a neutral expression. Experiments were reported for three prototypical expressions (i.e., happy, angry and surprise) of the BU-4DFE database. An extension of this work has been presented in [20], where results on the BU-4DFE database using the six universal facial expressions are reported.

In [26] a level curve based approach is proposed by Le et al. to capture the shape of 3D facial models. The level curves are parameterized using the arclength function. The Chamfer distance is applied to measure the distances between the corresponding normalized segments, partitioned from these level curves of two 3D facial shapes. These features are then used as spatio-temporal features to train HMM, and since the training data were not sufficient for learning HMM, the authors proposed to apply the universal background modeling to overcome the over-fitting problem. Results were reported for the happy, sad and surprise sequences of the BU-4DFE database.

Fang et al. [27] propose a fully automatic 4D facial expression recognition approach with a particular emphasis on 4D data registration and dense correspondence between 3D meshes along the temporal line. The variant of the Local Binary Patterns (LBP) descriptor proposed in [28], which computes LBP on three orthogonal planes is used as face descriptor along the sequence. Results are provided on the BU-4DFE database for all expressions and for the subsets of expressions used in [25] and [26], showing improved results with respect to competing solutions. In [29], the same authors propose a similar methodology for facial expression recognition from dynamic sequences of 3D scans, with an extended analysis and comparison of different 4D registration algorithms, including ICP and more sophisticated mesh matching algorithms, as Spin Images and MeshHOG. However, 12 manually annotated landmarks were used in this study.

Recently, Reale et al. [30] have proposed a new 4D spatio-temporal feature named *Nebula* for facial expressions and movement analysis from a volume of 3D data. After fitting the volume data to a cubic polynomial, a histogram is built for different facial regions using geometric features, as curvatures and polar angles. They have conducted several recognition experiments on the BU-4DFE database for posing expressions, and on a new database published in [31] for spontaneous expressions. However, manual intervention is used to detect the onset frame and just 15 frames from the onset one are used for classication, and these frames correspond to the most intense expression.

From the discussion above, it becomes clear that solutions specifically tailored for 4D facial expression recognition from dynamic sequences are still preliminary, being semi-automatic, or are capable of discriminating between only a subset of expressions.

B. Our Method and Contributions

Due to the increasing importance of shape analysis of objects in different applications, including 3D faces, a variety of mathematical representations and techniques have been suggested, as described above [20], [24], [29]. The difficulty in analyzing shapes of objects comes from the fact that: (1) Shape representations, metrics, and models should be invariant to certain transformations that are termed *shape preserving*. For instance, rigid motions and re-parameterizations of facial surfaces do not change their shapes, and any shape analysis of faces should be invariant to these transformations. (2)

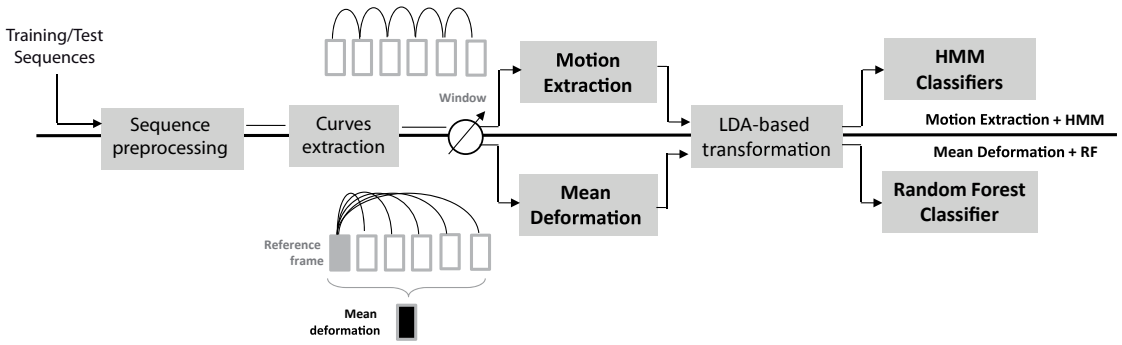


Fig. 1. Overview of the proposed approach. Four main steps are shown: Sequence preprocessing and extraction of the radial curves; Motion extraction and Mean deformation computation; Dimensionality reduction with LDA; HMM- and Random-Forest-based classification. Note that both train and test sequences can go through the upper and lower path in the block-diagram.

Registration of points across objects is an important ingredient in shape analysis. Specifically, in comparing shapes of faces, it makes sense that similar biological parts are registered to each other across different faces. Furthermore, it is important to use techniques that allow a joint registration and comparisons of surfaces in a comprehensive framework, rather than in two separate steps. These two issues— invariance and registration— are naturally handled using Riemannian methods where one can choose metrics that are invariant to certain transformations and form quotient spaces (termed shape spaces) by forming equivalence classes of objects that have the same shape. The elastic Riemannian metric used in this paper provides a nice physical interpretation of measuring deformations between facial curves using a combination of stretching and bending. These elastic deformations are captured by the Dense Scalar Field features used in this paper for classifications. In summary, the main motivation of using a Riemannian approach is to perform registration that matches corresponding anatomical features, and obtain deformation fields that are physically interpretable.

Based on these premises, in this work we extend the ideas presented in [32] to propose an automatic approach for facial expression recognition that exploits the facial deformations extracted from 3D facial videos. An overview of the proposed approach is given in Fig. 1. In the preprocessing step, the 3D mesh in each frame is first aligned to the previous one and then cropped. From the obtained subsequence, the 3D deformation is captured based on a Dense Scalar Field (DSF) that represents the 3D deformation between two frames. Linear Discriminant Analysis (LDA) is used to transform derived feature space to an optimal compact space to better separate different expressions. Finally, the expression classification is performed in two ways: (1) using the HMM models for temporal evolution; and (2) using mean deformation along a window with Random Forest classifier. Experimental results show that the proposed approaches are capable of improving the state of art performance on the BU-4DFE database. There are three main contributions in this work,

- Novel Dense Scalar Fields (DSFs) defined on radial curves of 3D faces using Riemannian analysis in shape spaces of curves. These scalar fields accurately capture deformations occurring between 3D faces represented as collections of radial curves;
- A new approach for facial expression recognition from 3D dynamic sequences, that combines the high descriptiveness of DSFs extracted from successive 3D scans of a sequence with the discriminant power of LDA features using HMM and multi-class Random Forest;
- An extensive experimental evaluation that compares the proposed solution with the state of the art methods using a common dataset and testing protocols. Results show that our approach outperforms the published state of the art results.

The rest of the paper is organized as follows: In Sect. II, we present a face representation model that captures facial features relevant to categorizing expression variations in 3D dynamic sequences. In Sect. III, the dynamic shape deformation analysis using LDA and classification using HMM and multi-class Random Forest are addressed. The main characteristics of the BU-4DFE and the preprocessing operations performed on the face scans are described in Sect. IV, with the experimental results and the comparative evaluation performed on the BU-4DFE database reported and discussed in the same Section. Finally, conclusions and future research directions are outlined in Sect. V.

II. Geometric Facial Deformation

One basic idea to capture facial deformations across 3D video sequences is to track mesh vertices densely along successive 3D frames. Since, as the resolution of the meshes varies across 3D video frames, establishing a dense matching on consecutive frames is necessary. For this purpose, Sun et al. [23] proposed to adapt a generic model (a tracking model) to each 3D frame using a set of 83 predefined facial landmarks to control the adaptation based on radial basis functions. A second solution is presented by Sandbach et al. [25], [33], where the authors used an existing non-rigid registration algorithm (FFD) [34] based on B-splines interpolation between a lattice of control points. In this case, dense matching is a preprocessing step used to estimate a motion vector field between 3D frames t and t-1. The problem of quantifying subtle deformations along the sequence still

remains a challenging task, and the results presented in [25] are limited to just three facial expressions: *happy*, *angry* and *surprise*.

In order to capture and model deformations of the face induced by different facial expressions, we propose to represent the facial surface through a set of parameterized radial curves that originate from the tip of the nose. Approximating the facial surface by an ordered set of radial curves, which locally captures its shape can be seen as a parameterization of the facial surface. Indeed, similar parameterizations of the face have shown their effectiveness in facial biometrics [35]. The mathematical setup for the shape theory offered here comes from Hilbert space analysis. A facial surface is represented by a collection of radial curves and a Riemannian framework is used to study shapes of these curves. We start by representing facial curves as absolutely continuous maps from $\beta : [0, 1] \rightarrow \mathbb{R}^3$ and our goal is to analyze shapes represented by these maps. The problem in studying shapes using these maps directly is that they change with re-parameterizations of curves. If γ is a re-parameterization function (typically a diffeomorphism from $[0, 1]$ to itself), then under the standard $\mathbb{L}^2$ norm, the quantity $\|\beta_1 - \beta_2\| \neq \|\beta_1 \circ \gamma - \beta_2 \circ \gamma\|$, which is problematic. The solution comes from choosing a Riemannian metric under which this inequality becomes equality and the ensuing analysis simplifies. As described in [36], we represent the facial curves using a new function q, called the square-root velocity function (SRVF) (see Eq. (1)). The advantage of using SRVF representation is that under this representation the elastic metric becomes the standard $\mathbb{L}^2$ metric and an identical re-parameterization of curves preserves the $\mathbb{L}^2$ norm of between their SRVFs. The mapping from a curve β to q is a bijection (up to a translation) and the space of all SRVFs is the Hilbert space of all square-integrable maps of the type $q : [0, 1] \rightarrow \mathbb{R}^3$. This space under the natural $\mathbb{L}^2$ inner product is actually a vector space and geodesics between points in this space are straight lines.

With the proposed representation, a facial surface is approximated by an indexed collection of radial curves β_α, where the index α denotes the angle formed by the curve with respect to a *reference* radial curve. In particular, the reference radial curve (i.e., the curve with $\alpha = 0$) is chosen as oriented along the vertical axis, while the other radial curves are separated each other by a fixed angle and are ordered in a clockwise manner. As an example, Fig. 2(a) shows the radial curves extracted for a sample face with happy expression. To extract the radial curves, the nose tip is accurately detected and each face scan is rotated to the upright position so as to establish a direct correspondence between radial curves having the same index in different face scans (the preprocessing steps, including nose tip detection and pose normalization are discussed in more detail in Sect. IV-A). In Fig. 2(b)-(c), two radial curves at $\alpha = 90°$ in the neutral and happy scans of the same subject are shown. As emerged in the plot (d) of the same figure, facial expressions can induce consistent variations in the shape of corresponding curves. These variations change in strength from expression to expression and for different parts of the face. In order to effectively capture these variations a Dense Scalar Field is proposed, which relies on a Riemannian analysis of facial shapes.

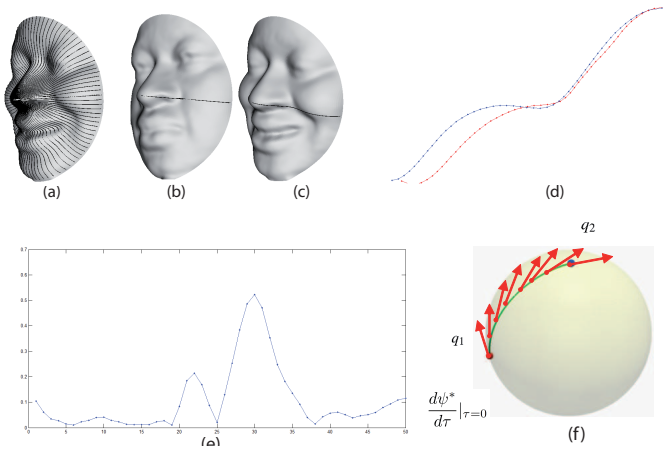


Fig. 2. The figure illustrates: (a) The extracted radial curves; (b)-(c) A radial curve on a neutral face, and the correspondent curve on the same face with happy expression, respectively; (d) The two radial curves are plotted together; (e) The values of the magnitude of $\frac{d\psi^*}{d\tau}|_{\tau=0}(k)$ computed between the curves in (d) are reported for each point k of the curves; (f) The parallel vector field across the geodesic between q_1 and q_2 in the space of curves $\mathcal{C}$.

Considering a generic radial curve β of the face, it can be parameterized as β: $I \rightarrow \mathbb{R}^3$, with $I = [0, 1]$, and mathematically represented through the *square-root velocity function* (SRVF) [36], [37], denoted by $q(t)$, according to:

$$q(t) = \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}, \quad t \in [0, 1]. \tag{1}$$

This specific representation has the advantage of capturing the shape of the curve and makes the calculus simpler. Let us define the space of the SRVFs as $\mathcal{C} = \{q : I \rightarrow \mathbb{R}^3, \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3)$, with $\|\cdot\|$ indicating the $\mathbb{L}^2$ norm. With the $\mathbb{L}^2$ metric on its tangent space, $\mathcal{C}$ becomes a Riemannian manifold. Basically, with this parametrization each radial curve is represented on the manifold $\mathcal{C}$ by its SRVF. According to this, given the SRVFs q_1 and q_2 of two radial curves, the shortest path ψ^* on the manifold $\mathcal{C}$ between q_1 and q_2 (called *geodesic path*) is a critical point of the following energy function:

$$E(\psi) = \frac{1}{2} \int ||\dot{\psi}(\tau)||^2 d\tau, \tag{2}$$

where ψ denotes a path on the manifold $\mathcal{C}$ between q_1 and q_2, τ is the parameter for traveling along the path ψ, $\dot{\psi} \in T_\psi(\mathcal{C})$ is the tangent vector field on the curve $\psi \in \mathcal{C}$, and $||.||$ denotes the $\mathbb{L}^2$ norm on the tangent space.

Since elements of $\mathcal{C}$ have a unit $\mathbb{L}^2$ norm, $\mathcal{C}$ is a hypersphere in the Hilbert space $\mathbb{L}^2(I, \mathbb{R}^3)$. As a consequence, the geodesic path between any two points $q_1, q_2 \in \mathcal{C}$ is simply given by the minor arc of the great circle connecting them on this hypersphere, $\psi^* : [0, 1] \rightarrow \mathcal{C}$. This is given by:

$$\psi^*(\tau) = \frac{1}{\sin(\theta)} \left(\sin((1 - \tau)\theta)q_1 + \sin(\theta\tau)q_2\right), \tag{3}$$

where $\theta = d_{\mathcal{C}}(q_1, q_2) = cos^{-1}(\langle q_1, q_2 \rangle)$. We point out that $sin(\theta) = 0$, if the distance between the two curves is zero, in other words $q_1 = q_2$. In this case, for each τ, $\psi^*(\tau) = q_1 = q_2$.

The tangent vector field on this geodesic is then written as $\frac{d\psi^*}{d\tau} : [0,1] \rightarrow T_\psi(\mathcal{C})$, and is obtained by the following equation:

$$\frac{d\psi^*}{d\tau} = \frac{-\theta}{\sin(\theta)} \left(\cos((1-\tau)\theta)q_1 - \cos(\theta\tau)q_2\right). \quad (4)$$

Knowing that on geodesic path, the covariant derivative of its tangent vector field is equal to 0, $\frac{d\psi^*}{d\tau}$ is parallel along the geodesic ψ^* and one can represent it with $\frac{d\psi^*}{d\tau}|_{\tau=0}$ without any loss of information. Accordingly, Eq. (4) becomes:

$$\frac{d\psi^*}{d\tau}|_{\tau=0} = \frac{\theta}{\sin(\theta)} \left(q_2 - \cos(\theta)q_1\right) \ (\theta \neq 0). \quad (5)$$

A graphical interpretation of this mathematical representation is given in Fig. 2. In Fig. 2(a), we show a sample face with happy expression and all the extracted radial curves. In Fig. 2(b) and Fig. 2(c) two corresponding radial curves (i.e., radial curves at the same angle α), respectively, on a neutral and a happy face of the same person are highlighted. These curves are reported together in Fig. 2(d), where the amount of deformation between them can be appreciated, although the two curves lie at the same angle α and belong to the same person. The amount of deformation between the two curves is calculated using Eq. (5), and the plot of the magnitude of this vector at each point of the curve is reported in Fig. 2(e) (i.e., 50 points are used to sample each of the two radial curves as reported on the x axis, while the magnitude of the vector field is reported on the y axis). Finally, Fig. 2(f) illustrates the idea to map the two radial curves on the hypersphere $\mathcal{C}$ in the Hilbert space through their SRVFs q_1 and q_2, and shows the geodesic path connecting these two points on the hypersphere. The tangent vectors of this geodesic path represent a vector field whose covariant derivative is zero. According to this, $\frac{d\psi^*}{d\tau}|_{\tau=0}$ becomes sufficient to represent this vector field, with the remaining vectors obtained by parallel transport of $\frac{d\psi^*}{d\tau}|_{\tau=0}$ along the geodesic ψ^*.

Based on the above representation, we define a *Dense Scalar Field* capable to capture deformations between two corresponding radial curves β^1_α and β^2_α of two faces approximated by a collection of radial curves.

Definition 1: **Dense Scalar Field (DSF)**
Let $x_\alpha(t) = ||\frac{d\psi^*_\alpha}{d\tau}|_{\tau=0}(t)||$ be the values of the magnitude computed for each point t of the curves q^1_α and q^2_α; let T be the number of sampled points per curve, and $|\Lambda|$ be the number of curves used per face. According to this, we define the function f by:

$$f : \mathcal{C} \times \mathcal{C} \longrightarrow (\mathbb{R}^+)^{\mathcal{T}},$$
$$f(q^1_\alpha, q^2_\alpha) = (x^1_\alpha, \ldots, x^k_\alpha, \ldots, x^T_\alpha).$$

Assuming that $\{\beta^1_\alpha | \alpha \in \Lambda\}$ and $\{\beta^2_\alpha | \alpha \in \Lambda\}$ be the collections of radial curves associated with the two faces F^1 and F^2 and let q^1_α and q^2_α be their SRVFs, the Dense Scalar Fields (DSF) vector is defined by:

$$DSF(F^1, F^2) = (f(q^1_0, q^2_0), \ldots, f(q^1_\alpha, q^2_\alpha), \ldots, f(q^1_{|\Lambda|}, q^2_{|\Lambda|})).$$

The dimension of the DSF vector is $|\Lambda| \times T$.

The steps to compute the proposed DSF are summarized in Algorithm 1.

Algorithm 1 – Computation of the Dense Scalar Field

Input: Facial surfaces F^1 and F^2; T, number of sample points on a curve; $\Delta\alpha$, angle between successive radial curves; $|\Lambda|$, number of curves per face
Output: $DSF(F^1, F^2)$, the DSF between the two faces

```
procedure COMPUTEDSF(F^1, F^2, T, Δα, |Λ|)
    n ← 0
    while n < |Λ| do
        α = n · Δα
        for i ← 1, 2 do
            extract the curve β^i_α
            compute the SRVF of β^i_α:
                q^i_α(t) ≐ β̇^i_α(t) / sqrt(||β̇^i_α(t)||) ∈ C, t = 1, ..., T
        end for
        compute the distance between q^1_α and q^2_α:
                θ = d_C(q^1_α, q^2_α) = cos^{-1}(⟨q^1_α, q^2_α⟩)
        compute the deformation vector (dψ*/dτ)|_{τ=0} using
        Eq. (5) as:
            f(q^1_α, q^2_α) = (x_α(1), x_α(2), ..., x_α(T)) ∈ R^T_+
            x_α(t) = |θ/sin(θ) (q^2_α − cos(θ) q^1_α)|, t = 1, ..., T
    end while
    compute DSF(F^1, F^2) as the magnitude
    of (dψ*/dτ)|_{τ=0}(k):
        DSF(F^1, F^2) = (f(q^1_0, q^2_0), ..., f(q^1_{|Λ|}, q^2_{|Λ|}))
    return DSF
end procedure
```

The first step to capture the deformation between two given 3D faces F^1 and F^2 is to extract the radial curves originating from the nose tip. Let β^1_α and β^2_α denote the radial curves that make an angle α with a reference radial curve on faces F^1 and F^2, respectively. The initial tangent vector to ψ^*, called also the shooting direction, is computed using Eq. (5). Then, we consider the magnitude of this vector at each point t of the curve in order to construct the DSFs of the facial surface. In this way, the DSF quantifies the local deformation between points of radial curves β^1_α and β^2_α, respectively, of the faces F^1 and F^2. In the practice, we represent each face with 100 radial curves, and T=50 sampled points on each curve, so that the DSFs between two 3D faces is expressed by a 5000-dimensional vector.

In Fig. 3 examples of the deformation fields computed between a neutral face of a given subject and the apex frames of the sequences of the six prototypical expressions of the same subject are shown. The values of the scalar field to be applied on the neutral face to convey the six different prototypical expressions are reported using a color scale. In particular, colors from green to red represent the highest deformations, whereas the lower values of the dense scalar field are represented in cyan/blue. As it can be observed, for different expressions, the high deformations are located in different regions of the face. For example, as intuitively expected, the corners of the mouth and the cheeks are mainly deformed for happiness expression, whereas the eyebrows are

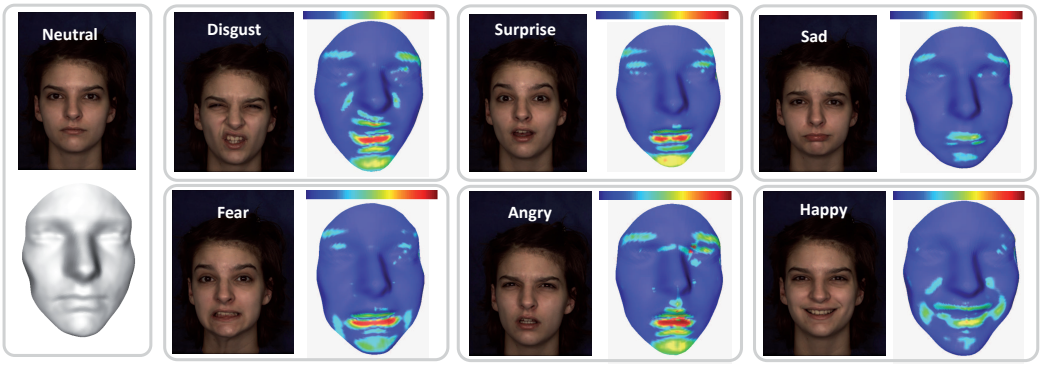


Fig. 3. Deformation Scalar Fields computed between a neutral face of a given subject and the apex frames of the sequences of the six prototypical expressions of the same subject. The neutral scan is shown on the left. Corresponding texture images are also illustrated with each DSFs colormap.

also strongly deformed for the angry and disgust expressions.

A. Effect of the Nose Tip Localisation Inaccuracy on the DSF Computation

In the following, we present a study on the effects that possible inaccuracies in the detection of the nose tip can have on the computation of the proposed dense scalar field. In particular, we consider the effects on the shooting directions of the geodesic paths and the radial curves originating from the nose tip.

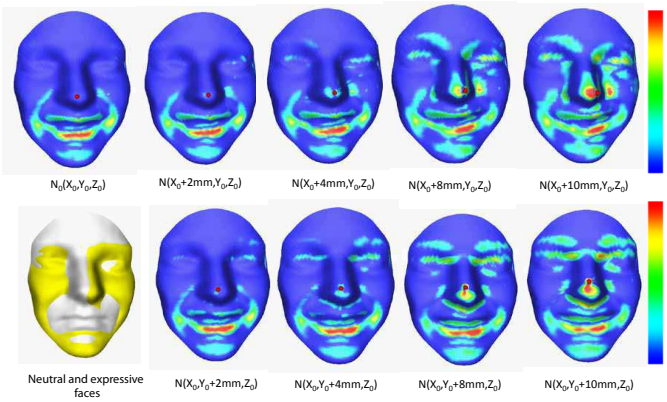


Fig. 4. Effect of the nose tip placement inaccuracy on the shooting directions of the geodesic paths (or the DSFs computation). The first row illustrates DSFs when varying the nose tip position along the X-direction; The second row shows DSFs when the variation is performed along the Y-direction.

We have changed the nose tip coordinates in the X- and Y-directions and have reported the DSFs compuation results (using colormaps on the expressive faces) in Fig. 4. As illustrated in this figure, a large localization error (> 4mm) of the nose tip generates false deformations, which could impact negatively the performance of the approach. In fact, our method is based on learning such geometric deformations to build HMMs or Random Forest classifiers. However, the left side of the figure illustrates the fact that the DSFs computation tolerates quite well errors up to 4mm.

B. DSF Compared to other Features

In order to compare the proposed DSF feature against other methods for extracting dense deformation features, we selected the Free-Form Deformation approach, which has been originally defined in Rueckert et al. [38] for medical images, and later on successfully applied to the problem of 3D dynamic facial expression recognition by Sandbach et al. [25], [33]. In particular, FFD is a method for non-rigid registration based on B-spline interpolation between a lattice of control points. In addition, we also compared our approach with respect to a baseline solution, which uses the point-to-point Euclidean distance between frames of a sequence. Figure 5 reports the results for an example case, where a frame of a happy sequence is deformed with respect to the first frame of the sequence. The figure shows quite clearly as the DSF proposed in this work is capable to capture the face deformations with smooth variations that include, in the example, the mouth, the chin and the cheeck. This result is important to discriminate between different expressions whose effects are not limited to the mouth region. Differently, variations captured by the other two solutions are much more concentrated in the mouth region of the face.

III. Expression Recognition using DSFs

Deformations due to facial expressions across 3D video sequences are characterized by subtle variations induced mainly by the motion of facial points. These subtle changes are important to perform effective expression recognition, but they are also difficult to be analyzed due to the face movements. To handle this problem, as described in the previous section, we propose a curve-based parametrization of the face that consists in representing the facial surface by a set of radial curves. According to this representation, the problem of comparing two facial surfaces, a reference facial surface and a target one, is reduced to the computation of the DSF between them.

In order to make possible to enter the expression recognition system at any time and make the recognition process possible from any frame of a given video, we consider subsequences of n frames. Thus, we chose the first n frames as the first subsequence. Then, we chose n-consecutive frames starting from

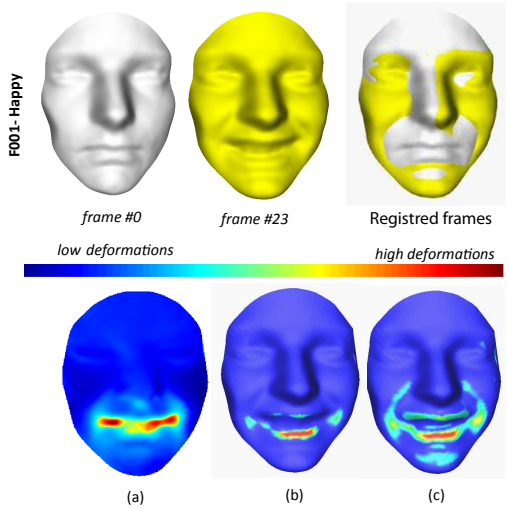


Fig. 5. Comparison of the different features extracted between two frames taken from subject 001 for the happy expression: (a) the Free Form-based Deformations (FFD); (b) the point-to-point Euclidean distances; and (c) the DSFs deformations.

the second frame as the second subsequence. This process is repeated by shifting the starting index of the sequence every one frame till the end of the sequence. In order to classify the resulting subsequences, we propose two different feature extraction and classification framework based on the DSF:

- **Mean Deformation-based features associated to Random Forest classifier**. The first frame of the subsequence is considered as a reference frame and the deformation is calculated from each of the remaining frames to the first one using the DSF. The average deformation of the n-1 resulting DSFs represents the feature vector in this classification scheme and is presented, after dimensionality reduction, to multi-class Random Forest classifiers;
- **3D Motion features combined with HMM classifiers**. The deformation between successive frames in a subsequence are calculated using the DSFs and presented to an HMM classifier preceded by LDA-based dimensionality reduction.

A. Mean Shape Deformation with Random Forest Classifier

The idea here is to capture a mean deformation of the face in the sliding window on the 3D expression sequence. In order to get this feature, the first frame of each subsequence is considered as the reference one, and the dense deformation is computed from this frame to each of the remaining frames of the subsequence. Let F_{ref} denote the reference frame of a subsequence and F_i the i-th successive frame in the subsequence; the successive frame, for example, is denoted by F_1. The DSF is calculated between F_{ref} and F_i, for different values of i ($i = 1, \ldots, n-1$), and the mean deformation is then given by:

$$\overline{DSF} = \frac{1}{n-1} \sum_{i=1}^{n-1} DSF(F_{ref}, F_i). \quad (6)$$

Figure 6 illustrates one subsequence for each expression with $n = 6$ frames. Each expression is illustrated in two rows: The upper row gives the reference frame of the subsequence and the n-1 successive frames of the subsequences. Below, the corresponding Dense Scalar Fields computed for each frame are shown. The mean deformation field is reported on the right of each plot and represents the feature vector for each subsequence. The feature vector for this subsequence is built based on the mean deformation of the n-1 calculated deformations. Thus, each subsequence is represented by a feature vector of size equal to the number of points on the face (i.e., the number of points used to sample the radial curves of the face). In order to provide a visual representation of the scalar fields, an automatic labeling scheme is applied: Warm colors (red, yellow) are associated with high $DSF(F_{ref}, F_t)$ values and correspond to facial regions affected by high deformations. Cold colors are associated with regions of the face that remain stable from one frame to another. Thus, this dense deformation field summarizes the temporal changes of the facial surface when a particular facial expression is conveyed.

According to this representation, the deformation of each subsequence is captured by the mean $\overline{DSF}$ defined in Eq. (6). The main motivation for using the mean deformation, instead of the maximum deformation for instance, is related to its greater robustness to the noise. In the practice, the mean deformation resulted more resistant to deformations due to, for example, inaccurate nose tip detection or the presence of acquisition noise. In Fig. 6, for each subsequence, the mean deformation field illustrates a smoothed pattern better than individual deformation fields in the same subsequence. Since the dimensionality of the feature vector is high, we use LDA-based transformation to map the present feature space to an optimal one that is relatively insensitive to different subjects, while preserving the discriminating expression information. LDA defines the within-class matrix S_w and the between-class matrix S_b. It transforms a n-dimensional feature to an optimized d-dimensional feature, where $d < n$. In our experiments, the discriminating classes are the 6 expressions, thus the reduced dimension d is 5.

For the classification, we used the multi-class Random Forest algorithm. The algorithm was proposed by Leo Breiman in [39] and defined as a meta-learner comprised of many individual trees. It was designed to operate quickly over large datasets and more importantly to be diverse by using random samples to build each tree in the forest. A tree achieves highly non-linear mappings by splitting the original problem into smaller ones, solvable with simple predictors. Each node in the tree consists of a test, whose result directs a data sample towards the left or the right child. During training, the tests are chosen in order to group the training data in clusters where simple models achieve good predictions. Such models are stored at the leaves, computed from the annotated data, which reached each leaf at train time. Once trained, a Random Forest is capable to classify a new expression from an input feature vector by putting it down each of the trees in the forest. Each tree gives a classification decision by voting for that class. Then, the forest chooses the classification having the most

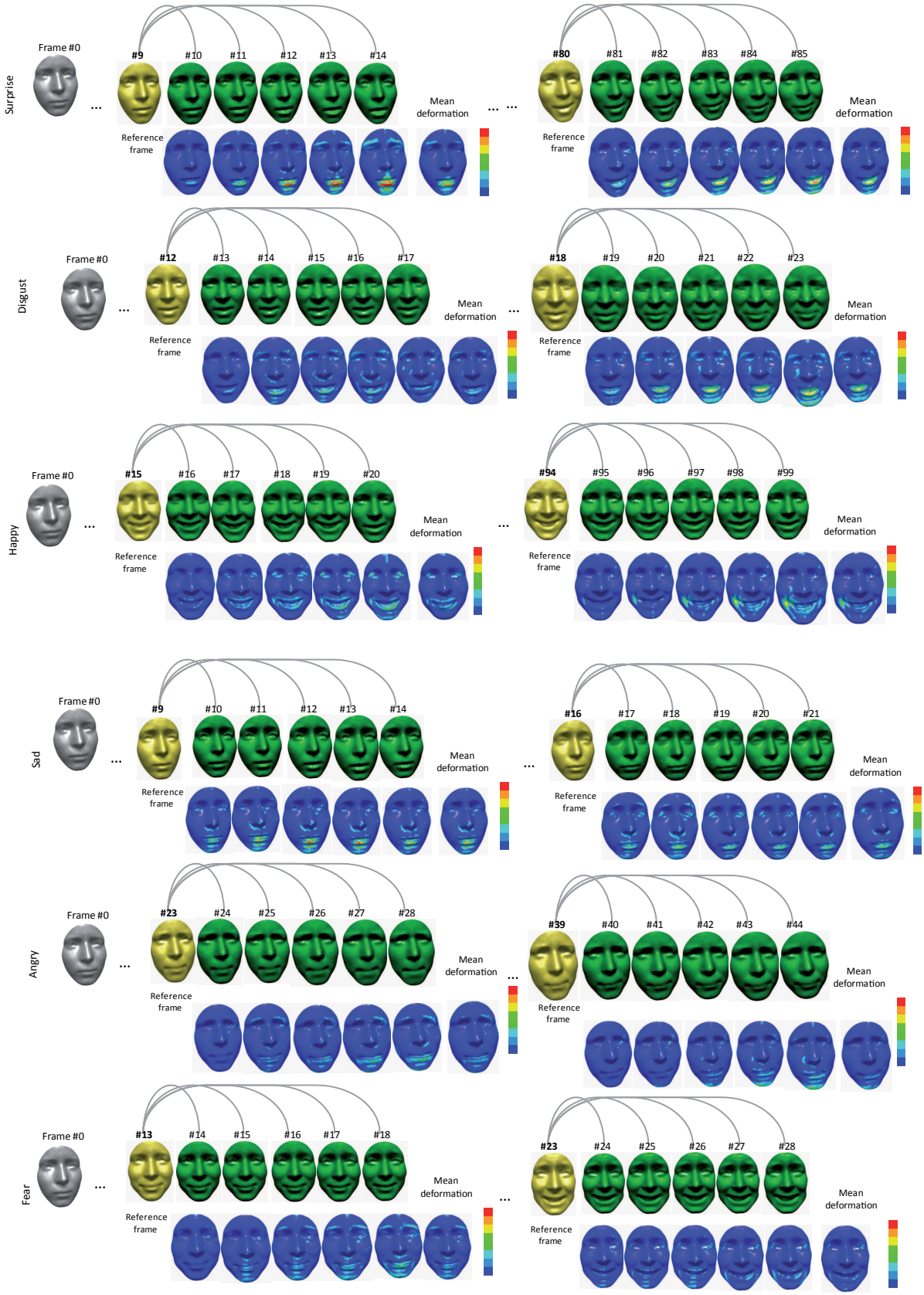


Fig. 6. Computation of dynamic shape deformation on different subsequences taken from the BU-4DFE database. Each expression is illustrated by two rows: the upper one gives the reference frame of the subsequence and the n-1 successive frames. The corresponding deformation fields computed for each frame with respect to the reference one are illustrated in the lower row. The mean deformation field is given on the right of each lower row.

votes (over all the trees in the forest).

B. 3D Motion Extraction with HMM Classifier

The DSF features described in Sect. II, can also be applied for expression recognition according to a different classification scheme. The deformations between successive frames in the subsequence are calculated using the DSF. In particular, the deformation between two successive 3D frames is obtained by computing the pairwise Dense Scalar Field $DSF(F_{t-1}, F_t)$ of correspondent radial curves. Based on this measure, we are able to quantify the motion of face points along radial curves and thus capture the changes in facial surface geometry.

Figure 7 illustrates a direct application of the $DSF(F_{t-1}, F_t)$ and its effectiveness in capturing deformation between one facial surface to another belonging to two consecutive frames in a 3D video sequence. This figure shows two subsequences extracted from videos in the BU-4DFE database (happy and surprise expressions are shown on the left and on the right, respectively). For each sequence, the 2D image and the 3D scans of some frames are shown in the upper row. In the lower row, the deformation scalar field $DSF(F_{t-1}, F_t)$ computed between consecutive frames (i.e., the current frame and the previous one) in the subsequence is reported. In order to provide a visual representation of the scalar field, an automatic labeling scheme is applied that includes only two colors: The red color is associated with high $DSF(F_{t-1}, F_t)$ values and corresponds to facial regions affected by high deformations. The blue color is associated with regions that remain more stable from one frame to another. As illustrated in Fig. 7, for different expressions, different regions are mainly deformed, showing the capability of the deformation fields to capture relevant changes of the face due to the facial expression. In particular, each deformation is expected to identify an expression, for example, as suggested by the intuition, the corners of the mouth and the cheeks are mainly deformed for the happiness expression.

With the proposed approach, the feature extraction process starts by computing for each 3D frame in a given video sequence the Dense Scalar Field with respect to the previous one. In this way, we obtain as many fields as the number of frames in the sequence (decreased by one), where each field contains as many scalar values as the number of points composing the collection of radial curves representing the facial surface. In practice, the size of $DSF(F_{t-1}, F_t)$ is 1×5000, considering 5000 points on the face, similarly to the feature vector used in the first scheme of classification (mean deformation-based). Since the dimensionality of the resulting feature vector is high, also in this case we use LDA to project the scalar values to a 5-dimensional feature space, which is sensitive to the deformations induced by different expressions. The 5-dimensional *feature vector* extracted for the 3D frame at instant t of a sequence is indicated as f^t in the following. Once extracted, the feature vectors are used to train HMMs and to learn deformations due to expressions along a temporal sequence of frames.

In our case, sequences of 3D frames constitute the temporal dynamics to be classified, and each prototypical expression is modeled by an HMM (a total of 6 HMMs λ^{expr} is required, with $expr \in \{an, di, fe, ha, sa, su\}$). Four states per HMM are used to represent the temporal behavior of each expression. This corresponds to the idea that each sequence starts and ends with a neutral expression (state S_1). The frames that belong to the temporal intervals where the face changes from neutral to expressive and back from expressive to neutral are modeled by the *onset* (S_2) and *offset* (S_4) states, respectively. Finally, the frames corresponding to the highest intensity of the expression are captured by the apex state (S_3). This solution has proved its effectiveness in clustering the expressive states of a sequence also in other works [33]. Figure 8 exemplifies the structure of the HMMs used in our framework.

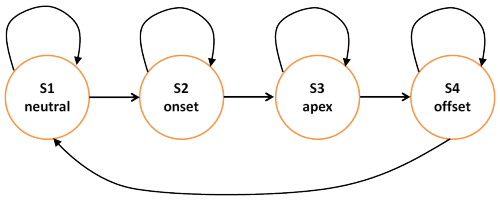


Fig. 8. Structure of the HMMs modeling a 3D facial sequence. The four states model, respectively, the *neutral*, *onset*, *apex* and *offset* frames of the sequence. As shown, from each state it is possible to remain in the state itself or move to the next one (this is known as *Bakis* or left-right HMM).

The training procedure of each HMM is summarized as follows:

- Feature vectors f^t of the training sequences are first clustered to identify a *codebook* of symbols using the standard LBG algorithm [40]. This provides the required mapping between multidimensional feature vectors, taking values in a continuous domain, with the alphabet of symbols emitted by the HMM states;
- Expression sequences are considered as observation sequences $O = \{O^1, O^2, \ldots, O^T\}$, where each observation O^t at time t is given by the feature vector f^t;
- Each HMM λ^{expr} is initialized with random values and the *Baum-Welch* algorithm [41] is used to train the model using a set of training sequences. This estimates the model parameters, while maximizing the conditional probability $P(O|\lambda^{expr})$.

Given a 3D sequence to be classified, it is processed as in Sect. II, so that each feature vectors f^t corresponds to a *test* observation $O = \{O^1 \equiv f^1, \ldots, O^T \equiv f^T\}$. Then, the test observation O is presented to the six HMMs λ^{expr} that model different expressions, and the *Viterbi* algorithm is used to determine the best *path* $\bar{Q} = \{\bar{q}^1, \ldots, \bar{q}^T\}$, which corresponds to the state sequence giving a maximum of likelihood to the observation sequence O. The likelihood along the best path, $p^{expr}(O, \bar{Q}|\lambda^{expr}) = \bar{p}^{expr}(O|\lambda^{expr})$ is considered as a good approximation of the true likelihood given by the more expensive *forward* procedure [41], where all the possible paths are considered instead of the best one. Finally, the sequence is classified as belonging to the class corresponding to the HMM whose log-likelihood along the best path is the greatest one.

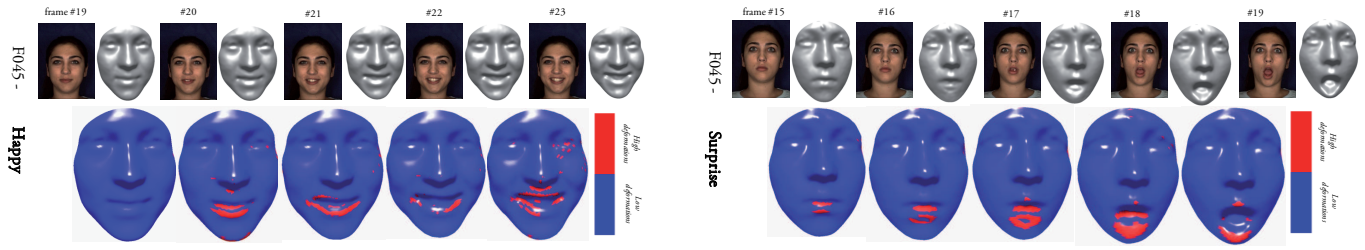


Fig. 7. Examples of DSF deformations between subsequent frames of 3D video sequences: Happy and surprise expressions are shown, respectively, on the left and right.

IV. Experimental Results

The proposed framework for facial expression recognition from dynamic sequences of 3D face scans has been experimented on the BU-4DFE database. Main characteristics of the database and results are reported in the following sections.

A. BU-4DFE Database: Description and Preprocessing

To investigate the usability and performance of 3D dynamic facial sequences for facial expression recognition, a dynamic 3D facial expression database has been created at *Binghamton University* [14]. The Dimensional Imaging's 3D dynamic capturing system [42], has been used to capture a sequence of stereo images and produce the depth map of the face according to a passive stereo-photogrammetry approach. The range maps are then combined to produce a temporally varying sequence of high-resolution 3D images with an RMS accuracy of 0.2mm. At the same time, 2D texture videos of the dynamic 3D models are also recorded. Each participant (subject) was requested to perform the six prototypical expressions (i.e., *angry*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*) separately. Each expression sequence contains neutral expressions in the beginning and the end, so that each expression was performed gradually from neutral appearance, low intensity, high intensity, and back to low intensity and neutral. Each 3D sequence captures one expression at a rate of 25 frames per second and each 3D sequence lasts approximately 4 seconds with about 35,000 vertices per scan (i.e., 3D *frame*). The database consists of 101 subjects (58 female and 43 male, with an age range of 18-45 years old) including 606 3D model sequences with 6 prototypical expressions and a variety of ethnic/racial ancestries (i.e., 28 Asian, 8 African-American, 3 Hispanic/Latino, and 62 Caucasian). More details on the BU-4DFE can be found in [14]. An example of a 3D dynamic facial sequence of a subject with "happy" expression is shown in Fig. 9, where 2D frames (not used in our solution) and 3D frames are reported. From left to right, the frames illustrate the intensity of facial expression passing from *neutral* to *onset*, *offset*, *apex* and *neutral* again.

It can be observed that the 3D frames present a near-frontal pose with some slight changes occurring mainly in the azimuthal plane. The scans are affected by large outliers, mainly acquired in the hair, neck and shoulders regions (see Fig. 9). In order to remove these imperfections from each 3D frame a preprocessing pipeline is performed. The main steps of this pipeline are summarized as follows (see also Fig. 10):

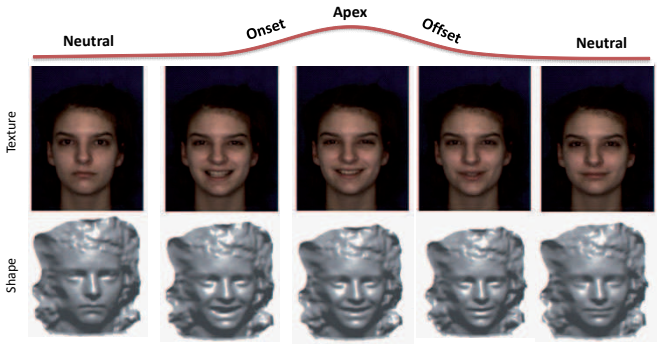


Fig. 9. Examples of 2D and 3D frames extracted from a dynamic 3D video sequence of the BU-4DFE dataset.

- Identify and fill the holes caused, in general, by self-occlusions or open mouth. The holes are identified by locating boundary edges, then triangulating them;
- Detect the nose tip on the face scan in the first frame of the sequence. The nose tip is detected by analyzing the peak point of the face scan in the depth direction. The nose tip is then tracked on all the subsequent frames when the search area is limited to a specific neighborhood around the nose tip detected on the first frame;
- Crop the facial area using a sphere centered on the detected nose tip with a constant radius set to 90mm based on some observations;
- Normalize the pose of a given frame according to its previous frame using the Iterative Closest Point (ICP)-based alignment. We point out that our implementation uses the following parameters to perform the ICP algorithm: (i) Match the nose tips of the faces first; (ii) Number of vertices considered to find the optimal tranformation=50; and (iii) Number of iterations=5. In addition to permit effective alignment, this set of parameters allows also an attractive computational cost.

In a real-world scenario of use, the head can move freely and rotate, whereas in our experiments only near-frontal faces are considered, as the BU-4DFE database does not contain non-frontal acquisitions. To account for the capability of our approach to cope with non-frontal scans, we report in Fig. 11 some registration results when applying an artificial rotation to one of the 3D faces to be aligned. It is clear that the registration method is able to handle with moderate pose variations (up to about 30/45 degrees). Instead, the registration method is not

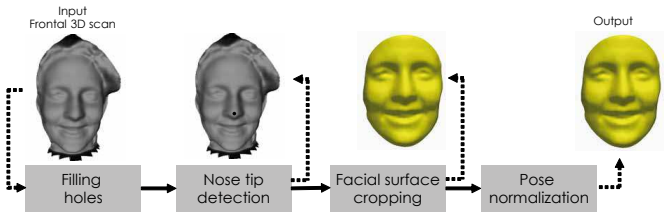


Fig. 10. Outline of the preprocessing steps. A pipeline of four filters is applied to each 3D sequence: (i) Filling holes, if any; (ii) Nose tip detection (for the first frame) and tracking (for remaining frames); (iii) Face cropping using a sphere centered on the nose tip and of radius 90mm; (iv) Pose normalization based on the ICP algorithm.

able to register a frontal face with a profile face (right side of the figure).

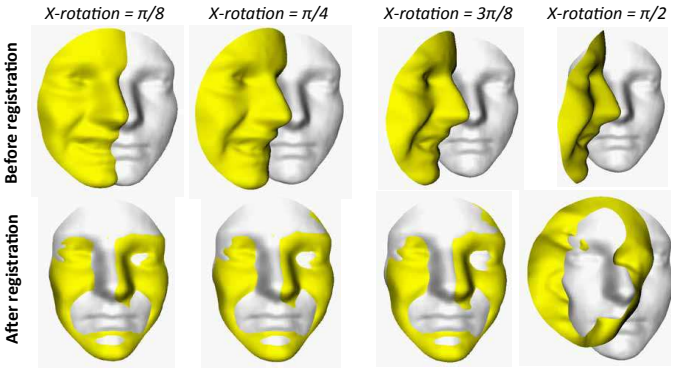


Fig. 11. Registration results using the ICP algorithm when rotating around the X-axis one of the 3D preprocessed faces. The first row shows the initial rotation applied on the yellow model (before the alignment) and the second row shows the alignment output (after alignment).

In the proposed framework, after preprocessing and Dense Scalar Fields computation across the 3D sequences, we designed two deformation learning and classification schemes. The first scheme consists on averaging, within a sliding window, the DSF computed on each frame with respect to the first frame of the window. This produces dense deformations across the sliding windows that are learned using a Multi-class Random Forest algorithm (see Sect. III-A). The second scheme consists on a dynamic analysis through the 3D sequences using conventional temporal HMMs-modeling. Here, the 3D motion (deformation) is extracted and then learned for each class of expression, as described in Sect. III-B. In both the cases, a common experimental set up has been used. In particular, data of 60 subjects have been selected to perform recognition experiments according to the evaluation protocol followed in other works [23], [26], [27]. The 60 subjects have been randomly partitioned into 10 sets, each containing 6 subjects, and 10-fold cross validation has been used for training/test, where at each round 9 of the 10 folds (54 subjects) are used for the training, while the remaining fold (6 subjects) is used for the test. In the following, we report experimental evaluation and comparative analysis with respect to previous studies.

B. Mean deformation-based Expression Classification

Following the experimental protocol proposed in [23], a large set of subsequences are extracted from the original expression sequences using a sliding window. The subsequences have been defined in [23] with a length of 6 frames with a sliding step of one frame from one subsequence to the following one. For example, with this approach, a sequence of 100 frames originates a set of $6 \times 95 = 570$ subsequences, each subsequence differing for one frame from the previous one. Each sequence is labelled to be one of the six basic expressions, thus extracted subsequences have the same label. This accounts for the fact that, in general, the subjects can enter the system not necessarily starting with a neutral expression, but with an arbitrary expression. The classification of these short sequences is regarded as an indication of the capability of the expression recognition framework to identify individual expressions. According to this, we first compute for each subsequence the Mean Deformation, which is then presented to multi-class Random Forest, as outlined in Sect. III.

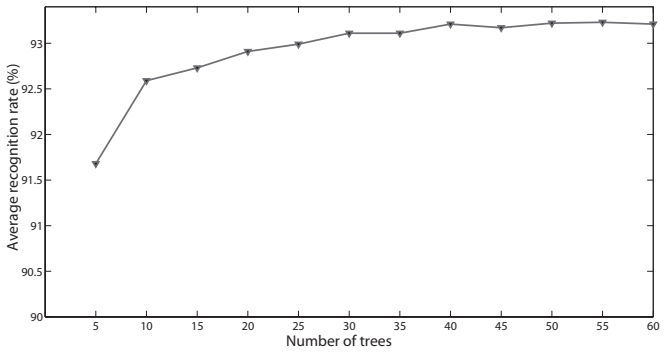


Fig. 12. 4D expression recognition results using a Random Forest classifier when varying the number of trees.

The performance of Random Forest classifier varies with the number of trees. Thus, we perform the experiments with different numbers of trees; the results of this experimentation is shown in Fig. 12. As illustrated in this figure, the average recognition rate raises with the increasing number of trees until 40, when the recognition rate reaches 93.21%, and then becomes quite stable. Thus, in the following we consider 40 trees and we report detailed results (confusion matrix) with this number of trees in Tab. I. We recall that the rates are obtained by averaging the results of the 10-independent runs (10-fold cross validation). It can be noted that the largest confusions are between the *disgust (Di)* expression and the *angry (An)* and *Fear (Fe)* expressions. Interestingly, these three expressions capture negative emotive states of the subjects, so that similar facial muscles can be activated. The best classified expressions are *happy (Ha)* and *Surprise (Su)* with recognition accuracy of 95.47% and 94.53%, respectively. The standard deviation from the average performance is also reported in the table. The value of this statistical indicator suggests that small variations are observed between different folds.

Effect of the Subsequence Size: We have also conducted additional experiments when varying the temporal size of the

TABLE I
CONFUSION MATRIX FOR MEAN DEFORMATION AND RANDOM FOREST CLASSIFIER (FOR 6-FRAMES WINDOW).

%	*An*	*Di*	*Fe*	*Ha*	*Sa*	*Su*
An	**93.11**	2.42	1.71	0.46	1.61	0.66
Di	2.3	**92.46**	2.44	0.92	1.27	0.58
Fe	1.89	1.75	**91.24**	1.5	1.76	1.83
Ha	0.57	0.84	1.71	**95.47**	0.77	0.62
Sa	1.7	1.52	2.01	1.09	**92.46**	1.19
Su	0.71	0.85	1.84	0.72	1.33	**94.53**
Average recognition rate **=93.21±0.81%**						

sliding window used to define the subsequences. In Fig. 13, we report results for a window size equal to 2, 5 and 6, and using the whole length of the sequence (on average this is about 100 frames). From the figure, it clearly emerges that the recognition rate of the six expressions increases when increasing the temporal length of the window. This reveals the importance of the temporal dynamics and shows that the spatio-temporal analysis outperforms a spatial analysis of the frames. By considering the whole sequences for the classification, the result reach 100%. In the paper, we decided to report detailed results when considering a window length of 6-frames to allow comparisons with previous studies.

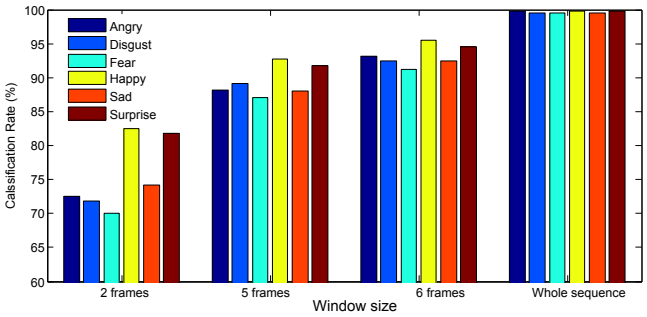


Fig. 13. Effect of the temporal size of the sliding window on the results. The classification rates increase when increasing the length of the temporal window.

Effect of the Spatial Resolution of 3D Faces: In the proposed face representation, the DSF is computed for the points of a set of radial curves originating from the nose tip. Due to this, the density of the scalar field depends on the number of radial curves and the number of points per curve. So, the resolution used for the number of curves and points per curve can affect the final effectiveness of the representation. To investigated this aspect, we have conducted experiments when varying the spatial resolution of the 3D faces (i.e., the number of radial curves and the number of points per curve). Figure 14 expresses quantitatively the relationship between the expression classification accuracy (on the BU-4DFE) and the number of radial curves and the number of points per curve. This can give an indication of the expected decrease in the performance in the case the number of radial curves or points per curve is decreased due to the presence of noise and spikes in the data. From these results, we can also observe that the resolution in terms of number of curves has more importance than the resolution in terms of points per curve.

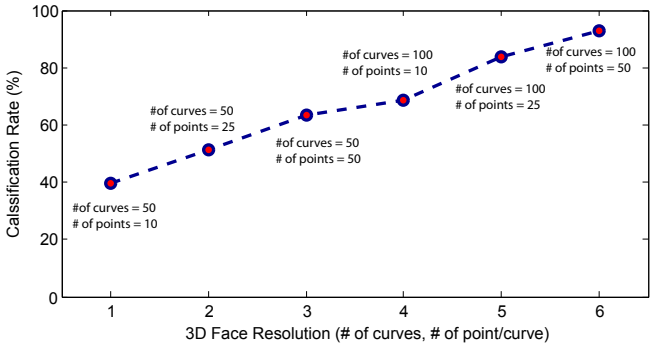


Fig. 14. Effects of varying the 3D face resolution on the classification results.

C. HMM-based Expression Classification

Following the same setup as in previous section (originally defined in [23]), for this experiment we trained the HMMs on 6 frames subsequences constructed as discussed above. The 4-state structure of the HMMs resulted adequate to model the subsequences. Also in this experiment, we performed 10-folds cross validation on the overall set of subsequences derived from the 60×6 sequences (31970 in total). The achieved results by classifying individual subsequences of the expression sequences (*frame-by-frame* experiment) are reported in the confusion matrix of Tab. II. Values in the table have been obtained by using features of 6-frames subsequences as input to the 6 HMMs and using the maximum emission probability criterion as decision rule. It is clear that the proposed approach is capable to accurately classify individual frames by analyzing the corresponding subsequence of previous 5 frames. The average recognition rate is equal to 93.83%, slightly higher than the one displayed by Mean Deformation plus Random Forest classification schema (though the standard deviation among different folds shows a greater value in this case). It can also be noted that, compared to the previous classifier, the same tendency of recognition rates is in general achieved. In fact, correct classification of *angry* is high despite the difficulty of this expression analysis. This learning scheme achieved better recognition than the first one for *angry (An)* expression. Actually, whereas the *angry (An)* expression is known for its subtle motions, our classifier achieved 93.95% of correct classification, which demonstrates the ability of the proposed DSF to capture subtle deformations across the 3D sequences. These similar good achievements are mainly the effect of the proposed deformation scalar field.

Comparion with the FFD feature: The proposed framework can also fit with different deformation fields than the proposed DSF. So, considering alternative features to densely capture the deformation fields on the lattice of points of the radial curves of the face can permit a direct comparison of our DSF feature with different ones. In particular, we considered the Free-Form Deformation (FFD) [38] feature, which is a standard method for non-rigid registration and has been successfully proved in the context of expression recognition [25] (see also Sect. II.B).

TABLE II
CONFUSION MATRIX FOR MOTION EXTRACTION AND HMM CLASSIFIERS (FOR 6-FRAMES WINDOW).

%	*An*	*Di*	*Fe*	*Ha*	*Sa*	*Su*
An	**93.95**	1.44	1.79	0.28	2.0	0.54
Di	3.10	**91.54**	3.40	0.54	1.27	0.15
Fe	1.05	1.42	**94.55**	0.69	1.67	0.62
Ha	0.51	0.93	1.65	**94.58**	1.93	0.40
Sa	1.77	0.48	1.99	0.32	**94.84**	0.60
Su	0.57	0.38	3.25	0.38	1.85	**93.57**
Average recognition rate = ***93.83±1.53%***						

Table III reports the confusion matrix obtained by posing FFD in our classification framework, using the same setting as in the experiments above (i.e., 100 radial curves, with 50 sampled points each, and LDA reduction of the deformation field from a 5000-dimensional vector to a 5-dimensional subspace). The overall result is that the proposed DSF feature provides a finer discriminative capability with respect to FFD, thus resulting in a better classification accuracy. This can be motivated by the nice invariant properties of the proposed Riemannian framework (as discussed in Sect. II).

TABLE III
CONFUSION MATRIX FOR FREE-FORM DEFORMATION (FFD) AND HMM CLASSIFIERS (FOR 6-FRAMES WINDOW).

%	*An*	*Di*	*Fe*	*Ha*	*Sa*	*Su*
An	**78.45**	4.51	5.72	1.97	6.49	2.86
Di	8.63	**76.1**	6.65	2.82	4.18	1.62
Fe	3.1	5.5	**80.23**	1.99	6.31	2.87
Ha	1.43	2.02	3.77	**86.12**	5.31	1.35
Sa	5.79	1.4	4.99	0.86	**85.83**	1.13
Su	1.73	2.04	6.13	1.55	3.9	**84.65**
Average recognition rate = 81.9 ± 2.35**%**						

D. Discussion and Comparative Evaluation

To the best of our knowledge, the works reporting results on expression recognition from dynamic sequences of 3D scans are those in [20], [24], [26], [29], and recently [30]. These works have been evaluated on the BU-4DFE dataset, but the testing protocols used in the experiments are sometimes different, so that a direct comparison of the results reported in these papers is not immediate. In the following, we discuss these solutions with respect to our proposal, also evidencing the different settings under which the expression recognition results are obtained.

Table IV summarizes approaches and results reported previously on the BU-4DFE dataset, compared to those obtained in this work. The testing protocols used in the experiments are quite different especially the number of verified expressions, all the six basic expressions in [23], [24], [27], [29], and [30] whereas [25], [26] reported primary results on only three expressions. The number of subjects considered is 60, except in [25] where the number of subjects is not specified. In general, sequences in which the required expressions are acted accurately are selected, whereas in [27] and [29] 507 sequences out of the 606 total are used for all subjects. In our experiments, we conducted tests by following the same setting proposed by the earliest and more complete evaluation described in [23]. The training and the testing sets were constructed by generating subsequences of 6-frames from all sequences of 60 selected subjects. The process were repeated by shifting the starting index of the sequence every one frame till the end of the sequence.

We note that the proposed approaches outperforms state-of-the-art solutions following similar experimental settings. The recognition rates reported in [23] and [24] based on temporal analysis only was 80.04% and spatio-temporal analysis was 90.44%. In both studies subsequences of constant window width including 6-frames ($win = 6$) is defined for experiments. We emphasize that their approach is not completely automatic requiring 83 manually annotated landmarks on the first frame of the sequence to allow accurate model tracking.

The method proposed in [25] and [20] is fully automatic with respect to the processing of facial frames in the temporal sequences, but uses *supervised* learning to annotate individual frames of the sequence in order to train a set of HMMs. Though performed off-line, supervised learning requires manual annotation and counting on a consistent number of training sequences that can be a time consuming operation. In addition, a drawback of this solution is the computational cost due to Free-Form Deformations based on B-spline interpolation between a lattice of control points for nonrigid registration and motion capturing between frames. Preliminary tests were reported on three expressions: *(An)*, *(Ha)* and *(Su)*. Authors motivated the choice of the happiness and anger expressions with the fact that they are at either ends of the valence expression spectrum, whereas surprise was also chosen as it is at one extreme of the arousal expression spectrum. However, these experiments were carried out on a subset of subjects accurately selected as acting out the required expression. Verification of the classification system was performed using a 10-fold cross-validation testing. On this subset of expressions and subjects, an average expression recognition rate of 81.93% is reported. In [20], the same authors have reported 64.6% classification rate when in their evaluation they consider all the six basic expressions.

In [26] a fully automatic method is also proposed, that uses an *unsupervised* learning solution to train a set of HMMs (i.e., annotation of individual frames is not required in this case). Expression recognition is performed on 60 subjects from the BU-4DFE database for the expressions of *happiness*, *sadness* and *surprise*. The recognition accuracy averaged on 10 rounds of 10-fold cross-validation show an overall value of 92.22%, with the highest performance of 95% obtained for the happiness expression. However, the authors reported recognition results on whole facial sequences, but this hinders the possibility of the methods to adhere to a real-time protocol. In fact, reported recognition results depends on the preprocessing of whole sequences unlike our approach and the one described in [23], which are able to provide recognition results when processing very few 3D frames.

In [27] and [29], results are presented for expression recognition accuracy on 100 subjects picked out from BU-4DFE

TABLE IV
COMPARISON OF THIS WORK TO EARLIER STUDIES. PROTOCOL DESCRIPTION: #SUBJECTS (S), #EXPRESSIONS (E), WIN SIZE (WIN). T: TEMPORAL ONLY/S-T: SPATIO-TEMPORAL. ACCURACY ON SLIDING WINDOW/WHOLE SEQUENCE (OR SUBSEQUENCE).

Authors	Method	Features	Classification	Protocol	T/S-T	RR (%)
Sun et al. [23]	MU-3D	12 Motion Units	HMM	60 S, 6 E, Win=6	T	70.31, —
Sun et al. [23]	T-HMM	Tracking model	HMM	60 S, 6 E, Win=6	T	80.04, —
Sun et al. [23]	P2D-HMM	Curvature+Tracking model	T-HMM+S-HMM	60 S, 6 E, Win=6	S-T	82.19, —
Sun et al. [23]	R-2DHMM	Curvature+Tracking model	2D-HMM	60 S, 6 E, Win=6	S-T	90.44, —
Sandbach et al. [25]	3D Motion-based	FFD+Quad-tree	GentleBoost+HMM	—, 3 E, Win=4	T	73.61, 81.93
Sandbach et al. [20]	3D Motion-based	FFD+Quad-tree	GentleBoost+HMM	—, 6 E, variable Win	T	64.6, —
Le et al. [26]	Level curve-based	pair- and segment-wise distances	HMM	60 S, 3 E, —	S-T	—, 92.22
Fang et al. [27], [29]	AFM Fitting	LBP-TOP	SVM-RBF	100 S, 6 E, —	T	—, 74.63
Fang et al. [27], [29]	AFM Fitting	LBP-TOP	SVM-RBF	100 S, 3 E, —	T	—, 96.71
Reale et al. [30]	Spatio-temporal volume	"Nebula" Feature	SVM-RBF	100 S, 6 E, Win=15	S-T	—, 76.10
This work	Geometric Motion Extraction	3D Motion Extraction	LDA-HMM	60 S, 6 E, Win=6	T	93.83, —
This work	Geometric Mean Deformation	Mean Deformation	LDA-Random Forest	60 S, 6 E, Win=6	T	93.21, —

database. However, 507 sequences are selected manually according to the following criteria: (1) the 4D sequence should start by neutral expression, and (2) sequences containing corrupted meshes are discarded. In addition, to achieve recognition rate of 75.82%, whole sequences should be analyzed. The authors reported highest recognition rates when only (Ha), (An), and (Su) expressions (96.71%) or (Ha), (Sa) and (Su) (95.75%) are considered.

The protocol used in [30] is quite different from the others. First, the onset frame for each of the six canonical expressions has been marked manually on each sequence of the BU-4DFE database. Then, a fixed size window of 15 frames starting from the onset frame has been extracted from each expression of 100 subjects. So, although sequences from 100 subjects are used by this approach, it also uses a manual intervention to detect the onset frame and just 15 frames from the onset one are used for the classification (and these typically correspond to the most intense expression, including the apex frames).

According to this comparative analysis, the proposed framework compares favorably with state-of-the-art solutions. It consists of two geometric deformation learning schemes with a common feature extraction module (DSF). This demonstrates the effectiveness of the novel mathematical representation called Dense Scalar Field (DSF), under the two designed schemes.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an automatic approach for identity-independent facial expression recognition from 3D video sequences. Through a facial shapes representation by collections of radial curves, a Riemannian shape analysis was applied to quantify dense deformations and extract motion from successive 3D frames. Two different classification schema were performed, a HMM-based classifier and a Mean deformation-based classifier. An LDA-based transformation was applied to decrease the dimensionality of the resulting feature vectors. The proposed approach outperforms previous ones; it is capable to accurately classify short sequences containing very different 3D frames, with an average accuracy of 93.83% using HMM-based classifier and 93.21% using mean deformation-based classifier, following state-of-the-art settings on the BU-4DFE dataset. This emphasizes the capability of the proposed geometric shape deformation analysis to capture subtle deformations in 4D videos.

One limitation of the approach is the nose tip detection in case of non-frontal views and/or the presence of occlusions (by glasses, hand, hair, etc.). The BU-4DFE database contains frontal 3D faces without occlusion, however, in realistic scenarios, more elaborated techniques should be applied to detect the nose tip. As future perspectives of the presented work are, first, its extension to spatio-temporal analysis by introducing a spatial (intra-frame) analysis beside the temporal analysis (inter-frame). Second, its adaptation to low resolution 3D videos, outputs of the depth-consumer cameras like the Kinect, is a distant goal mainly due the presence of large noise and the low resolution of the acquired scans.

ACKNOWLEDGMENTS

This work was supported by the French research agency ANR through the 3D Face Analyzer project under the contract ANR 2010 INTB 0301 01.

REFERENCES

[1] A. Mehrabian and M. Wiener, "Decoding of inconsistent communications," Journal of Personality and Social Psychology, vol. 6, no. 1, p. 109?114, May 1967.

[2] P. Ekman, "Universals and cultural differences in facial expressions of emotion," in Proc. Nebraska Symposium on Motivation, vol. 19, Lincoln, NE, 1972, pp. 207–283.

[3] P. Ekman and W. V. Friesen, Manual for the the Facial Action Coding System. Palo Alto, CA: Consulting Psychologist Press, 1977.

[4] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[5] I. A. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable approach," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 640–649, April 2007.

[6] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," Int. Journal of Computer Vision, vol. 79, no. 1, pp. 1–12, Aug. 2008.

[7] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An intrinsic framework for analysis of facial surfaces," Int. Journal of Computer Vision, vol. 82, no. 1, pp. 80–95, April 2009.

[8] S. Berretti, A. Del Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 12, pp. 2162–2177, Dec. 2010.

[9] Y. Wang, J. Liu, and X. Tang, "Robust 3D face recognition by local shape difference boosting," IEEE Transactions on Pattern Analysis and Machine Untelligence, vol. 32, no. 10, pp. 1858–1870, Oct. 2010.

[10] A. Maalej, B. Ben Amor, M. Daoudi, A. Srivastava, and S. Berretti, "Shape analysis of local facial patches for 3D facial expression recognition," Pattern Recognition, vol. 44, no. 8, pp. 1581–1589, Aug. 2011.

[11] S. Berretti, B. Ben Amor, M. Daoudi, and A. del Bimbo, "3D facial expression recognition using sift descriptors of automatically detected keypoints," The Visual Computer, vol. 27, no. 11, pp. 1021–1036, Nov. 2011.

[12] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D facial expression database for facial behavior research," in Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition, Southampton, UK, Apr. 2006, pp. 211–216.

[13] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in Proc. First COST 2101 Workshop on Biometrics and Identity Management, May 2008.

[14] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," in Proc. Int. Conf. on Automatic Face and Gesture Recognition (FGR08), Amsterdam, The Netherlands, Sep. 2008, pp. 1–6.

[15] B. Matuszewski, W. Quan, and L.-K. Shark, "High-resolution comprehensive 3-D dynamic database for facial articulation analysis," in Proc. IEEE Int. Conf. on Computer Vision Workshops, Barcelona, Spain, Nov. 2011, pp. 2128–2135.

[16] B. J. Matuszewski, W. Quan, L.-K. Shark, A. S. McLoughlin, C. E. Lightbody, H. C. Emsley, and C. L. Watkins, "Hi4D-ADSIP 3-D dynamic facial articulation database," Image and Vision Computing, vol. 30, no. 10, pp. 713–727, 2012 2012.

[17] D. Cosker, E. Krumhuber, and A. Hilton, "A FACS valid 3d dynamic action unit database with applications to 3D dynamic morphable facial modeling," in Proc. IEEE Int. Conf, on Computer Vision, Barcelona, Spain, Nov. 2011, pp. 2296–2303.

[18] S. Berretti, A. Del Bimbo, and P. Pala, "Superfaces: A super-resolution model for 3D faces," in Proc. of Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment, Florence, Italy, Oct. 2012, pp. 73–82.

[19] Y. Li, A. Mian, W. Lu, and A. Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in Proc. IEEE Workshop on Applications of Computer Vision, Tampa, FL, Jan. 2013, pp. 186–192.

[20] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," Image and Vision Computing, vol. 30, no. 10, pp. 683–697, 2012.

[21] L. Benedikt, V. Kajic, D. Cosker, P. Rosin, and D. Marshall, "Facial dynamics in biometric identification," in Proc. British Machine Vision Conf., Leeds, UK, Sep. 2008, pp. 1–10.

[22] L. Benedikt, D. Cosker, P. L. Rosin, and D. Marshall, "Assessing the uniqueness and permanence of facial actions for use in biometric applications," IEEE Transactions on Systems, Man and Cybernetics - Part A, vol. 40, no. 3, pp. 449–460, May 2010.

[23] Y. Sun and L. Yin, "Facial expression recognition based on 3D dynamic range model sequences," in Proc. Eur. Conf. on Computer Vision, Marseille, France, Oct. 2008, pp. 58–71.

[24] Y. Sun, X. Chen, M. J. Rosato, and L. Yin, "Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis," IEEE Transactions on Systems, Man, and Cybernetics, Part A, vol. 40, no. 3, pp. 461–474, 2010.

[25] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "A dynamic approach to the recognition of 3D facial expressions and their temporal models," in Proc. IEEE Conf. on Automatic Face and Gesture Recognition, Santa Barbara, CA, Mar. 2011, pp. 406–413.

[26] V. Le, H. Tang, and T. S. Huang, "Expression recognition from 3D dynamic faces using robust spatio-temporal shape features," in IEEE Conference on Automatic Face and Gesture Recognition, Santa Barbara, CA, Mar. 2011, pp. 414–421.

[27] T. Fang, X. Zhao, S. Shah, and I. Kakadiaris, "4D facial expression recognition," in Proc. IEEE Int. Conf. on Computer Vision Workshop, Barcelona, Spain, Nov. 2011, pp. 1594–1601.

[28] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[29] T. Fang, X. Zhao, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "3d/4d facial expression analysis: An advanced annotated face model approach," Image and Vision Computing, vol. 30, no. 10, pp. 738 – 749, 2012, ¡ce:title¿3D Facial Behaviour Analysis and Understanding¡/ce:title¿.

[30] M. Reale, X. Zhang, and L. Yin, "Nebula feature: A space-time feature for posed and spontaneous 4d facial behavior analysis," 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 0, pp. 1–8, 2013.

[31] X. Zhang, L. Yin, J. F. Cohn, S. J. Canavan, M. Reale, A. Horowitz, and P. Liu, "A high-resolution spontaneous 3d dynamic facial expression database," in FG, 2013, pp. 1–6.

[32] H. Drira, B. Ben Amor, M. Daoudi, A. Srivastava, and S. Berretti, "3D dynamic expression recognition based on a novel deformation vector field and random forest," in Proc. Int. Conf. on Pattern Recognition, Tsukuba, Japan, Nov. 2012, pp. 1104–1107.

[33] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "Recognition of 3D facial expression dynamics," Image and Vision Computing, vol. 30, no. 10, pp. 762–773, 2012.

[34] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid registration using free-form deformations:application to breast mr images." IEEE Transactions on medical imaging, vol. 18, no. 8, pp. 712–721, 1999.

[35] H. Drira, B. Ben Amor, M. Daoudi, and A. Srivastava, "Pose and expression-invariant 3D face recognition using elastic radial curves," in Proc. British Machine Vision Conference, Aberystwyth, UK, August 2010, pp. 1–11.

[36] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn, "Shape analysis of elastic curves in euclidean spaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 7, pp. 1415–1428, 2011.

[37] S. Joshi, E. Klassen, A. Srivastava, and I. Jermyn, "A novel representation for Riemannian analysis of elastic curves in $\mathbb{R}^n$," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Minneapolis, MN, Jun. 2007, pp. 1063–6919.

[38] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," IEEE Transactions on Medical Imaging, vol. 18, no. 8, p. 712721, 1999.

[39] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[40] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," IEEE Transactions on Communications, vol. 28, no. 1, pp. 84–94, Jan. 1980.

[41] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," Proceedings of IEEE, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[42] Di3D, "http://www.di3d.com," 2006.



Boulbaba Ben Amor received the MS degree in 2003 and the PhD degree in Computer Science in 2006, both from Ecole Centrale de Lyon, France. Before that, he obtained the engineer degree in computer science from ENIS, Tunisia, in 2002. He jointed the Mines-Tlcom Institute as Associate Professor, in 2007. Since then, he is also 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. 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. He is now visiting researcher at the Florida State University (FSU) for the period January-August, 2014.



Hassen Drira is an assistant Professor of Computer Science at Institut Mines-Télécom Lille/Télécom Lille, LIFL UMR (CNRS 8022) since September 2012. He received his engineering degree in 2006 and his M.Sc. degrees Computer Science in 2007 from National School of Computer Science (ENSI), Manouba, Tunisia. He obtained his Ph.D degree in Computer Science in 2011, from University of Lille 1, France. He spent the year 2011-2012 in the MIIRE research group within the Fundamental Computer Science Laboratory of Lille (LIFL) as a Post-Doc. His research interests are mainly focused on pattern recognition, statistical analysis, 3D face recognition, biometrics and more recently 3D facial expression recognition. He has published several refereed journal and conference articles in these areas.



Stefano Berretti received the Ph.D. in Information and Telecommunications Engineering in 2001 from University of Florence, Italy. Currently, he is an Associate Professor at Department of Information Engineering and at Media Integration and Communication Center of University of Florence, Italy. His research interests are mainly focused on content modeling, retrieval, and indexing of image and 3D object databases. Recent researches have addressed 3D object retrieval and partitioning, 3D face recognition, 3D and 4D facial expression recognition. He has been visiting researcher at the Indian Institute of Technology (IIT), in Mumbai, India (2000), and visiting professor at the Institute TELECOM, TELECOM Lille 1, in Lille, France (2009), and at the Khalifa University, Sharjah, UAE (2013). Stefano Berretti is author of more than 100 papers appeared in conference proceedings and international journals in the area of pattern recognition, computer vision and multimedia. He is in the program committee of several international conferences and serves as a frequent reviewer of many international journals. He has been co-chair of the Fifth Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA 2012), held in conjunction with ECCV 2012.



Mohamed Daoudi Mohamed Daoudi is a Professor of Computer Science at Télécom Lille and LIFL (UMR CNRS 8022). He is the head of Computer Science department at Telecom Lille. He received his Ph.D. degree in Computer Engineering from the University of Lille 1, France, in 1993 and Habilitation from the University of Littoral, France, in 2000. He was the founder and the scientific leader of MIIRE research group http://www-rech.telecom-lille.fr/miire/. His research interests include pattern recognition, image processing, three-dimensional analysis and retrieval and 3D face analysis and recognition. He has published over 140 papers in some of the most distinguished scientific journals and international conferences. He was in the program committee of several international conferences. He has been co-chair of several workshops including 3D Objects Retrieval ACM Workshop (ACM MM 2010), 3D Object Retrieval Eurographics Workshop (EG 2010) and 3D Face Biometrics (IEEE FG 2013). He is the author of several books, including 3D processing: Compression, Indexing and Watermarking (Wiley, 2008) and 3D Face Modeling, Analysis and Recognition (Wiley, 2013). He is Senior member IEEE.



Anuj Srivastava is a Professor of Statistics at the Florida State University in Tallahassee, FL. He obtained his MS and PhD degrees in Electrical Engineering from the Washington University in St. Louis in 1993 and 1996, respectively. After spending the year 1996-97 at the Brown University as a visiting researcher, he joined FSU as an Assistant Professor in 1997. His research is focused on pattern theoretic approaches to problems in image analysis, computer vision, and signal processing. Specifically, he has developed computational tools for performing statistical inferences on certain nonlinear manifolds and has published over 200 refereed journal and conference articles in these areas.