



HAL
open science

Novel Generative Model for Facial Expressions Based on Statistical Shape Analysis of Landmarks Trajectories

Paul Audain Desrosiers, Mohamed Daoudi, Maxime Devanne

► **To cite this version:**

Paul Audain Desrosiers, Mohamed Daoudi, Maxime Devanne. Novel Generative Model for Facial Expressions Based on Statistical Shape Analysis of Landmarks Trajectories. 23rd International Conference on Pattern Recognition, ICPR 2016, International Association for Pattern Recognition, Dec 2016, Cancún, Mexico. pp.961 - 966, 10.1109/ICPR.2016.7899760 . hal-01521379

HAL Id: hal-01521379

<https://hal.science/hal-01521379>

Submitted on 11 May 2017

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.

Novel Generative Model For Facial Expressions Based On Statistical Shape Analysis of Landmarks Trajectories

Paul Audain Desrosiers, Mohamed Daoudi, Maxime Devanne
Télécom Lille, CRISTAL UMR (CNRS 9189), France
Email: (desrosiers, mohamed.daoudi, maxime.devanne)@telecom-lille.fr

Abstract—We propose a novel geometric framework for analyzing spontaneous facial expressions, with the specific goal of comparing, matching, and averaging the shapes of landmarks trajectories. Here we represent facial expressions by the motion of the landmarks across the time. The trajectories are represented by curves. We use elastic shape analysis of these curves to develop a Riemannian framework for analyzing shapes of these trajectories. In terms of empirical evaluation, our results on two databases: UvA-NEMO and Cohn-Kanade CK+ are very promising. From a theoretical perspective, this framework allows formal statistical inferences, such as generation of facial expressions.

I. INTRODUCTION

Shape analysis of motion trajectories of features in videos have been proposed recently. One motivation comes from human action recognition where the motions of the features are represented by trajectories as elements of nonlinear manifolds. Matikainen et al. [1] present a method for using the trajectories of tracked feature points in a bag of words paradigm for video action recognition. Despite of the promising results obtained, the authors do not take into account the geometric information of the trajectories. More recently, in the case of human skeleton in RGB-Depth images, Devanne et al. [2] propose to formulate the human action recognition as the problem of computing a distance between trajectories generated by the joints moving during the action. They consider the shape of such trajectories and perform shape analysis and comparison on the shape space where each action is characterized by a trajectory shape. This approach obtained very promising results on human action recognition. Similar to the ideas of Devanne et al., Su et al. [3] propose a metric which takes into account time-warping on a Riemannian manifold. They propose a metric, which allows the registration of trajectories and compute statistics of the trajectories. Su et al. [4] apply this framework to the problem of visual speech recognition. Motivated by the very good results obtained by Devanne et al. [2], we propose to extend this work to model not only one trajectory but a set of trajectories corresponding to the set of landmarks.

In particular, we are interested in the problem of facial expression recognition and classification. With the widespread diffusion of devices endowed with on-board cameras (e.g., hand-held devices, entertainment consoles, personal computers, surveillance and monitoring sensors), there is now an

increasing interest in performing online detection and recognition of spontaneous emotions rather than the traditional posed expressions [5]. This trend has potential applications in the diagnostics of pathologies, such as Alzheimer and Autism, human-computer interaction, gaming, augmented and virtual reality, drivers fatigue detection and many others. However, spontaneous facial expression recognition is significantly more challenging than recognizing posed ones [6], [7]. In [7], the authors present a smile classifier that can distinguish posed and spontaneous enjoyment smiles. This method employs a SVM classifier to classify a set of features which incorporate facial cues such as D-marker, symmetry, and dynamics. In [6], the authors proposed a metric learning method for spontaneous facial expression recognition. These approaches do not propose a statistical model of facial expressions.

In the present paper, we propose a new theoretical framework which uses the shape information of landmarks trajectories. The main contributions of this paper are:

- We propose a statistical shape analysis framework of trajectories in \mathbb{R}^2 . This framework is able to align/register, compare and compute statistics (mean, covariance and distribution) of a set of trajectories. Figure 1 shows an overview of the proposed framework.
- We propose a generative model of trajectories of the landmarks. We will demonstrate how it is able to infer facial expressions using statistical model of trajectories.
- We demonstrate the use of this framework theory using two data sets. The UvA-NEMO data set has been used to show how this framework is able to discriminate between spontaneous and deliberate smile expressions. Cohn-Kanade (CK+) data set is used to generate facial expressions.

The rest of this paper is organized as follows. Section II proposes a mathematical formulation of trajectory shapes as point on a non-linear manifold. Section III presents statistical framework for modeling variations of trajectory shapes. In section IV, we evaluate the proposed approach on UvA-NEMO and Cohn-Kanade CK+ data sets. Experimental results show its effectiveness in the classification of spontaneous and deliberate facial expressions, and in modeling statistical shapes of trajectories. Section V concludes and propose some future directions of this work.

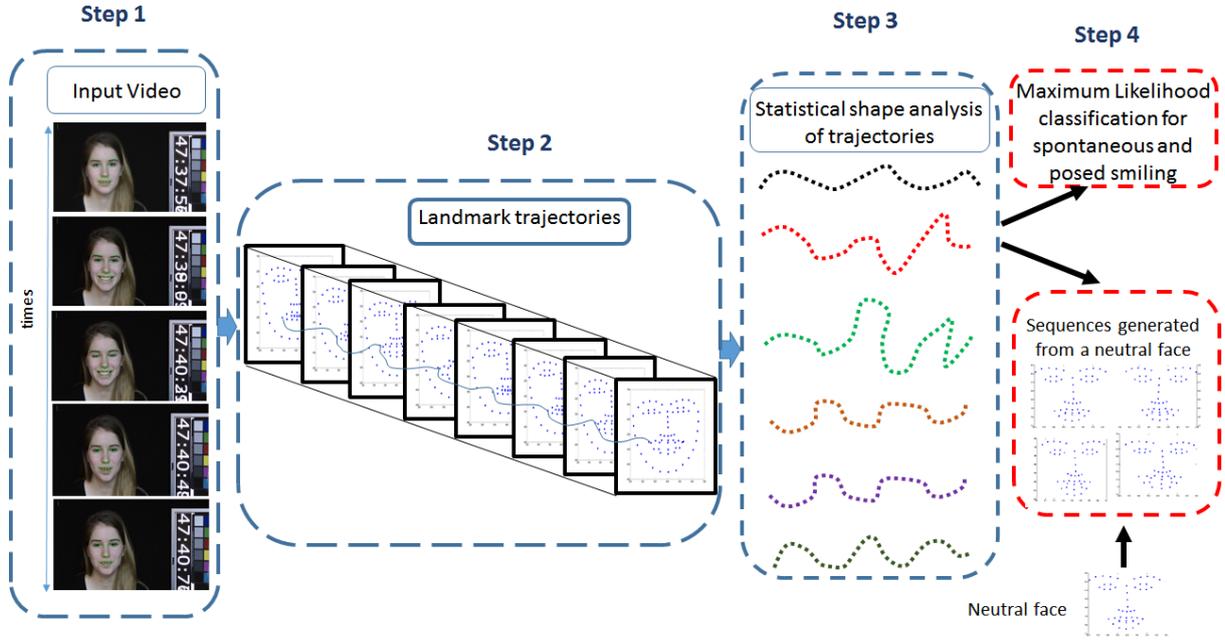


Fig. 1. Schematic description of the proposed pipeline

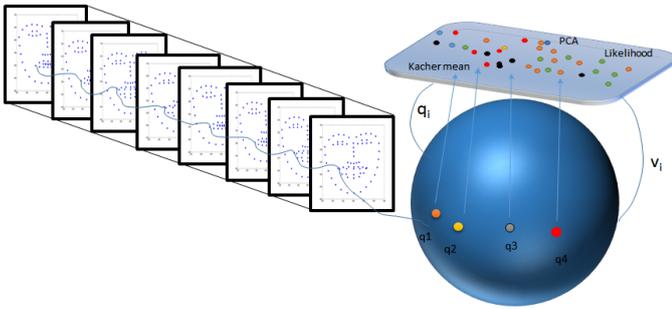


Fig. 2. Riemannian Trajectories of the Landmarks

II. SHAPE ANALYSIS OF TRAJECTORIES

In the last few years, many approaches have been developed to analyze shapes of 2D curves. We can cite approaches based on Fourier descriptors [8], Curvature Scale Space [9], moments or the median axis. More recent works in this area consider a formal definition of shape spaces as a Riemannian manifold of infinite dimension on which they can use classic tools for statistical analysis. Motivated by the promising results obtained in 3D facial recognition [10] and human action recognition [11], [2], we propose to use the shape analysis framework proposed by [12]. Each facial expression is represented by a set of trajectories of landmarks. Formally, we start by considering a given trajectory as a continuous parameterized function $\beta(t) \in \mathbb{R}^2$, $t \in [0, 1]$. β is first represented by its *Square-Root Velocity Function* (SRVF), q , according to :

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

Then, with the \mathbb{L}^2 -norm of the q functions scaled to 1 ($\|q\| = 1$), the space of such representation: $\mathcal{C} = \{q : [0, 1] \rightarrow \mathbb{R}^2, \|q\| = 1\}$ becomes a Riemannian manifold and have a spherical structure in the Hilbert space $\mathbb{L}^2([0, 1], \mathbb{R}^2)$. The spherical structure of \mathcal{C} is illustrated in figure 2. Given two curves β_1 and β_2 represented by their SRVFs q_1 and q_2 on the manifold, the geodesic path connecting q_1, q_2 is given analytically by the minor arc of the great circle connecting them on \mathcal{C} (see [13] for further details). It has been proved in [13] that under the \mathbb{L}^2 -norm, the quantities $\|q_1 - q_2\|$ and $\|q_1 \circ \gamma - q_2 \circ \gamma\|$ are same, where the composition $(q \circ \gamma)$ denotes the function q with a new parameterization dictated by a non-linear function $\gamma : [0, 1] \rightarrow [0, 1]$. This important property allows curves registration by re-parameterization, and thus makes the curves registration easier. In fact, it allows to consider one of the curves as reference and search for a $\gamma^* = \operatorname{argmin}_{\gamma \in \Gamma} (\|q_1 - \sqrt{\gamma} q_2 \circ \gamma\|)$ which optimally registers the two curves. This optimization is resolved by Dynamic Programming, as described in [13]. After the registration (let q_2 becomes q_2^* with the optimal re-parameterization function γ^*), we need to quantify the shape difference of the two curves. The distance between two elements q_1 and q_2^* is defined as $d_{\mathcal{C}}(q_1, q_2^*) = \cos^{-1}(\langle q_1, q_2^* \rangle)$. Such distance represents the similarity between the shape of two curves in \mathbb{R}^2 . Basically, it quantifies the amount of deformation between two shapes. This distance, also called elastic distance, is invariant to rotation, scaling and it takes into account the stretching and the bending of the curves [12].

III. STATISTICAL SHAPE ANALYSIS OF TRAJECTORIES

An important advantage of our Riemannian approach is its ability to compute summary statistics of a set of trajectories.

While a mean within a vector space is easy to compute, performing the task on a non-linear manifold such as \mathcal{C} is not obvious.

For example, one can use the notion of Karcher mean [14] to define an average trajectory that can serve as a representative trajectory of a group of trajectories. Given a set of data points q_i on a the manifold \mathcal{C} sufficiently close to each others, one way to define their geometric mean is via the minimization of a certain cost function defined by :

$$\mu = \underset{q_i}{\operatorname{argmin}} \sum_{i=1}^N d_{\mathcal{C}}(\mu, q_i)^2 \quad (1)$$

The algorithm 1 summarizes a Karcher mean computation of the trajectories corresponding to the same landmarks. For each landmarks trajectories, a Karcher mean is computed. The figure 2 shows this idea for one Karcher mean and one tangent space.

Algorithm 1: Karcher mean computation

Input: $\{q_1, q_2 \dots q_N\}$: points belonging to \mathcal{C} ,
 $\epsilon = 0.5$, τ : threshold which is a very small number
Output: μ_j : mean of $\{q_i\}_{i=1:N}$

- 1 **1-** μ_0 : initial estimate of Karcher mean, for example one could just take $\mu_0 = q_1$
- 2 **repeat**
- 3 **for** $i \leftarrow 1$ **to** N **do**
- 4 **2-** Compute $v_i = \exp_{\mu_j}^{-1}(q_i)$
- 5 **3-** Compute the average direction $\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$
- 6 **4-** Move μ_j in the direction of \bar{v} by ϵ :
 $\mu_{j+1} = \exp_{\mu_j}(\epsilon \bar{v})$
- 7 **end**
- 8 **5-** $j=j+1$
- 9 **until** $\|\bar{v}\| < \tau$;

Since manifolds lack a vector space structure and other Euclidean structures such as norm and inner product, machine learning algorithms including, principal component analysis (PCA) and Maximum Likelihood clustering algorithm cannot be applied in their original forms on the manifold \mathcal{C} . A common approach used to cope with its non-linearity consists in approximating the manifold valued data with its projection to a tangent space at a particular point on the manifold, for example, the mean of the data μ . Then, each sample shape q_i is mapped in the tangent space at the mean shape $T_{\mu}\mathcal{S}$ using the inverse exponential map [15] defined as:

$$v_i = \exp_{\mu}^{-1}(q_i) = \frac{\theta}{\sin\theta}(q_i - \cos(\theta)\mu), \quad (2)$$

where $\theta = d_{\mathcal{C}}(\mu, q_i)$. The original shape q_i can be retrieved from the velocity vector v_i by using the exponential map operator [15] defined as:

$$q_i = \exp_{\mu}(v_i) = \cos(\|v_i\|)\mu + \sin(\|v_i\|)\frac{v_i}{\|v_i\|}. \quad (3)$$

Such tangent space is a linear vector space which is more convenient to compute statistics. Hence, in order to learn the

distribution of tangent vectors on the tangent space, we can first perform PCA to learn a principal subspace denoted \mathcal{B} . Then, the covariance matrix on this principal basis is computed as $\Sigma = \sum_{i=1}^N v_i v_i^T$, where v_i are the tangent vectors projected into the principal subspace \mathcal{B} .

Finally, the multivariate normal distribution of facial expression c_k , $p(v|c_k; \Sigma)$ is learned using the covariance matrix Σ computed from the set of v_i where $|\Sigma|$ is the determinant of the covariance matrix Σ .

$$p(v|c_k; \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} v^T \Sigma^{-1} v} \quad (4)$$

Such a distribution is computed separately for each landmark trajectory from training videos. Then, the maximum-likelihood of a test video is computed by:

$$\hat{c}_k = \underset{c_k}{\operatorname{argmax}} \prod_{j=1}^L p(v_j|c_k; \Sigma) = \underset{c_k}{\operatorname{argmax}} \sum_{j=1}^L \log p(v_j|c_k; \Sigma) \quad (5)$$

Where L corresponds to the number of trajectories which also corresponds to the number of the landmarks. Such a maximum likelihood is employed for classification of test videos.

In addition, the previous learned distributions can be employed to generate random trajectory shapes representing random facial expressions. The algorithm 2 summarizes our approach for generating facial expressions from neutral face. This approach consists in generating random trajectories corresponding to the landmarks by using our generative model. Each point of these trajectories correspond to a moving point along the trajectories.

Algorithm 2: Generation of facial expressions

Input: Number of landmarks L , principal basis \mathcal{B} , covariance matrix Σ
Output: Set of random shapes $\{q_l\}_{l=1:L}$

- 1 **for** $i \leftarrow 1$ **to** L **do**
- 2 **1-** Compute a vector $v_l = \tau \Sigma \mathcal{B}$, where τ is a random vector with values between -1 and 1 .
- 3 **2-** Compute the shape $q_l = \exp_{\mu}(v_l)$
- 4 **end**

All the scripts in this framework are written in C/C++ and Matlab.

IV. EXPERIMENTAL RESULTS

To validate the proposed method, we use the UVA-NEMO smile database and the Cohn-Kanade data set (CK+). UVA-NEMO smile database is used to classify spontaneous/deliberate facial expressions and Cohn-Kanade data set is used to learn and to generate random facial expressions.

A. UvA-NEMO data set

The UvA-NEMO smile database [7] contains two types of dynamic smiles which are spontaneous and posed. Thus, 400 subjects (185 females, 215 males) have been required for

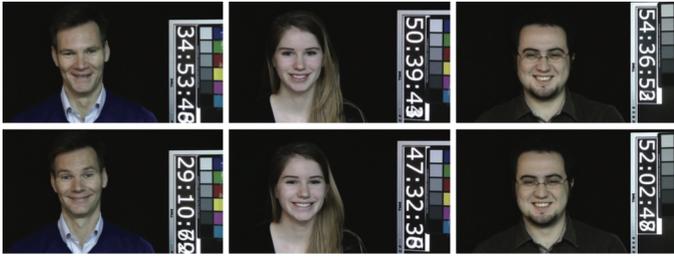


Fig. 3. Spontaneous (top) and posed (bottom) enjoyment smiles from the UvA-NEMO Smile Database.

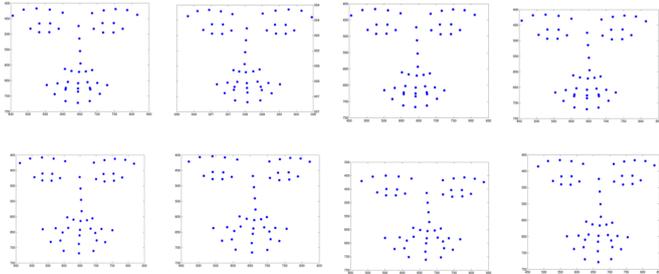


Fig. 4. A Video sequence UvA-NEMO database for a happy expression by using IntraFace.

this acquisition. The collected database contains 1240 smiles videos that are divided in two parts, 597 spontaneous and 643 posed smiles 2D videos. RGB videos are recorded with a resolution of 1920x1080 pixels at a rate of 50 frames per second. The age of the subjects varies from 8 to 76 years. At this time, no landmark annotation is released to the public for this database (See 3).

To Fully explore this database with the proposed method, we need a tool that can extract the facial features for each frame in each video. The Supervised Descent Method (SDM) recently developed in [16] helps us to extract those features in each video. We can report that in the UvA-NEMO database the length of videos varies from 1s to 14s. However, in the videos less than 4s, the subjects perform directly the smile, while in video greater than 4s the subjects do not only perform the smile but also move their head or their body. In order to validate our propose method on this database, we choose 400 videos, 200 that are used for spontaneous (100 for testing and 100 for training) and we do the same work for posed videos.

B. CK+ data set

The Cohn-Kanade dataset (CK+) [17] contains seven facial basic emotions (anger, disgust, fear, happy, sadness, surprise and contempt). CK+ is an annotated data set and also provides the landmark annotations. To validate the proposed method, we used only sequences annotated for the facial expressions. However, only 327 sequences out of 593 sequences have landmarks annotations, and duration of sequences varies from 7 to 60 frames. All sequences begin at the onset and stop at the apex because the maximum emotion is found in the apex. Moreover, the Active Appearance Model (AAM) is used as the

baseline. To extract the facial features of all the 327 sequences, the authors used the AAM to track the faces and extract the features (68 landmarks) as reported in figure 5.

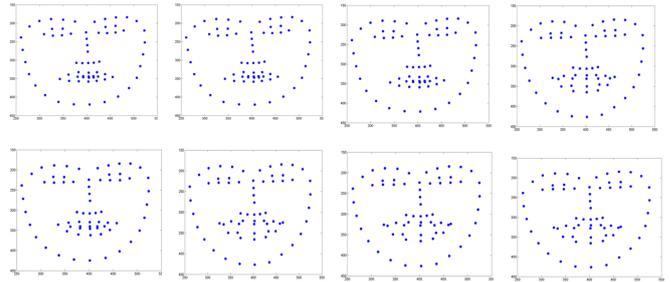


Fig. 5. A Video sequence on the CK+ for a happy expression, landmarks are released by the dataset.

The Supervised Descent Method (SDM) is largely used for its robustness to frame alignment and facial features detection. The available source code of IntraFace (IF) based on the SDM, allows to extract features from pictures and videos. Among all the possibilities that offered by IF, we used the latter to extract the facial features in a video sequences. Thus, we set each selected video in the database of UvA-NEMO as the input of IF. The output of IF corresponds to 49 facial features for each frame in each video, see figure 4. The variation of each facial features in each video represents 49 features trajectories. We report in figure 6 the trajectories of the landmarks for a smiling sequence in data set of CK+.

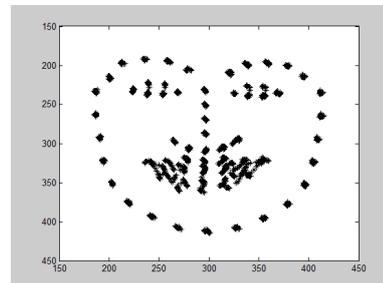


Fig. 6. Landmarks trajectories on CK+ database for a happy expression.

C. Results

In the proposed method, each facial feature in a video sequence represents a single trajectory. The shape of each trajectory is interpreted in a Riemannian manifold to achieve shape analysis and learn shape distributions as described in section III. On one hand, these distributions are used to classify and to test the expressions. On another hand, these distributions are employed to generate a set of random shapes resulting in random facial expressions. To compare those facial features from a video to another one, it is important that all the videos have the same length (or same number of frame). However, the length of those videos in the UvA-NEMO database is different from a video to another one. In our case, we empirically find a compromise between the

videos less than 4s (short video) and greater than 4s (long video), because the length of the video sequence vary from 55 frames to 705 frames.

We selected 100 shorts spontaneous and 100 posed videos for training, and 100 short spontaneous and 100 posed videos for testing. All the short videos are re-sampled empirically at the same length that is 160 frames, and the number of eigenvalues for the PCA is fixed to 50. After learning their distribution, we test 100 spontaneous and posed videos. We find 86% for spontaneous and 63% for posed videos, while in [7], they obtained 85.73% recognition by using Eyelid Features and 87.02% by using Mid-level fusion (voting). However these results are not easy to analyze. Indeed, the authors do not provide how many spontaneous expressions are classified as spontaneous. Moreover, this discriminative approach is not able to generate facial expressions.

The Cohn-Kanade data set contains many short videos that vary from 7 to 60 frames. We apply the same method described above. The data set is re-sampled to a length of 50. With this re-sampled data, we obtain best result for the expressions such as happy (100%), surprise (97,92%) and contempt (100%). In order to generate a new sequence video from the learned distributions, and show the performance of the method, we used these facial expressions giving the best rate of recognition in CK+.

By applying the algorithm 2, we can generate a set of random trajectory shapes from a distribution corresponding to a given expressions. These trajectories shapes represent the deformation of each landmark along the time. By applying such deformation to an initial neutral face, we can visualize a random sequence corresponding to the chosen expression. The main motivation to generate facial expressions is this ability to animate avatar in Unity, unreal engine with a chosen expression, and to make the relation between human and avatar become more realistic.

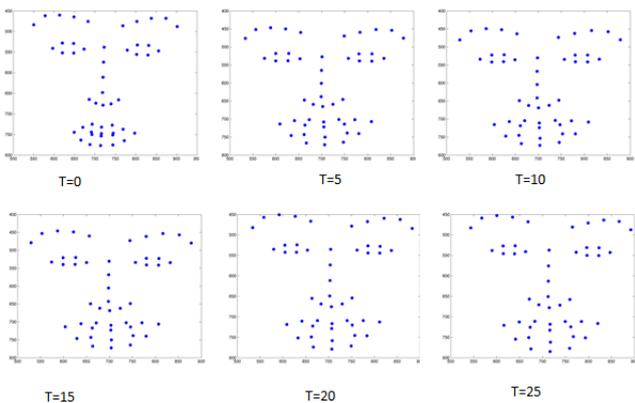


Fig. 7. Random spontaneous smile sequence generated from the UvA-NEMO database

From the distributions learned from the UVA-NEMO database, we first generate random sequences corresponding to a spontaneous smile (Fig. 7) and a posed smile (Fig. 8). Each plot of the figures corresponds to a mask at different

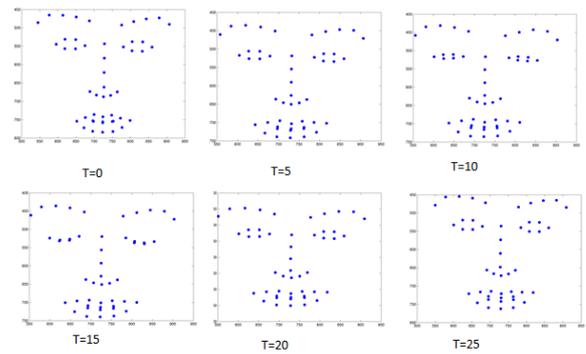


Fig. 8. Random posed smile sequence generated from the UvA-NEMO database

time step T . In a second time, we employ CK+ dataset to generate random sequences of each facial expressions from corresponding learned distribution. Figures 9, 10 and 11 show examples of the expressions contempt, happy and surprise.

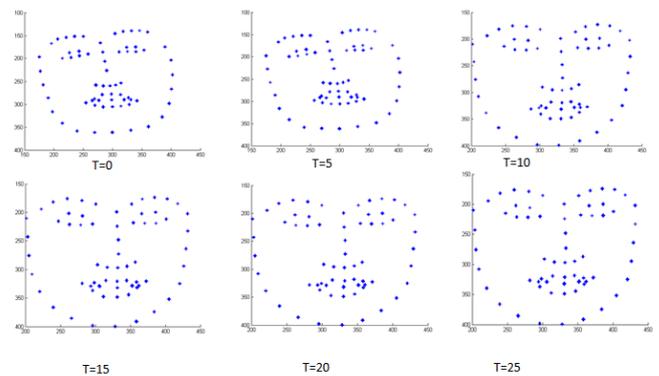


Fig. 9. Random contempt expression sequence generated from the CK+ database

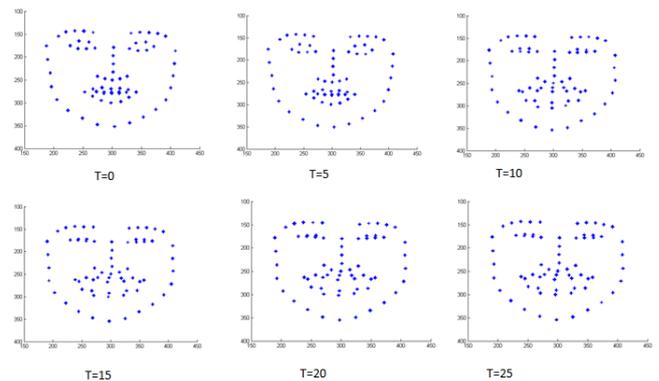


Fig. 10. Random Happy expression sequence generated from the CK+ database

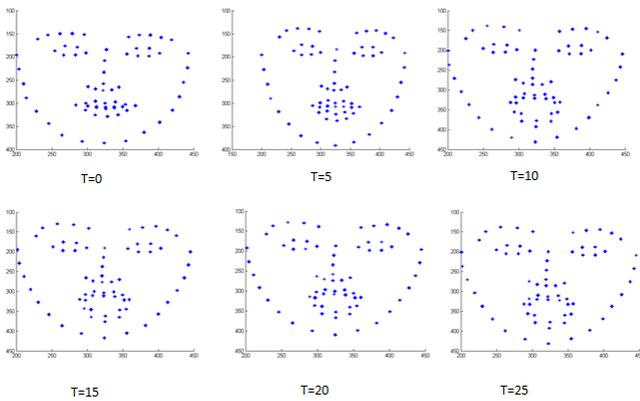


Fig. 11. Random surprise expression sequence generated from the CK+ database

V. CONCLUSIONS

In this paper, we have proposed one of the first framework for modeling the trajectories corresponding to the motion of a set of landmarks extracted from facial expressions sequences. To take into account the non-linearity of temporal evolution of trajectories, we proposed a statistical shape analysis of these trajectories. The first results obtained on two well known datasets UvA-NAMO and CK+ show that the results obtained by the proposed approach are very promising for two important applications in computer visions: spontaneous/posed facial expression classification and the generation of facial expressions from neutral face. In the future, we would like to deeply analyze the landmark trajectories to identify which are the best trajectories more contributing to the facial expressions. This framework could be a first step for the new generation of advanced avatars, which can make the conversation between human and avatars more realistic.

ACKNOWLEDGMENT

This work was partially supported by the FUI project MAGNUM 2 and the Programme d'Investissements d'Avenir (PIA) and Agence Nationale pour la Recherche (grant ANR-11-EQPX-0023), and European Funds for the Regional Development (Grant FEDER- Presage 41779).

REFERENCES

- [1] P. Matikainen, M. Hebert, and R. Sukthankar, "Trajectons: Action recognition through the motion analysis of tracked features," in *Workshop on Video-Oriented Object and Event Classification, ICCV 2009*, September 2009.
- [2] M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. Del Bimbo, "3D human action recognition by shape analysis of motion trajectories on riemannian manifold," *IEEE Trans. on Cybernetics*, vol. 45, no. 7, pp. 1340–1352, 2014.
- [3] J. Su, S. Kurtek, E. Klassen, and A. Srivastava, "Statistical analysis of trajectories on riemannian manifolds: Bird migration, hurricane tracking and video surveillance," *Annals of Applied Statistics*, vol. 8, no. 1, 2014.
- [4] J. Su, A. Srivastava, F. D. M. de Souza, and S. Sarkar, "Rate-invariant analysis of trajectories on riemannian manifolds with application in visual speech recognition," June 2014.
- [5] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, 2009.

- [6] S. Wan and J. K. Aggarwal, "Spontaneous facial expression recognition: A robust metric learning approach," *Pattern Recognition*, vol. 47, no. 5, pp. 1859–1868, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.patcog.2013.11.025>
- [7] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Are you really smiling at me? spontaneous versus posed enjoyment smiles," in *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III*, 2012, pp. 525–538.
- [8] A. E. Oirrak, M. Daoudi, and D. Aboutajdine, "Estimation of general 2d affine motion using fourier descriptors," *Pattern Recognition*, vol. 35, no. 1, pp. 223–228, 2002.
- [9] M. Daoudi and S. Matusiak, "Visual image retrieval by multiscale description of user sketches," *J. Vis. Lang. Comput.*, vol. 11, no. 3, pp. 287–301, 2000.
- [10] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3d face recognition under expressions, occlusions, and pose variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, 2013. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2013.48>
- [11] M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. Del Bimbo, "Space-time pose representation for 3d human action recognition," in *ICIAP Workshops*, ser. Lecture Notes in Computer Science, A. Petrosino, L. Maddalena, and P. Pala, Eds., vol. 8158. Springer, 2013, pp. 456–464.
- [12] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn, "Shape analysis of elastic curves in euclidean spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1415–1428, 2011.
- [13] —, "Shape analysis of elastic curves in euclidean spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1415–1428, 2011.
- [14] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, vol. 30, pp. 509–541, 1977.
- [15] S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding, "Statistical modeling of curves using shapes and related features," *Journal of the American Statistical Association*, vol. 107, no. 499, pp. 1152–1165, 2012.
- [16] X. Xiong and F. D. la Torre, "Supervised descent method and its applications to face alignment," in *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, 2013, pp. 532–539. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2013.75>
- [17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, June 2010, pp. 94–101.