Magnifying subtle facial motions for 4D Expression Recognition

Qingkai Zhen, Di Huang, Yunhong Wang, Hassen Drira, Ben Boulbaba, Mohamed Daoudi

Magnifying Subtle Facial Motions for 4D Expression Recognition

Qingkai Zhen
IRIP Lab, School of Computer Science and Engineering, Beihang University, China, Email: qingkai.zhen@buaa.edu.cn

Di Huang
IRIP Lab, School of Computer Science and Engineering, Beihang University, China, Email: dhuang@buaa.edu.cn

Yunhong Wang
IRIP Lab, School of Computer Science and Engineering, Beihang University, China, Email: yhwang@buaa.edu.cn

Hassen Drira
Institut Mines-Télécom/Télécom Lille, CRIStAL (UMR CNRS 9189), France
Email: hassen.drira@telecom-lille.fr

Boulbaba Ben Amor
Institut Mines-Télécom/Télécom Lille, CRIStAL (UMR CNRS 9189), France
Email: boulbaba.benamor@telecom-lille.fr

Mohamed Daoudi
Institut Mines-Télécom/Télécom Lille, CRIStAL (UMR CNRS 9189), France
Email: mohamed.daoudi@telecom-lille.fr

***Abstract*—In this paper, an effective method is proposed to address the problem of automatic 4D Facial Expression Recognition (*4D FER*). The flow of 3D faces is first modeled to capture the spatial deformations based on the recently-developed Riemannian approach, namely *Dense Scalar Fields (DSF)*, where registration and comparison of neighboring 3D face frames are jointly led. The deformations obtained are then fed into a temporal filtering based magnification step in order to amplify the slight facial actions over time. The proposed method allows revealing subtle (hidden) deformations which enhance the performance of emotion classification. We evaluate our approach on the BU-4DFE dataset, and the state-of-art accuracy up to 94.18% is achieved, which is superior to the top one so far reported, clearly demonstrating its effectiveness.**

I. Introduction

Facial expressions analysis and recognition from 3D data has attracted lots of researchers due to its diverse applications in the past decade, such as facial animation, human-computer interaction, *etc.*In recent years, there has been tremendous interest in tracking and recognition facial expressions from dynamic 3D facial expression sequence (4D data), it is suggested that the dynamics of facial expressions provides important cues about the underlying emotions that are not available in static 3D images.

There are a few works that use 4D data for facial expression analysis. Sun and Yin, the pioneers of 4D *FER*, extracted a *Spatio-Temporal* (*ST*) descriptor from dynamic sequences of 3D facial scans [1], and applied *HMM* classifier to predict the expression type. In [2], the tracking-model-based approach is presented for vertex correspondences, vertex motion estimation, and *HMM* is trained to learn the spatial and temporal information of the 3D model sequence. Canavan *et al.* [3] presented a dynamic curvature based approach for facial activity analysis, then constructed the dynamic curvature descriptor from local regions as well as temporal domains, and *SVM* classifier is adopted for classification. Sandbach *et al.* [4] exploited 3D motion-based features (*Free-Form Deformation, FFD*) between neighboring 3D facial geometry frames for *FER*. A feature selection step was applied to localize the features of each of the onset and offset segments of the expression. The *HMM* classifier was used to model the full temporal dynamics of each expression. In [5], the entire expressive sequence is modelled to contain an *Onset* followed by an *Apex* and an *Offset*. Feature selection methods are applied to extract features for each of the onset and offset segments of the expression. These features are then used to train *GentleBoost* classifiers and build an *HMM* to model the full temporal dynamics of the expression. Ben Amor *et al* [6], [7], [8] presented the facial expression deformation by collections of radial curves, Dense Scalar Fields (*DSFs*) features are feed into *Random Forest* or *HMM* classifier for classification. Xue *et al.* [9] applied three dimension discrete cosine transform (*3D-DCT*) on local depth patch-sequences to extract spatio-temporal features, and selected nearest-neighbor classifier to make decision.

Even though the performance of 4D *FER* has been great boosted in recent years, amplify the subtle movement on the facial surface is still an unsolved problem. We present a novel and effective approach to handle this problem, our contributions are two-folds:

- A comprehensive pipeline of spatio-temporal processing for effective facial expression recognition from 4D data.
- A method to amplify subtle movements on facial surfaces which contributes to distinguish similar expressions.

The rest of the paper is structured as follows. Section II introduces the background of the *DSF* based geometry feature. The magnification of subtle facial deformation is described in Section III. The experimental results are presented and analyzed in Section IV, followed by Section V where we conclude the paper.

II. Dense Scalar Fields

Following the geometric approach recently-developed in [6], we represent 3D facial surfaces by a collection of radial curves emanating from the nose tip. It is a parameterization imposed for 3D face description, registration, comparison, *etc.*The amount of deformation from one shape to another

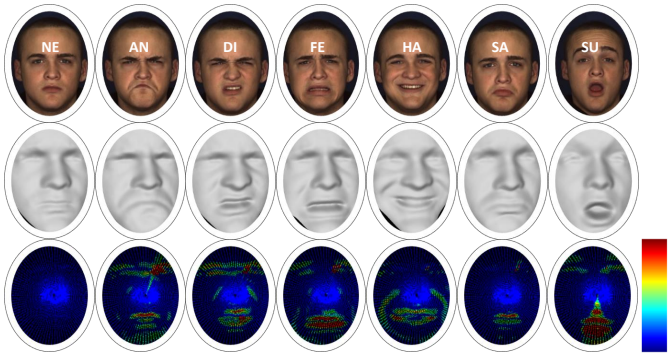


Fig. 1: Top row: facial texture images of an individual with different expressions for visualization; middle row: depth maps of each 3D face frame used for *FER* in this study; and bottom row: facial deformations in the Riemannian space. Warm colors are associated to bigger χ values and correspond to the facial regions with high deformations, and cold colors reflect the most static facial parts.

(across the 3D video) is computed through analyzing the differences between the corresponding radial curves based on the theory of differential geometry.

Specifically, in the pre-processing step, the 3D face mesh in each frame is first aligned to the first one. The facial surfaces are then approximated by indexed collections of radial curves β_α, where the index α denotes the angle formed by the curve with respect to a reference one. These curves are further uniformally resampled. Given a radial curve β of the facial surface with an arbitrary orientation $\alpha \in [0, 2\pi]$, it can be parameterized as $\beta : I \rightarrow \mathbb{R}^3$, with $I = [0, 1]$, and mathematically represented using the square-root velocity function (*SRVF*), denoted by $q(t)$, according to:

$$q(t) = \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}}, t \in I. \tag{1}$$

This geometry representation has the advantage of capturing the curve shape and makes the calculus simpler. While there are several ways to measure the curve shape, an elastic analysis of the parametrized curves is appropriate in this application, particularly under facial expression variations. This is because (1) such analysis uses the square-root velocity function representation which allows for the comparison of local facial shapes in the presence of deformations; (2) this method employs a square-root representation under which the elastic metric is reduced to the standard $\mathbb{L}^2$ metric and thus simplifies the analysis; (3) based on this metric, the group of re-parametrization acts by isometry on the curves manifold, and a Riemannian re-parametrization metric can thus be set between two facial curves. Shown in Fig. 1 are examples of apex frames taken from the 3D videos on the BU-4DFE dataset as well as the dense 3D deformations computed with respect to their neutral frames. Let us define the space of the *SRVFs* as

$$\mathcal{C} = \{q : I \rightarrow \mathbb{R}^3, \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3) \tag{2}$$

where $\|\cdot\|$ indicates the $\mathbb{L}^2$ norm. With the $\mathbb{L}^2$ metric on its tangent space, $\mathcal{C}$ becomes a Riemannian manifold. Basically, using this parametrization, each radial curve is represented on the manifold $\mathcal{C}$ by its *SRVF*. Accordingly, given *SRVF* q_1 and q_2 of two curves, the shortest path ψ^* on the manifold $\mathcal{C}$ between them (called geodesic path) is a critical point of the following energy function:

$$E(\psi) = \frac{1}{2} \int \|\dot{\psi}(\tau)\|^2 d\tau \tag{3}$$

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, and $\dot{\psi}(\tau) \in T_{\psi(\tau)}(\mathcal{C})$ is the tangent vector field on the curve $\psi(\tau) \in \mathcal{C}$. Since elements of $\mathcal{C}$ have a unit $\mathbb{L}^2$ norm, $\mathcal{C}$ is an hypersphere in the Hilbert space $\mathbb{L}^2(I, \mathbb{R}^3)$. As a consequence, the geodesic path between any two points q_1 and $q_2 \in \mathcal{C}$ is given by the minor arc of the great circle connecting them. The tangent vector field on this geodesic between the curves β_1 and β_2 making the angle α with the reference curve is parallel along the geodesic and one can represent it with the initial velocity vector (called shooting vector) without any loss of information.

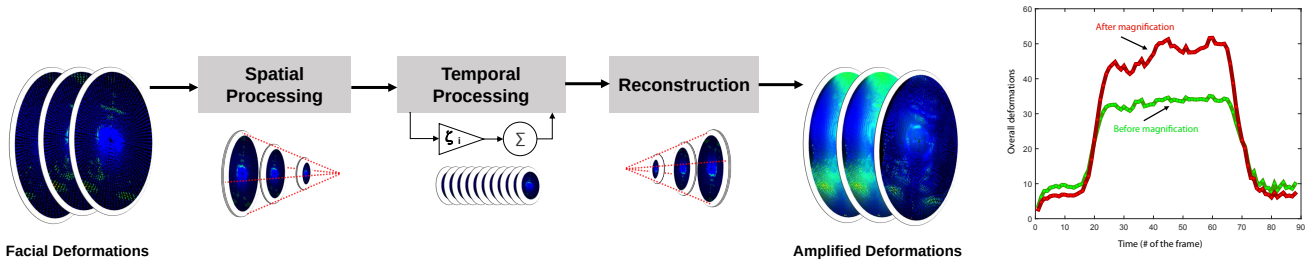
$$\frac{d\psi^*_\alpha}{d\tau}_{|\tau=0} = \frac{\theta}{\sin(\theta)}(q_2 - \cos(\theta)q_1), \ (\theta \neq 0). \tag{4}$$

where $\theta = d_{\mathcal{C}}(q_1, q_2) = cos^{-1}(\langle q_1, q_2 \rangle)$ represents the length of the geodesic path connecting q_1 to q_2. In practice, the curves are re-sampled to a specified number of points, say T, and the face is approximated by a collection of $|\Lambda|$ curves. The norm of the quantity at each discrete point r is computed to measure the amount of 3D deformation in this position of the surface parameterized by the pair (α, r), termed Dense Scalar Fields (*DSFs*). The final feature vector is of the size $T \times |\Lambda|$. We will refer to this quantity at a given time t of the 3D video by $\chi(t)$ (see bottom row of Fig. 1 for illustration). It provides the amplitude of the deformations between two facial surfaces in a dense way.

III. Subtle Facial Deformation Magnification

As described in Section II, χ reveals the shape difference between two facial surfaces by deforming one mesh to another through an accurate registration step. However, there exists another challenge to capture certain facial movements, especially the slight ones, with low spatial amplitude, reflected by the limited performance in distinguishing similar 3D facial expressions in the literature. To solve this problem, we propose a novel approach to highlight the subtle geometry changes of the facial surface in χ by adapting the Eulerian spatio-temporal processing [10] to the 3D domain.

The Eulerian spatio-temporal processing was introduced for motion magnification in 2D videos [10]. Its basic idea is to amplify the variation of pixel values over time in a spatially-multiscale manner without explicitly estimating motion but exaggerating it by amplifying temporal color changes at fixed



(a) (b)

Fig. 2: (a) Overview of 3D video magnification. The original facial deformation features are first decomposed into different spatial frequencies, and the temporal filter is applied to all the frequency bands. The filtered spatial bands are then amplified by a given factor ζ, added back to the original signal, and collapsed to the output sequence. (b) An example of facial expression deformation (norm of the velocity vector) before (green) and after (red) magnification.

positions. It relies on a linear approximation related to the brightness constancy assumption that forms the basis of the optical flow algorithm. However, the case is not that straightforward in 3D, because the vertex correspondence across frames cannot be achieved as easy as that in 2D. Fortunately, during the computation of χ, such correspondence is established by surface registration and remeshing, and we can thus apply Eulerian spatio-temporal processing to 3D video. We take into account the values of the time series χ at any spatial location and highlight the differences in a given temporal frequency band of interest. It therefore combines spatial and temporal processing to emphasize subtle changes in a 3D face video.

The process is shown in Fig. 2(a). Specifically, the video sequences are first decomposed into different spatial frequency bands by Gaussian pyramid, and these bands tend to be magnified differently. We consider that the time series correspond to the values of χ on the mesh surfaces in a frequency band and apply a band pass filter to extract the frequency bands of interest. The temporal processing, $\mathfrak{T}$, is uniform for all spatial levels and all χ within each level. We then multiply the extracted band passed signal by a magnification factor ζ, and add the magnified signal to the original and collapse the spatial pyramid to obtain the final output.

For the translational motion of the facial mesh, we express the observed $\chi(s,t)$ value with respect to a displacement function $\delta(t)$, such that $\chi(s,t) = \chi(s) + \delta(t)$ and $\chi(s,0) = \chi(s)$. By using a first-order Taylor series expansion, at time t, $\chi(s+\delta(t))$ can be approximated as

$$\chi(s,t) \approx \chi(s) + \delta(t)\frac{\partial \chi(s)}{\partial s} \tag{5}$$

Let $\phi(s,t)$ be the result of applying a broadband temporal band pass filter to $\chi(s)$ at each position (s). Assume that the motion signal $\delta(t)$ is within the pass band of the temporal filter.

$$\phi(s,t) = \delta(t)\frac{\partial \chi(s)}{\partial s} \tag{6}$$

Amplify the band pass signal by factor ζ and add it back to $\chi(s)$.

$$\hat{\chi}(s,t) = \chi(s,t) + \zeta\phi(s,t) \tag{7}$$

Algorithm 1: Online 3D Deformation Magnification

Input: χ, l-Gaussian pyramid levels, ζ-amplification factor, ξ-sample rate, γ-attenuation rate, ƒ-video frequency

Step1. Spatial Processing

for $i = 1; i \leq n$ **do**

  $\mathfrak{D}(i,:,:,:)$ = decompose the $\chi(i)$, with l level Gaussian pyramid.

Step2. Temporal Processing

$\mathfrak{S}$ = $\mathfrak{T}(\mathfrak{D}$, ƒ, $\xi)$

Step3. Magnification

for $i = 1; i \leq 3$ **do**

  $\mathfrak{S}(:,:,:,\mathrm{i})$ = $\mathfrak{S}(:,:,:,\mathrm{i}) * \zeta * \gamma$

Step3. Reconstruction

for $i = 1; i \leq n$ **do**

  $\hat{\chi}(i)$ = $\mathfrak{S}(i,:,:,:)$ + $\chi(i)$

Output: $\hat{\chi}(t)$

By combining (5), (6), and (7), we reach

$$\hat{\chi}(s,t) \approx \chi(s) + (1+\zeta)\delta(t)\frac{\partial \chi(s)}{\partial s} \tag{8}$$

Assuming that the first-order Taylor expansion holds for the amplified larger perturbation $(1+\zeta)\delta(t)$, the motion magnification of 3D face video can be simplified as:

$$\hat{\chi}(s,t) \approx \chi(s + (1+\zeta)\delta(t)) \tag{9}$$

This shows that the spatial displacement $\delta(t)$ of the $\chi(s)$ at time t, is amplified to a magnitude of $(1+\zeta)$. Sometimes $\delta(t)$ is not entirely within the pass band of the temporal filter. In this case, let $\delta_k(t)$, indexed by k, represent different temporal spectral components of $\delta(t)$. The result in a band pass signal is,

$$\phi(s,t) = \sum_k \gamma_k \delta_k(t)\frac{\partial \chi(s)}{\partial s} \tag{10}$$

where γ is an attenuation factor. Temporal frequency dependent attenuation can be equivalently interpreted as a frequency-

dependent motion magnification factor, $\zeta_k = \gamma\zeta$, and the amplified output signal is computed by

$$\hat{\chi}(s,t) \approx \chi(s + \sum_k (1+\zeta k)\delta_k(t)) \quad (11)$$

Fig. 2(b) displays an example of facial deformation trajectory before (green) and after (red) magnification.

IV. Experimental Results

A. Dataset and Protocol

The BU-4DFE dataset [1] is a dynamic 3D facial expression dataset which consists of 3D facial sequences of 58 females and 43 males. It includes in total 606 3D sequences possessing the 6 universal expressions. Each 3D sequence captures a facial expression at a rate of 25 fps (frames per second) and lasts approximately 3-4 seconds.

In our experiments, at a time t, the 3D face model f^t is approximated by a set of 200 elastic radial curves originating from the nose tip, and a total of 50 points on each curve are sampled. Based on this parameterization, each 3D face geometry in the video sequence is compared to a reference frame f^0 to derive $\chi(t)$ at time t. Then, within the spatial processing step, a Gaussian pyramid decomposition is used to decompose χ into 4 band levels. Finally, a temporal processing to all the bands is applied. The factor ζ is set to 10, the sample rate ξ is set to 25, $\mathfrak{f} \in [0.3, 0.4]$, and the attenuation rate γ is set to 1.

It should be noted that the proposed approach can either be evaluated when making use of full sequences [6], [9], [11] or sliding window-based sub-sequences [1], [2], [5], [12], while as pointed in [13], the latter can bias the final result. As a result, we exploit the former, and our experiments are conducted on two sub-pipelines: (1) the whole video sequence (denoted as *WV*) and (2) the magnified whole video sequence (denoted as *MWV*).

A multi-class Support Vector Machine (*SVM*) is exploited where $\bar{\chi}$ is treated as a feature vector to predict the expression label. We also adopt *HMM* to encode the temporal behavior of the sequence for decision. To allow fair comparison with the previous studies, we randomly select 60 subjects from the BU-4DFE dataset under 10-fold cross-validation protocol.

B. Performance

Table I shows that the magnification procedure achieves an improvement of around 10% in the accuracy for both classifiers, *i.e. SVM* and *HMM*. Without magnification, our approach reaches the performance of 82.49% and 83.19% using the *SVM* and *HMM* classifiers, while the results are improved to 93.39% and 94.18%, which highlights the effectiveness of magnification in 3D face videos. Table II shows the confusion matrices (*WV*, *MWV*) achieved by using the *SVM* and *HMM* classifiers. From this table, we can see that the *SU* and *HA* sequences are better predicted than the others. This is mainly due to the clear patterns and high intensities of their deformations. The remaining expressions (*DI*, *FE*, *AN* and *SA*) are harder to distinguish. We believe that two major reasons induce this difficulty: (1) intra-class variations make similar classes confusing, such as *DI*/*AN*/*FE*; and (2) lower deformation magnitude is often exhibited when these expressions are performed. Furthermore, it can be seen from these confusion matrices, the accuracies in distinguishing *AN*, *DI* and *FE* expressions are all significantly ameliorated. Fig. 3 gives more illustrations of deformation magnification on the sequences of the same subject possessing the six prototype expressions.

C. Comparison with state-of-the-art

Several studies report their *FER* results on the BU-4DFE dataset; however they differ in the experimental setting. In this section, we compare our results with the one of the existing approaches considering these differences.

Top results on BU-4DFE are shown in Table III. In this table, #E denotes the number of expressions, #S is the number of subjects, #-CV provides the number of cross-validation, and *Full Seq./Win* means the decision is made based on full sequence or sub-sequences captured using a sliding window. [2] reports the highest accuracy when using a sliding window of 6 frames; nevertheless, the approach requires manual annotation of 83 landmarks on the first frame. Moreover, the vertex-level dense tracking is time consuming. In a more recent work from the same group developed by Reale *et al.* [12], the authors deliver a classification rate of 76.9% using the sequences of 100 subjects with a fixed size of window of 15 frames, when segmentation is manually applied to the 3D face video to extract the expressive time interval. In [14], Fang *et al.* reach an accuracy of 74.63% with 507 sequences of 100 subjects, but they do not provide details on the classification scheme. Le *et al.* [11] evaluate their algorithm on the sequences of 60 subjects only on three expressions (*HA*, *SA* and *SU*) and display the result of 92.22%. Regarding on the tasks that conduct classification under the same protocol [6], [9], [13], the proposed method outperforms them, demonstrating its competency at 4D *FER*. Besides, it also possesses the advantages: (1) no manual landmark is required; and (2) no dimensionality reduction or feature selection techniques are applied.

V. Conclusions

In this paper, an effective approach is presented for 4D *FER*. It focuses on improving the performance by 3D video magnification which reveals subtle facial deformations. After a preprocessing step, the flow of 3D faces is first modeled to capture spatial shape changes in the *DSF* based Riemannian geometry space, where registration and comparison are jointly achieved. Such deformations are then amplified using the temporal filter over time. The prediction is finally carried out using these magnified features. Experimental results on the BU-4DFE dataset demonstrate the effectiveness of the proposed method.

TABLE I: Average accuracy with standard deviation achieved by *SVM* and *HMM* using full sequence before and after magnification on the BU-4DFE database.

Algorithm	**Magnification?**	**Performance** (%)
SVM on $\bar{\chi}$	N	82.49 ± 3.10
	Y	93.39 ± 3.54
HMM on $\chi(t)$	N	83.19 ± 2.83
	Y	94.18 ± 2.46

TABLE II: Confusion matrices (*WV*, *MWV*) achieved by the *SVM* and *HMM* classifiers respectively on the BU-4DFE database.

SVM on $\bar{\chi}$	**Whole Video (*WV*)**						**Magnified Whole Video (*MWV*)**					
%	AN	DI	FE	HA	SA	SU	AN	DI	FE	HA	SA	SU
AN	**73.86**	9.18	6.49	1.75	6.11	2.51	**91.07**	2.73	2.01	1.59	2.08	0.51
DI	8.76	**71.27**	9.29	3.51	4.84	2.21	2.05	**92.62**	2.63	1.07	1.38	0.24
FE	5.79	5.37	**73.14**	4.59	5.39	5.61	1.66	1.53	**92.33**	1.31	1.54	1.62
HA	0.81	1.18	2.42	**93.6**	1.08	0.88	0.91	0.88	2.36	**94.29**	0.97	0.58
SA	2.54	2.27	2.99	1.63	**88.75**	1.77	1.36	1.22	1.62	0.9	**93.93**	0.96
SU	0.74	0.88	1.91	0.75	1.38	**94.32**	0.51	0.61	1.29	0.52	0.96	**96.11**
Average	**82.49 ± 3.10**						**93.39 ± 3.54**					
HMM on $\chi(t)$												
%	AN	DI	FE	HA	SA	SU	AN	DI	FE	HA	SA	SU
AN	**75.29**	5.88	7.31	1.14	8.17	2.21	**91.87**	1.91	2.41	0.38	2.69	0.73
DI	10.42	**71.55**	11.43	1.82	4.27	0.5	2.11	**94.22**	2.32	0.29	0.86	0.19
FE	5.07	6.86	**73.69**	3.33	8.06	2.99	1.37	1.86	**92.85**	0.91	2.19	0.81
HA	0.48	0.87	1.54	**94.93**	1.81	0.37	0.47	0.77	1.43	**95.3**	1.67	0.35
SA	3.71	1.01	4.17	0.65	**89.19**	1.26	1.84	0.51	2.07	0.33	**94.61**	0.63
SU	0.49	0.33	2.79	0.32	1.59	**94.47**	0.33	0.22	1.89	0.22	1.08	**96.25**
Average	**83.19 ± 2.83**						**94.18 ± 2.46**					

TABLE III: Comparative results with the state-of-the-art on BU-4DFE.

Method	**Experimental Settings**	**Accuracy**
Sun *et al.* [1]	6E, 60S, 10-CV, Win=6	90.44%
Sun *et al.* [2]	6E, 60S, 10-CV, Win=6	94.37%
Reale *et al.* [12]	6E, 100S, –, Win=15	76.9%
Sandb. *et al.* [5]	6E, 60S, 6-CV, Win	64.6%
Fang *et al.* [14]	6E, 100S, 10-CV, –	74.63%
Le *et al.* [11]	3E, 60S, 10-CV, Full seq.	92.22%
Xue *et al.* [9]	6E, 60S, 10-CV, Full seq.	78.8%
Berretti *et al.* [13]	6E, 60S, 10-CV, Full seq.	79.4%
Berretti *et al.* [13]	6E, 60S, 10-CV, Win=6	72.25%
Ben Amor *et al.* [6]	6E, 60S, 10-CV, Full seq.	93.21%
Ben Amor *et al.* [6]	6E, 60S, 10-CV, Win=6.	93.83%
This work – *SVM* on $\bar{\chi}$	6E, 60S, 10-CV, Full seq.	**93.39%**
This work – *HMM* on $\chi(t)$	6E, 60S, 10-CV, Full seq.	**94.18%**

REFERENCES

[1] Y. Sun and L. Yin, "Facial expression recognition based on 3d dynamic range model sequences," in *Europeon Conference on Computer Vision*, 2008, pp. 58–71.

[2] Y. Sun, X. Chen, M. 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: Systems and Humans*, vol. 40, no. 3, pp. 461–474, 2010.

[3] S. Canavan, Y. Sun, X. Zhang, and L. Yin, "A dynamic curvature based approach for facial activity analysis in 3d space," in *Computer Vision and Pattern Recognition Workshops*, 2012, pp. 14–19.

[4] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "A dynamic approach to the recognition of 3d facial expressions and their temporal models," in *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 406–413.

[5] ——, "Recognition of 3d facial expression dynamics," *Image and Vision Computing*, vol. 30, no. 10, pp. 762–773, 2012.

[6] B. Ben Amor, H. Drira, S. Berretti, M. Daoudi, and A. Srivastava, "4-d facial expression recognition by learning geometric deformations," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2443–2457, 2014.

[7] H. Drira, B. B. Amor, M. Daoudi, A. Srivastava, and S. Berretti, "3d dynamic expression recognition based on a novel deformation vector field and random forest," in *International Conference on Pattern Recognition*, 2012, pp. 1104–1107.

[8] M. Daoudi, H. Drira, B. B. Amor, and S. Berretti, "A dynamic geometry-based approach for 4d facial expressions recognition," in *European Workshop on Visual Information Processing*, 2013, pp. 280–284.

[9] M. Xue, A. Mian, W. Liu, and L. Li, "Automatic 4d facial expression recognition using dct features," in *IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 199–206.

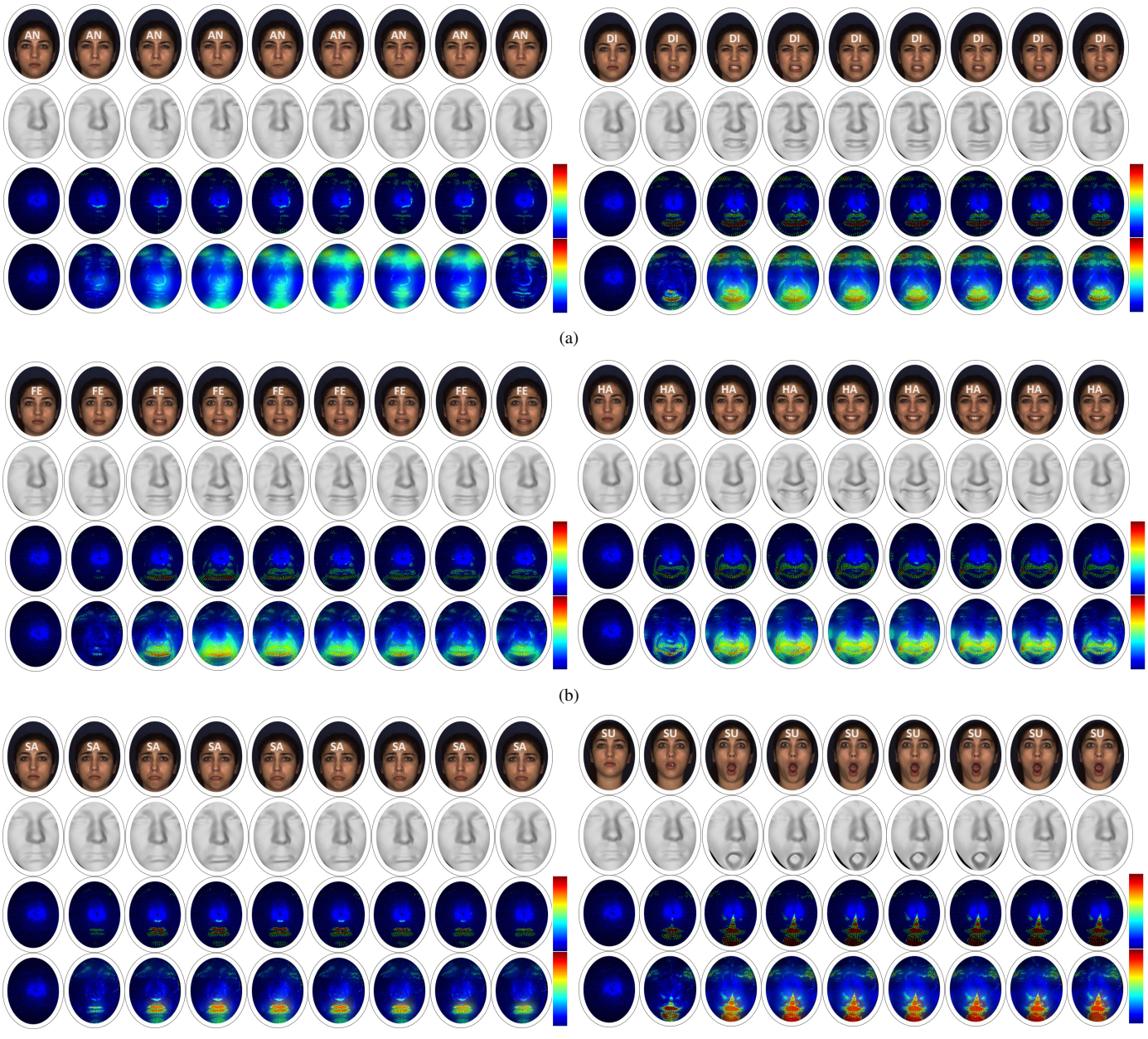
[10] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–8, 2012.

[11] V. Le, H. Tang, and T. Huang, "Expression recognition from 3d dynamic faces using robust spatio-temporal shape features," in *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 414–421.

[12] M. Reale, X. Zhang, and L. Yin, "Nebula feature: A space-time feature for posed and spontaneous 4D facial behavior analysis." in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2013, pp. 1–8.

[13] S. Berretti, A. Del Bimbo, and P. Pala, "Automatic facial expression recognition in real-time from dynamic sequences of 3d face scans," *The Visual Computer*, vol. 29, no. 12, pp. 1333–1350, 2013.

[14] 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.



(a)

(b)

(c)

Fig. 3: Illustrations of the deformation magnification on the sequences of the same subject performing the six universal expressions. One can appreciate the magnification effects on 3D deformations compared to those of the original *DSF* feature. From up to bottom, each row presents the texture image, the depth map, the original *DSF* feature, and the amplified feature, respectively.