High Quality Facial Expression Recognition in Video Streams using Shape Related Information only

Laszlo A. Jeni
The University of Tokyo, Japan
laszlo.jeni@ieee.org

Daniel Takacs
Realeyes Data Services Ltd., UK
takacsd@gmail.com

Andras Lorincz
Eotvos Lorand University, Hungary
andras.lorincz@elte.hu

Abstract

Person independent and pose invariant facial emotion classification is important for situation analysis and for automated video annotation. Shape and its changes are advantageous for these purposes. We estimated the potentials of shape measurements from the raw 2D shape *data of the CK+ database. We used a simple Procrustes transformation and applied the multi-class SVM leave-one-out method. We found close to 100% classification performance demonstrating the relevance of details in shape space. Precise, pose invariant 3D shape information can be computed by means of constrained local models (CLM). We used this method: we fitted 3D CLM to CK+ data and derived the frontal views of the 2D shapes. Performance reached and sometimes surpassed state-of-the-art results. In another experiment, we studied pose invariance: we rendered 3D emotional database with different poses using BU 4DFE database, fitted 3D CLM, transformed the 3D shape to frontal pose and evaluated the outputs of our classifier. Results show that the high quality classification is robust against pose variations. The superior performance suggests that shape, which is typically neglected or used only as side information in facial expression categorization, could make a good benchmark for future studies.*

1. Introduction

Our everyday communication is highly influenced by the emotional information available to us about those who we communicate with. Facial expression and body language are the main sources of this information. Thus, recognition of facial expression is highly relevant for human-computer interaction and may gain broad applications in video annotation, situation analysis of social interactions.

In the last decade many approaches have been proposed for automatic facial expression recognition. We are experiencing a breakthrough in this field due to the availability of high quality marked databases, like the Cohn-Kanade Extended Facial Expression Database (CK+) [4] and the advance of learning algorithms, most notably the advance of constrained local models (CLM) [2,7]. Recently, very good results have been achieved by means of textural information [5,9]. On the other hand, shape of the face extracted by active appearance models (AAM) (see, e.g., [4] and references therein) showed relatively poor performance.

Line drawings, however, can express facial expressions very well, so shape information could also be a good descriptor of emotions. Shape – as opposed to texture – is attractive for facial expression recognition since it should be robust against rotations and may be robust against light conditions that influence the texture of wrinkles.

We studied facial expression recognition using all available landmarks of the shape. We found close to 100% performance, indicating that the compression inherent in AAM was responsible for the relatively poor performance. We then used the more expressive CLM method and studied the behavior of CLM fits with respect to head pose directions. Our main result is that shape information and CLM based automated marker generation gives rise to, sometimes surpasses state-of-the-art performance and it could be improved further with more precise automated marker identification.

Beyond the theoretical interest that shape alone may give rise to 100% performance, we note that one may safely replace the estimated personal AU0 normalization by the easily measurable personal mean shape normalization and – as we show here – results outperform known results of texture and shape AAM [4], texture based CLM [9] and 2D shape based method using the Grassmann manifold [12]. In turn, we suggest shape identification for a standard benchmark.

The paper is built as follows. Theoretical components and the datasets are reviewed in Section 2 and 3, respectively. Results are detailed in Section 4 followed by some discussion and our summary in Section 5.

2. Theory

2.1. Active Appearance and Constrained Local Models

AAM and CLM methods are generative parametric models for person-independent face alignment. They usually apply either separated or joined shape and appearance models to generate candidate faces (for AAM) or region templates (for CLM) and use fast gradient algorithms in order to optimize them. The shape model of a 3D CLM, for example, is defined by a 3D mesh and in particular the 3D vertex locations of the mesh. Consider shape $\mathbf{s}$ of a 3D CLM as the coordinates of N 3D vertices that make up the mesh:

$$\mathbf{s} = (x_1, y_1, z_1 \ldots, x_N, y_N, z_N)^T \tag{1.1}$$

This model allows linear shape variation: shape $\mathbf{s}$ can be expressed as a base shape $\mathbf{s}_0$ plus a linear combination of m shape vectors $\mathbf{s}_i \in \mathbb{R}^{3N}$, $(i = 1, \ldots, m)$:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{m} p_i \mathbf{s}_i, \tag{1.2}$$

where coefficients p_i are the shape parameters and vectors $\mathbf{s}_i$ are set to orthonormal.

In our work, we used the 3D CLM method [7], which fits its model to an unseen image in an iterative manner by generating templates using the current parameter estimates, correlating the templates with the target image to generate response images and optimizing the shape parameters so as to maximize the sum of responses. The interested reader is referred to [2] and [7] for the details of the CLM algorithm. We note that the 3D CLM of [7] is using 6 rigid and 24 non-rigid parameters $(m = 30)$, where the non-rigid parameters are determined by principal component analysis (PCA) by starting from 66 marker points (also called landmarks), i.e., from 3*66=198 dimensions. On the other hand, the task of 2D AAM [6] is more demanding from the point of view of rigid parameters, but it typically applies about 10 parameters to represent texture and shape, respectively.

2.2. Procrustes' transformation

For any shape $\mathbf{s}$, Procrustes' transformation applies translation, uniform scaling and rotation to match the reference shape $\mathbf{s}_r$ in Euclidean norm. The minimum of this cost function is called the Procrustes distance. We applied this transformation in 2D.

2.3. Extracted Features

In 2D, we used the 2D shape coordinates of the CK+ database. There is a slight difference between the set of the CK+ marker points and the set of CLM marker points: the latter dropped marker points #60, #64 corresponding to the inner points of the left and right corners of the mouth, respectively. That is, we had 2*68=136 dimensional vectors for classification.

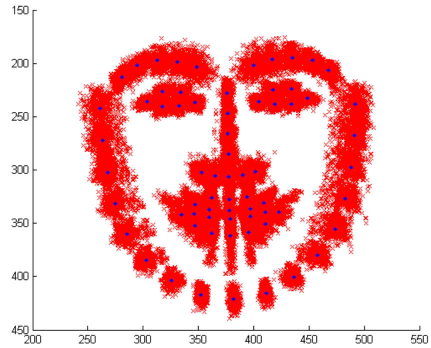

Figure 1: All shapes of the CK+ database fitted to the mean shape by Procrustes' transformation. Red crosses: marker points (or landmarks) of the individual shapes, blue dots: marker points of the mean shape.

We used the Procrustes method to compute the mean shape and normalize all shapes to this mean (Fig. 1). For the classification task, we used the so called AU0 normalization; we computed the differences between the features of the apex frame and the features of the first (neutral) frame.

We also used the 3D CLM that we fitted to the facial expression series of the CK+ database in order to have a good estimate of the rigid transformation parameters. Then we extracted the normalized 2D shape parameters by first removing the rigid transformation and then projecting to 2D.

In another set of experiments, we used the BU 4DFE dataset [3], rendered 3D facial expressions, and rotated those to different poses. This procedure was followed by the 3D CLM analysis and we extracted the 2D shape parameters alike in the previous case.

2.4. AU0 Normalization

AU0 normalization is crucial for shape based facial expression recognition, because it removes the personal variation, however it is person dependent and it is not available for a single frame. We assume that we have videos (frame series) about the subject like in the case of the BU 4DFE and we can compute the personal mean shape. We found that the mean shape is almost identical to the neutral shape, i.e., to AU0. Since the BU 4DFE contains highly distorted shapes of different kind, we believe that time averaged shapes are either close to neutral, or if not, the difference should be taken into account. Quantitative analysis of this conjecture with naïve subjects, however, is to be conducted in future studies.

We show in Section 4 how to replace the unavailable AU0 information by means of the CLM method.

2.5. Multi-class Support Vector Machine for Emotion Classification

Support Vector Machines (SVMs) are very powerful for binary and multi-class classification as well as for regression problems. They are robust against outliers. For two-class separation, SVM estimates the optimal separating hyper-plane between the two classes by maximizing the margin between the hyper-plane and closest points of the classes. The closest points of the classes are called support vectors; they determine the optimal separating hyper-plane, which lies at half distance between them.

We are given sample and label pairs $(\mathbf{x}^{(k)}, y^{(k)})$ with $\mathbf{x}^{(k)} \in \mathbb{R}^m$, $y^{(k)} \in \{-1, 1\}$, and $k = 1, \ldots, n$. Here, for class 1 and for class 2 $y_k = 1$ and $y_k = -1$, respectively. Assume further that we have a set of feature vectors $\phi_j : \mathbb{R}^m \to \mathbb{R}^M$, where M might be infinite. The support vector classification seeks to minimize the cost function

$$J(\mathbf{w}, b) = \frac{1}{2}||\mathbf{w}||^2, \tag{2.1}$$

where $\mathbf{w} \in \mathbb{R}^M$, subject to the condition

$$y^{(k)} \left(\sum_{i=1}^{M} w_i \phi_i(\mathbf{x}^{(k)}) - b \right) \geq 1 \tag{2.2}$$

for $k = 1, \ldots, m$. For a linear SVM, (4.2) simplifies to

$$y^{(k)} \left(\mathbf{w}^T \mathbf{x}^{(k)} - b \right) \geq 1. \tag{2.3}$$

In this study we used the LIBSVM software [1]. We used multi-class classification, where decision surfaces are computed for all class pairs, i.e., for k classes one has $k(k-1)/2$ decision surfaces and then applies a voting strategy for decisions. We note that multi-class SVM is considered competitive to other SVM methods [1]. In all cases, we used only linear classifiers.

3. Datasets

3.1. The Cohn-Kanade Extended Dataset

During the simulation we used the Cohn-Kanade Extended Facial Expression Database [4]. This database was developed for automated facial image analysis and synthesis and for perceptual studies. The database is used by researchers to compare the performance of their models.

The database contains 123 different subjects and 593 image sequences. From these, 118 subjects are annotated with the seven universal emotions (anger, contempt, disgust, fear, happy, sad and surprise). The image sequences are annotated with 68 landmark points. We used these landmark points as the “ground truth”.

Emotion	N
Angry	45
Contempt	18
Disgust	59
Fear	25
Happiness	69
Sadness	28
Surprise	83 (78)

Table 1: Distribution of emotion labels for 118 subjects with five samples (S010_002, S011_001, S055_001, S058_001) that we removed.

In the Procrustes experiment we removed S010_002, S011_001, S055_001, S058_001 and S124_001 image sequences (from surprise set), because the landmarks do not match the corresponding images in these sequences. In the CLM experiments, landmarks were provided by the CLM itself.

3.2. The BU-4DFE Dynamic Facial Expression Database

For our studies on pose dependence we used the BU-4DFE dataset [3]. This dataset is a high-resolution 3D dynamic facial expression database. It contains 3D video sequences of 101 different subjects, with a variety of ethnic/racial ancestries.

Each subject was asked to perform six prototypic facial expressions (anger, disgust, happiness, fear, sadness, and surprise), therefore the database contains 606 3D facial expression sequences.

In the pose invariant experiment we marked a neutral frame and an apex frame of each sequence and rendered short video sequences with different yaw rotations. For these sequences the landmarks were provided by the CLM tracker itself.

4. Experimental Results

4.1. Experiment on the CK+ dataset with Procrustes’ method and original landmarks

In this experiment we used the CK+ dataset with the original 68 CK+ landmarks. First, we calculated the mean shape using Procrustes’ method. Then we normalized all shapes by minimizing the Procrustes distance between individual shapes and the mean shape.

We trained a multi-class SVM using the leave-one-subject-out cross validation method. The result of the classification is shown in Table 2: emotions with large distortions, such as disgust, happiness and surprise, gave rise to nearly 100% classification performance. Even for the worst case (fear), performance was 92%. This is comparable to human performance [8].

	An.	Co.	Di.	Fe.	Ha.	Sa.	Su.
Anger	93.3	2.2	4.4	0	0	0	0
Contempt	0	94.4	0	0	0	5.6	0
Disgust	0	0	100	0	0	0	0
Fear	0	0	0	92	8	0	0
Happy	0	1. 5	0	0	98.5	0	0
Sadness	0	3.6	0	0	0	96.4	0
Surprise	0	2.6	0	0	0	0	97.4

Table 2: Confusion matrix for the Procrustes method with AU0 normalization shown both in a figure and in a table averaged for the 118 subjects of the CK+ database.

	An.	Co.	Di.	Fe.	Ha.	Sa.	Su.
Anger	95.6	0	2.2	0	0	2.2	0
Contempt	5.6	94.4	0	0	0	0	0
Disgust	0	0	100	0	0	0	0
Fear	0	4	0	80	8	4	4
Happy	0	0	0	0	100	0	0
Sadness	0	0	0	3.5	0	96.4	0
Surprise	0	1.3	0	1.3	0	0	97.4

Table 3: Confusion matrix using the personal mean shape instead of the AU0 normalization shown both in a figure and in a table averaged for the 118 subjects of the CK+ database.

	An.	Co.	Di.	Fe.	Ha.	Sa.	Su.
Anger	73.3	0	17.8	2.2	0	6.7	0
Contempt	5.6	72.2	0	0	0	16.7	5.6
Disgust	8.5	0	89.8	0	1.7	0	0
Fear	4	4	4	68	8	12	0
Happy	1.4	0	2.9	0	95.7	0	0
Sadness	17.9	7.1	0	14.3	0	50	10.7
Surprise	0	1.2	0	2.4	0	2.4	94

Table 4: Confusion matrix for the 3D CLM method with AU0 normalization shown both in a figure and in a table averaged for the 118 subjects of the CK+ database.

	An.	Co.	Di.	Fe.	Ha.	Sa.	Su.
Anger	77.8	0	13.3	0	2.2	6.7	0
Contempt	0	94.4	0	0	0	5.6	0
Disgust	6.8	0	91.5	0	0	1.7	0
Fear	0	8	0	80	4	4	4
Happy	0	1.4	0	0	98.6	0	0
Sadness	14.3	10.7	0	3.6	0	67.9	3.6
Surprise	1.2	1.2	0	0	0	0	97.6

Table 5: Confusion matrix for the 3D CLM method using the personal mean shape instead of the AU0 normalization averaged for the 118 subjects of the CK+ database.

Replacing AU0 normalization by personal mean shape slightly decreases average performance: recognition on the CK+ database drops from 96% to 94.8% (see Table 2 and 3). As a point of reference, we note that without any AU0 normalization, average recognition rate reached 88.6% only.

4.2. Experiment on the CLM-tracked CK+ dataset

In this experiment we studied the performance of the multi-class SVM using CLM method on the CK+ dataset.

First, we tracked facial expressions with the CLM tracker and annotated all image sequences starting from the neutral expression to the peak of the emotion. 3D CLM estimates the rigid and non-rigid transformations. We removed the rigid ones from the faces and projected the frontal view to 2D. We also applied the Procrustes normalization in this experiment. We did not find significant differences between the two cases and report the results without Procrustes normalization.

The recognition performance of the system can be seen in Table 4: classification performance is affected by imprecision of the CLM tracking.

Emotions with large distortions can still be recognized in about 90% of the cases, whereas more subtle emotions are sometimes confused with others.

We evaluated the personal mean shape normalization (see Table 5). We found that this method compensates for the estimation error of the CLM method. Correct classification percentage rises from 77.57% to 86.82% for the CLM tracked CK+. This result is better than the available best AAM result that uses texture plus shape information [4] and the best CLM result that utilizes only textural information [9]. There is comparison between different learning methods, including SVMs and the Grassmann manifold, showing that considerable improvements can be gained by the latter [12]. We note that the results in [12] were reported on the first version of the CK dataset with different emotion labels so direct comparison is misleading. Table 6 gives an overview of these works.

Methods based on 3D shape promise robustness against head pose. Our studies on pose invariance are detailed in the next subsections.

		An.	Co.	Di.	Fe.	Ha.	Sa.	Su	Avg.
Taheri et al [12]	E-LDA	60	–	76.3	70.2	91.3	75	96.1	78.2
	G-LDA	68	–	80.5	74.4	88.9	78.2	97.3	81.2
	G-KLDA	65.7	–	86.8	83	95.1	85.7	98.6	85.8
	E-SVM	62.8	–	78.9	74.4	91.3	80.3	97.2	80.8
	G-SVM	65.7	–	78.9	74.5	95	85.7	97.2	82.8
Lucey et al. [4]	AAM-SVM(S)	35	25	68.4	21.7	98.4	4	100	50.3
	AAM-SVM(T)	70	21.9	94.7	21.7	100	60	98.7	66.7
	AAM-SVM(ST)	75	84.4	94.7	65.2	100	68	96	83.3
Chew et al. [9]	CLM-SVM(T)	70.1	52.4	92.5	72.1	94.2	45.9	93.6	74.4
This work	CLM-SVM(S)	73.3	72.2	89.8	68	95.7	50	94	77.6
This work	CLM-SVM(S)*	77.8	94.4	91.5	80	98.6	67.9	97.6	86.8

(S) - shape information (T) - texture information
* - personal mean shape normalization

Table 6: Comparison of results on CK dataset.

	An.	Di.	Fe.	Ha.	Sa.	Su.
Anger	71.8	10.6	0	2.6	15.4	0
Disgust	23.1	61.5	0	12.8	0	2.6
Fear	0	5.1	61.5	7.7	12.8	12.8
Happy	5.1	5.1	7.7	79.5	2.6	0
Sadness	20.5	5.1	15.4	0	59	0
Surprise	0	2.6	7.9	0	0	89.5

Table 7: Confusion matrix for the BU-4DFE dataset using AU0 normalization.

	An.	Di.	Fe.	Ha.	Sa.	Su.
Anger	82.1	10.3	0	0	7.7	0
Disgust	12.8	74.4	2.6	7.7	2.6	0
Fear	0	5.1	61.5	7.7	15.4	10.3
Happy	2.6	5.1	5.1	87.2	0	0
Sadness	17.9	2.6	5.1	2.6	71.8	0
Surprise	0	0	7.9	0	0	92.1

Table 8: Confusion matrix for the BU-4DFE dataset using personal mean shape instead of AU0 normalization.

	An.	Co.	Di.	Fe.	Ha.	Sa.	Su.
Anger	56.4	5.1	28.2	2.6	0	7.7	0
Contempt	-	-	-	-	-	-	-
Disgust	20.5	0	61.5	7.7	7.7	0	2.6
Fear	5.1	23.1	0	38.5	5.1	12.8	15.4
Happy	2.6	7.7	0	0	89.7	0	0
Sadness	20.5	17.9	0	2.6	0	59	0
Surprise	0	2.6	0	7.9	0	2.6	86.8

Table 9: Confusion matrix with training on the CK+ and testing on the BU-4DFE dataset.

4.3. Experiment on the CLM-tracked BU-4DFE dataset

We characterized the BU-4DFE database by using the CLM technique. First, we selected a frame with neutral expression and an apex frame of the same frame series. We used these frames and all frames between them for the evaluations. We applied CLM tracking for the intermediate frames in order, since it is more robust than applying CLM independently for each frames. We removed the rigid transformation after the fit and projected the frontal 3D shapes to 2D. We applied a 6 class multi-class SVM (the BU-4DFE database does not contain contempt) and evaluated the classifiers by the leave-one-subject-out method. We compared the normalization using the CLM estimation of the AU0 values with the normalization based on the personal mean shape. Note that for the expert annotation, i.e., for the CK+ database, performance drop for personal mean shape was only 1% in average (see Tables 2 and 3).

For the BU-4DFE database, however, we found an 8% improvement on the average in favor of the mean shape method (see Tables 7 and 8).

We executed cross evaluations (Table 9). We used the CK+ as the ground truth, since it seemed more precise: the target expression for each sequence is fully FACS coded, emotion labels have been revised and validated, and CK+ utilizes FACS coding based emotion evaluation and this method is preferred in the literature [4].

We note however, that both the CK+ and the BU-4DFE facial expressions are posed and not spontaneous. We will return to this point in the next section. Our results depicted in Table 9 show considerable discrepancies between the two databases.

4.4. Pose Invariant Experiment using BU dataset

In this experiment we studied CLM's performance as a function of pose, since we are interested in pose invariant emotion recognition for situation analysis. We used the BU-4DFE dataset to render 3D faces with six emotions (anger, disgust, fear, happiness, sadness, and surprise), which are available in the database. We randomly selected 25 subjects and we rendered rotated versions of every emotion. We covered rotation angles between 0 and 44 degrees of anti-clockwise rotation around the yaw axis.

As illustrated on the subfigures of Fig. 3, CLM based classification is robust against large pose variations, including the hard cases like anger. However, misclassification types change as a function of angle.

As illustrated on Fig. 4, as the angle of rotation increases, the error of the landmark position estimation accumulates. It may reach 10 RMSE unit on average (1 pixel error for all landmarks corresponds to 1 RMSE unit.) This error influences emotion recognitions only slightly.

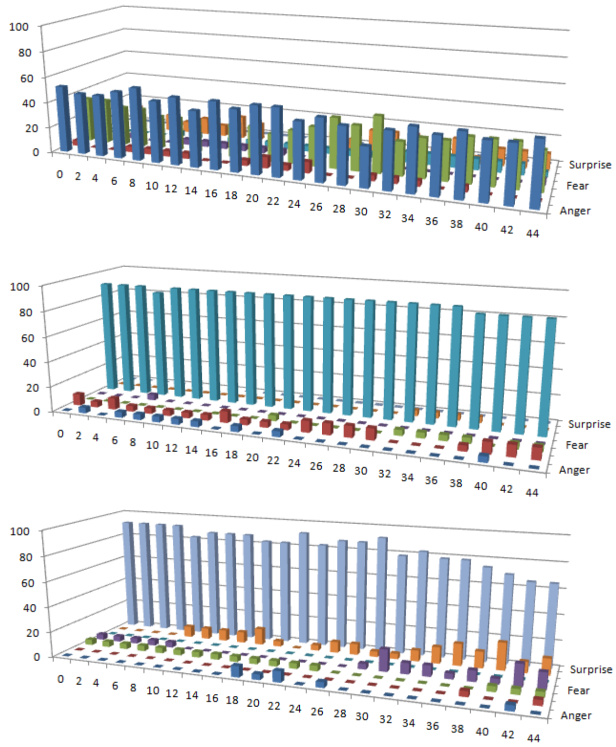

Figure 3: Measurements on pose dependence using the BU-4DFE database. Top: classification as a function of rotational degree for angry faces. Middle: Same for happy faces. Bottom: Same for surprised faces.

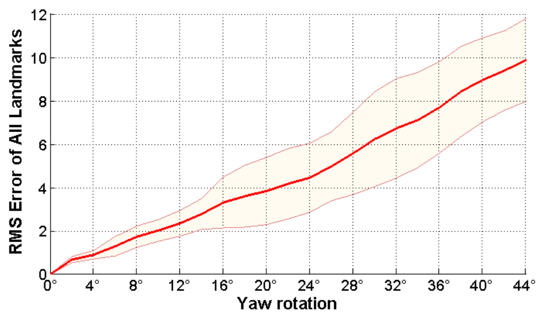

Figure 4: RMSE of reconstructed CLM estimations of the landmark positions of the 2D meshes in pixels as a function of rotation angle and using every landmark. The bold line denotes the mean. Shaded region around the mean shows the standard deviation. Distortion was compared to the initial frame (0 degrees of yaw rotation). 1 pixel error for all landmarks corresponds to 1 RMSE unit.

Other works also address pose invariant recognition on the BU dataset, like [13,14,15,16], however they are using either the static version of the database [13,14] or only a limited number of subjects and emotion categories from the dynamic dataset [15,16] and therefore direct comparison cannot be provided.

5. Discussion and Conclusions

Motivated by highly precise 2D facial expression classification using shape information only, we engaged in similar estimations using the recent 3D CLM method.

We used a number of methods to study performance of shape representations for facial expression recognition. In all studies, we applied multi-class SVM classification [1]. We used expert annotated frontal databases as well as 3D dynamic datasets [3,4].

We found that full shape information, without PCA compression and with Procrustes normalization gives rise to excellent results, similar to human performance [8] so this direction is highly promising.

We used CLM method to extract shape data, since it is more precise and may preserve more information than the AAM method [4]. We received state-of-the-art results for the CK+ database based on the shape information we extracted.

We replaced normalization using an estimation of the AU0 parameters with the personal mean shape that gave rise to considerable improvements. We think that the difference is in the noise of CLM based AU0 estimation, which is larger than the discrepancy between AU0 values (as determined by the experts) and mean shape values (as determined by averaging over the shapes of the same person). We received very good performance with shape information alone. Our results surpass performances of the best available AAM [4] and CLM [9] methods utilizing shape plus texture and texture information, respectively.

This method has practical values since – as we showed – the average shape for all facial expressions is very close to the AU0 values and the average face can be computed by averaging over time.

From the point of view of situation analysis and human-computer interaction, angle dependence of facial expression recognition is of great importance. We studied the robustness of the CLM method for yaw rotations. We rendered rotated 3D faces using the BU-4DFE database [3] and found that CLM based shape estimation and shape based emotion recognition are highly robust against such pose variations. However, we note that both the CK+ and the BU 4DFE databases contain posed facial expressions that may differ considerably from the natural ones [11].

We suspect that CLM based shape estimations may also be robust against light conditions due to the strength of approach that multiples probability estimations of experts [10].

In sum, shape information is very efficient for facial expression recognition provided that details of shape changes are determined precisely. 3D CLM method is promising in this respect, since 3D CLM shape estimation is robust against pose variations as we showed here. This can be of high value in situation analysis. We also suspect that CLM based shape estimation is robust against light variations due to the multiplication of probability estimations of different experts, each of which can be trained to resist to light variations. In turn we suggest that shape based facial expression recognition could be used as a benchmark to measure progress of the different methods. This benchmark can also be extended by texture based algorithms to improve performance.

Acknowledgments

We are grateful to Jason Saragih for providing his CLM code for our studies. The European Union, the European Social Fund and the European Regional Development Fund have provided financial support to the project under the grant agreements no. TÁMOP 4.2.1/ B-09/ 1/ KMR-2010-0003 and KMOP-1.1.2-08/1-2008-0002..

References

[1] C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.

[2] D. Cristinacce and T. Cootes, Automatic feature localisation with constrained local models, *Pattern Recognition*, 41: 3054-3067, 2008.

[3] Lijun Yin; Xiaochen Chen; Yi Sun; Tony Worm; Michael Reale: "A High-Resolution 3D Dynamic Facial Expression Database", *The 8th International Conference on Automatic Face and Gesture Recognition (FGR08)*, 17-19 September 2008 (Tracking Number: 66)

[4] 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: *3rd IEEE Workshop on CVPR for Human Communicative Behavior Analysis*, (CVPR4HB 2010).

[5] P. Lucey, J. F. Cohn, K. M. Prkachin, P. Solomon, and I. Matthews, Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database. *9th IEEE Int. Conf. on Automatic Face and Gesture Recogn*. (FG2011).

[6] I. Matthews and S. Baker. Active appearance models revisited. *Int. J. Comp. Vision,* 60(2): 135-164, 2004.

[7] J. M. Saragih, S. Lucey, and J. F. Cohn, Deformable model fitting by regularized landmark mean-shift. *Int. J. Comp. Vision,* 91(2): 200-215, 2011.

[8] J. M. Susskind, G. Littlewort, M. S. Bartlett, J. Movellan, and A. K. Anderson, Human and computer recognition of facial expressions of emotion, *Neuropsychologia*, 45(1): 152-162, 2007.

[9] S. W. Chew, P. J. Lucey, S. Lucey, J. Saragih, J. F. Cohn and S. Sridharan, Person-independent facial expression detection using constrained local models. *In Proceedings of FG 2011 Facial Expression Recognition and Analysis Challenge*, Santa Barbara, CA, 2011.

[10] G. E. Hinton, Training Products of Experts by Minimizing Contrastive Divergence, *Neural Computation*, 14(8) 1771-1800, 2002.

[11] T. McLellan, L. Johnston. J. Dalrymple-Alford, and R. Porter, Sensitivity to genuine versus posed emotion specified in facial displays, *Cognition and Emotion,* 24(8): 1277-1292, 2010.

[12] Sima Taheri, Pavan Turaga, Rama Chellappa, "Towards View-Invariant Expression Analysis Using Analytic Shape Manifolds," *Int. Conf. Automatic Face and Gesture Recognition*, 2011

[13] W. Zheng, H. Tang, Z. Lin, and T. S. Huang, "Emotion recognition from arbitrary view facial images," *European Conf. Computer vision*, 2010, pp. 490–503.

[14] O. Rudovic, I. Patras, and M. Pantic, "Coupled gaussian process regression for pose-invariant facial expression recognition," *European Conf. Computer vision*, 2010, pp. 350–363.

[15] G. Sandbach, S. Zafeiriou, M. Pantic, D. Rueckert, "A dynamic approach to the recognition of 3D facial expressions and their temporal models," *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* , vol., no., pp.406-413, 21-25 March 2011.

[16] Vuong Le, Hao Tang, Huang, T.S., "Expression recognition from 3D dynamic faces using robust spatio-temporal shape features," *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* , vol., no., pp.414-421, 21-25 March 2011