Efficient, Pose Invariant Facial Emotion Classification using 3D Constrained Local Model and 2D Shape Information

Laszlo A. Jeni
Dept. of Electrical Engineering
University of Tokyo
laszlo.jeni@ieee.org

Hideki Hashimoto
Dept. of Electrical Engineering
Chuo University
hashimoto@elect.chuo-u.ac.jp

Andras Lorincz
Dept. of Softw Techn & Meths
Eotvos Lorand University
andras.lorincz@elte.hu

Abstract

Pose invariant facial emotion classification is important for situation analysis and for automated video annotation. We started from the raw 2D shape data of the CK+ database and used a simple Procrustes transformation and the multi-class SVM leave-one-out method for classification. We found close to 100% performance demonstrating the potentials of shape based methods. We applied a 3D constrained local model (CLM) and generated a 3D emotionally modulated database with different poses using FaceGen. We fitted 3D CLM and used it in an iterative manner to exclude the potentially occluded landmarks. We transformed the 3D shape to frontal pose and evaluated the outputs of our classifier. Excellent pose invariant performance with considerable improvement over the non-iterative method was achieved.

1. Introduction

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) [2,6]. Recently, very good results have been achieved by means of textural information [5], while shape of the face extracted by active appearance models (AAM) showed relatively poor performance.

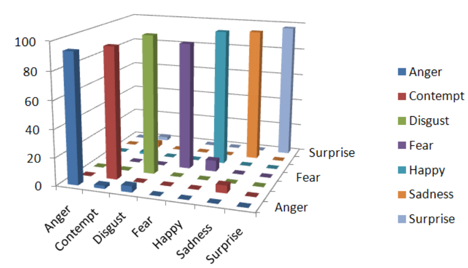
Line drawings, however, can express facial expressions very well, so shape information should also be a good descriptor of emotions. Shape – as opposed to texture – is attractive since it should be robust against light conditions and pose variations.

Using precise shape information of frontal views and applying the Procrustes method, we found close to 100% performance on the CK+ database. Our main results are that (i) 2D shape information gives close to 100% performance and (ii) 3D CLM based iterative facial expression recognition compensates for pose variations robustly with excellent performance on an artificial database .

2. Experimental Results

2.1. Procrustes for frontal views

We used the Cohn-Kanade Extended Facial Expression Database (CK+) [4] with the original 68 CK+ landmarks, calculated the mean shape and normalized all shapes by minimizing the Procrustes distance to it. We trained a multi-class SVM [1] using the leave-one-subject-out cross validation method. Results of Table 1 show that shape is an excellent gauge if information is precise in 2D.



	An.	Co.	Di.	Fe.	Ha.	Sa.	Su.
Anger	**93.4**	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 1: Confusion matrix for the Procrustes method shown in a figure (upper) and in a table (lower) for the 118 subjects of the CK+ database.

2.2. Algorithmic design for 3D

We generated an artificial database with FaceGen [3].

First algorithm. We fitted 3D CLM to faces of artificial database. We rotated the shape to frontal pose, projected it to 2D and used a 2D classifier trained for FaceGen.

Second algorithm. We dropped unstable marker points of the profile and then fitted the 3D CLM. We determined the approximate pose, identified and dropped the occluded marker points, fitted CLM, rotated the shape to the frontal pose, projected it to 2D and used the 2D classifier.

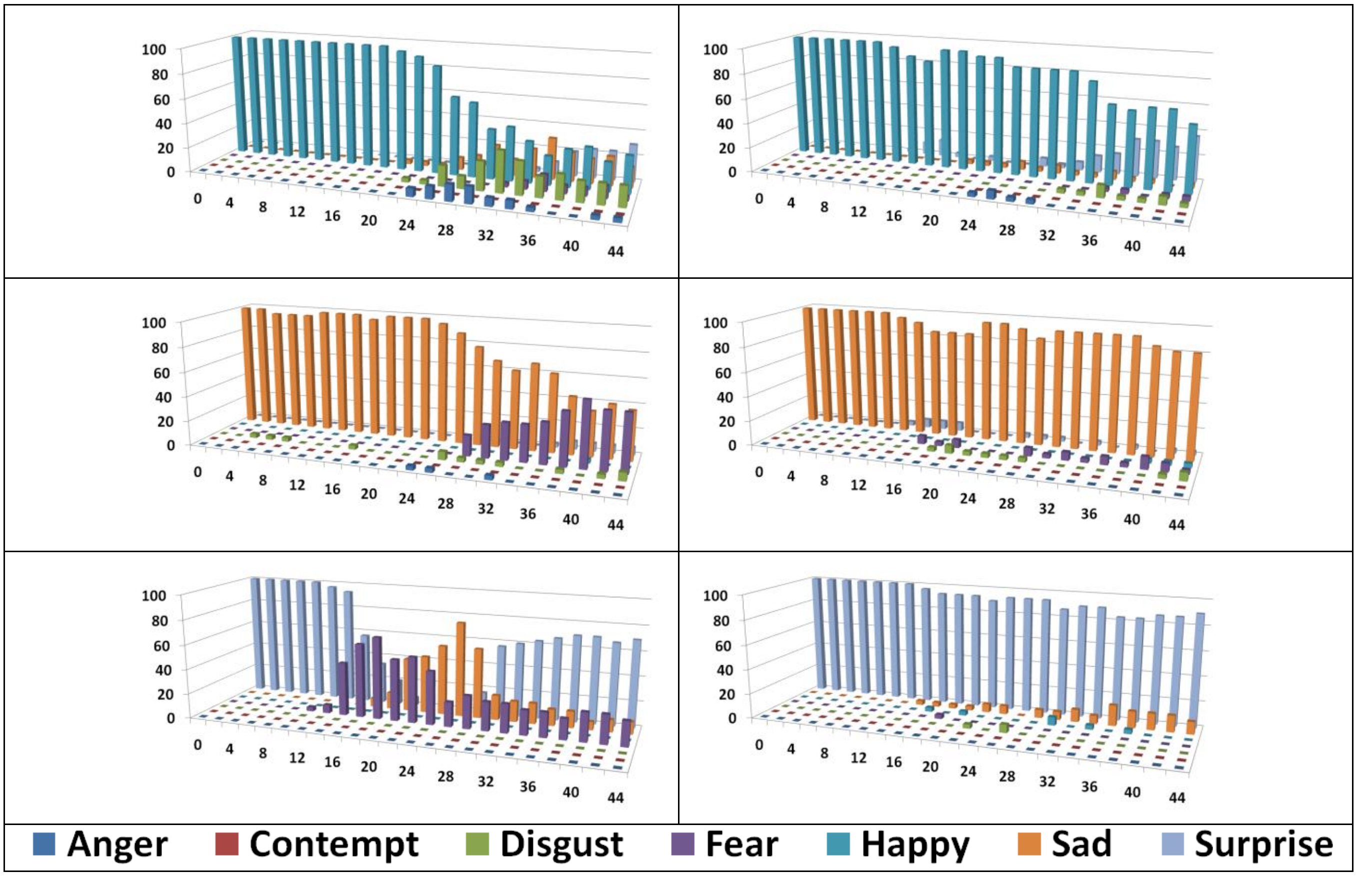


Fig. 1: Decision probabilities as a function of yaw angle for different FaceGen emotions with a multi-class SVM classifier. **Left:** (1) estimate pose with CLM, (2) rotate to frontal pose, (3) project to 2D, (4) decide. **Right:** (1) drop profile markers, (2) estimate pose with CLM, (3) drop (potentially) occluded markers, (4) re-estimate shape and pose with CLM, (4) rotate to frontal pose, (5) fill-in missing markers with PCA, (6) project to 2D, (7) decide. Top: happy, middle: sad, bottom: surprise.

3. Results and discussion

In the experiment we used FaceGen generated database and developed a new classifier. Results were somewhat better than those shown in Table 1. We note that 1. FaceGen and CK+ expressions differ; FaceGen expressions are moderate, 2. at around 24 degrees of yaw angle, CLM switched to a model trained for the half of the face (Fig. 1).

Results show that some of the facial expressions are robust against yaw angle variations up to about 20 degrees without and up to 40-44 degrees with removal of pose estimated exclusion of occluded landmarks. Switch of the model also improves results (not shown). Further improvements are expected when shape and texture are combined or for higher precision CLM models.

Acknowledgement. We are grateful to Jason Saragih for providing his CLM code for our studies. The European Union and the European Social Fund have provided financial support to the project under the grant agreement no. TÁMOP 4.2.1./ B-09/ 1/ KMR- 2010-0003.

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] FaceGen, Singular Inversions, http://en.wikipedia.org/wiki/FaceGen

[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), in press.

[6] 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.