# Combining appearance and geometric features for facial expression recognition

Hui  Yu
Honghai  Liu

# Combining Appearance and Geometric Features for Facial Expression Recognition

Hui Yu, Honghai Liu
University of Portsmouth, Portsmouth, UK

## ABSTRACT

This paper introduces a method for facial expression recognition combining appearance and geometric facial features. The proposed framework consistently combines multiple facial representations at both global and local levels. First, covariance descriptors are computed to represent regional features combining various feature information with a low dimensionality. Then geometric features are detected to provide a general facial movement description of the facial expression. These appearance and geometric features are combined to form a vector representation of the facial expression. The proposed method is tested on the CK+ database and shows encouraging performance.

**Keywords:** Geometric features, covariance descriptors, facial expression, facial patches.

## 1. INTRODUCTION

Facial expression recognition has been an active research topic in over last twenty years. It has various applications including robotics, emotion analysis, image understanding and facial animation among others [1, 2]. Automatic facial expression recognition can help understand emotional activities of the target identities with relatively less intrusiveness than most other contacted methods. Human-machine interaction is one of the main applications of facial expression recognition since facial expressions convey significant human emotions. According to early psychological research, 55% communicative cues can be judged from facial expressions [3]. There is extensive research in facial expression recognition and understanding since the pioneering work by Mase et al. [4]. The current work can be generally classified into two categories: appearance based methods and geometric based methods. A more detailed survey on facial expression recognition can be found in [5].

Appearance based methods usually extract the textural variations of face images through various descriptors including Principal Component Analysis (PCA) [6], Independent Component Analysis (ICA) [7] and Gabor filters [8]. Garber filter provide better performance than many state-of-the-art methods [9], but it subjects to expensive computational cost in both memory and time. Local Binary Patterns (LBP) has been recently proposed as effective appearance descriptors for facial analysis [10, 11]. It achieved a convincing performance compared with Gabor filters but with a light computational cost [12, 13]. A variation of LBP has been proposed by Rajamanoharan et al. [14] for 3D facial action unit detection. Different from LBP, Jiang et al. [15] recently proposed to use irregular patches based on domain knowledge to define face regions for facial action recognition.

Geometric methods represent facial movements using predefined geometric landmark positions on salient facial features [16, 17]. Since this kind of methods describes facial movements based on limited fiducial points on the face, it required accurate facial feature detection.

Facial Action Coding System (FACS) has attracted attention since it was invented by Ekman et al. [18] in 1970s. FACS postulated six primary emotions each possessing distinctive contents together with a unique facial expression. These facial expressions are referred to as basic emotional facial expression which used to be thought as universal across human ethnicity and cultures, but recent research suggested these facial expressions are not culturally universal [19, 20, 21]. FACS action unit has been widely used in facial action recognition and optimization [22, 23, 24, 25, 26]. One potential disadvantage of AU based methods is that the errors accumulated from AU classification may propagate which can affect the performance of facial expression recognition.

With the development of technologies and higher demands for human-machine interaction, it requires a higher precision in recognizing effective facial expressions. Though much progress has been made in recent years, it is still challenging in improving the recognition accuracy. In this paper, we propose a framework to effectively combine the appearance descriptors and geometric features of the image for facial expression recognition. A covariance matrix for
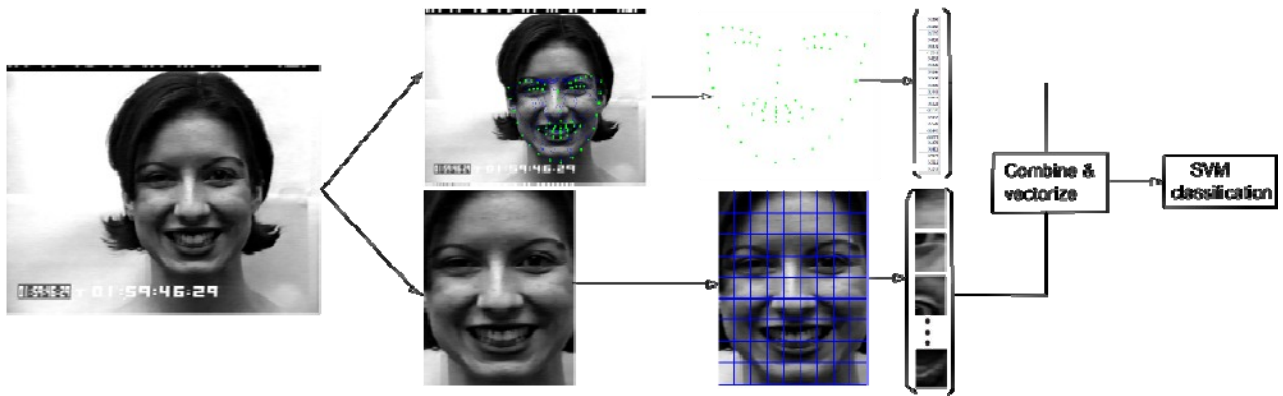
the ordering and patch size of the image. The covariance matrices have low dimension compared with many feature descriptors since the symmetry matrix $C_{i,p}$ contains only $N = \frac{1}{2}(d^2 + d)$ different values. The noise corrupting individual samples can be largely filtered out with an average filter during the covariance computation. In this paper, the covariance descriptor is calculated based on an eight-dimensional feature vector for each pixel:

$$f(x,y) = \left[ x, y, I, I_x, I_y, \sqrt{I_x^2 + I_y^2}, I_{xx}, I_{yy} \right] \tag{3}$$

This feature vector contains the pixel coordinate $(x, y)$, the gray scale pixel value I at $(x, y)$, the derivative $I_x, I_y$ along x and y respectively, the second derivative $I_{xx}, I_{yy}$ along x and y respectively and the magnitude of the gradient $\sqrt{I_x^2 + I_y^2}$ .

To eliminate the effect of some large values, the Covariance descriptors are normalized for each patch. The descriptors are then vectorized as a unique vector for each face image. This vector is normalized and can be expressed as: $C = [\overline{C}_1; \overline{C}_2, ..., \overline{C}_M]$ , Where $\overline{C}_p$ is the normalized covariance descriptor for the $p^{th}$ patch, M being the total number of patches.

Inspired by the distance measure method to compute the dissimilarity of two covariance matrices by Forstner et al. [28], we compute the difference of the covariance descriptor pair of the two decomposed layers:

$$\rho_p(C_{p,1}, C_{p,2}) = \sqrt{\sum_{i=1}^{m} \ln^2 \lambda_{p,i}(C_{p,1}, C_{p,2})} \tag{4}$$

Where $\lambda_{p,i}(C_{p,1}, C_{p,2})$ $(i = 1,2,...,m)$ are the generalized eigenvalues of $C_{p,1}$ and $C_{p,2}$ . The covariance descriptor distance items of all the patches of the face image are stacked into a vector and normalized as $\rho = [\overline{\rho}_1, \overline{\rho}_2, ..., \overline{\rho}_N]$, $N = d \times d$ .

## 2.2 Geometric features

We employ the landmark detection method by Yu et al. [29] for detecting feature points on the face. Since the landmarks on the face profile carry little information in facial expressions, we only take the landmarks around eyes, nose and mouth as the geometric feature points. Fig. 1 shows an example of the detected facial features. After detecting the landmarks, Procrustes transformation is applied to align with the reference face shape. The aligned shape features are then normalized and vectorized for the combination with the covariance descriptors.

To improve the performance of the recognition, we just concatenate the appearance vector and geometric vector into one single vector representation for each image. The geometric vector is then normalized as a vector $G$ .

These appearance and geometric features are then concatenated to form one vector for the representation of the face image:

$$F = [C; \rho; G] \tag{5}$$

## 3. EXPERIMENT

The proposed method is evaluated using the Cohn-Kanade Extended Facial Expression Database (CK+) [30].

Fig. 2. Examples of the six primary facial expressions from the CK+ database



Fig. 3. An example of facial image patches division. Left block: the detected surprise facial expression; right block: divided 64 patches of the facial expression.

Fig. 2 illustrates examples of six types of facial expressions. We only use the peak frame for each facial expression. This database consists of 123 subjects and 593 facial expression sequences in frontal view. Among those 118 subjects are annotated with seven primary facial expressions including surprise, anger, fear, happy, sad, disgust and contempt. In this paper we only test the algorithm for the first 6 facial expressions.

In the experiment, we use SVM as the classifier and the leave-one-out cross validation method [31]. Each facial expression image is divided into 64 patches as shown in Fig. 3. We compare the proposed method with the state-of-the-art LBP method. The experiment on the CK+ facial expression database shows a superior performance to the LBP method. Table 1 demonstrates the comparison result.

TABLE I Comparison of facial expression recognition methods on CK+ database

|  | Surprise | Sad | Fear | Anger | Disgust | Happy |
|---|---|---|---|---|---|---|
| LBP | 0.91 | 0.89 | 0.96 | 0.91 | 0.93 | 0.90 |
| Ours | 0.99 | 0.93 | 0.96 | 0.93 | 0.95 | 1.00 |

## 4. CONCLUSION

In this paper, we propose to recognize facial expressions using a combination of appearance and geometric features of the peak frames. The Covariance descriptors of appearance itself have shown good performance in the recognition of six primary facial expressions. The experiment demonstrated that when combining the geometric features of the facial expression, the performance can be clearly improved. In future, we will test the proposed method on more databases. As research suggested dynamics of facial expression may improve the recognition performance [32], we will improve the algorithm for dynamic facial expression analysis.

## REFERENCES

[1] P. S., Aleksic, & Katsaggelos, A. K. (2006). Automatic facial expression recognition using facial animation parameters and multistream hmms. IEEE Transactions on Information Forensics and Security, 1(1), 3–11.

[2] H., Yu, H. Liu, (2014) Regression-Based Facial Expression Optimization, , IEEE Transactions on Human-Machine Systems, Vol 44 Issue 3, pp 386-394;

[3] A, Mehrabian. "Communication without Words", Psychology Today, 1968. Vo1.2, No.4, pp 53-56.

[4] K., Mase, & Pentland, A. (1991). Recognition of facial expression from optical flow. IEICE Transactions, 74(10), 3474–3483

[5] Z. Zeng, Pantic, M.; Roisman, G.I; Huang, T.S., "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.31, no.1, pp.39,58, Jan. 2009

[6] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition.Journal of Cognitive Neuroscience, 1, 71–86

[7] P. N., Belhumeur, Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711–720

[8] M. J. Lyons, , J. Budynek& Akamatsu, S. (1999). Automatic classification of single facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence,21(12), 1357–1362.

[9] S. Xiea, , S. Shana,., Chena, X., Mengc, X., & Gao, W. (2009). Learned local gabor patterns for face representation and recognition. Signal Processing, 89(12), 2333–2344.

[10] T., Ojala, M., Pietikainen, & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distribution.Pattern Recognition,29(1), 51–59

[11] Tan, X., & Triggs, B. (2007). Fusing gabor and lbp feature sets for kernel-based face recognition. In Proceedings of the third international workshop analysis and modelling of faces and gestures(pp. 235–249), Rio de Janeiro, Brazil.

[12] C. Shan, S. Gong, & McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 27,803–816.

[13] Zavaschi, T., Oliveira, L., & Koerich, A. (2011). Facial expression recognition using ensemble of classifiers. In Proceedings of 36th international conference on acoustics, speech and signal processing(pp. 1489–1492).

[14] G. Rajamanoharan, S. Zafeiriou, M. Pantic. Local Normal Binary Patterns for 3D Facial Action Unit Detection, Proceedings of the IEEE International Conference on Image Processing (ICIP 2012). Orlando, FL, USA, pp. 1813 - 1816, October 2012.

[15] B. Jiang, B. Martinez, M. F. Valstar, M. Pantic. Decision level fusion of domain specific regions for facial action recognition, International Conference on Pattern Recognition. Stockholm, Sweden, 2014.

[16] Besinger, A., Sztynda, T., Lal, S., Duthoit, C., Agbinya, J., Jap, B., et al. (2010). Optical flow based analyses to detect emotion from human facial image data.Expert Systems with Applications, 37, 8897–8902.

[17] B. Martinez, M. F. Valstar, X. Binefa, M. Pantic. Local Evidence Aggregation for Regression Based Facial Point Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence. 35(5): pp. 1149 - 1163, 2013.

[18] P., Ekman, & Friesen, W. (1971). Constants across cultures in the face and emotion. Journal of Personality Social Psychology, 17(2), 124–129

[19] R., Jack, Garrod, O., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. Proceedings of the National Academy of Sciences, 109 (19), 7241-4.

[20] R., Jack, O., Garrod , H.,Yu,, R., Caldara, and P.,Schyns, (2013) Reply to Sauter and Eisner: differences outweigh commonalities in the communication of emotions across human cultures. Proceedings of the National Academy of Sciences, 110 (3). E181-E182. ISSN 1091-6490.

[21] H., Yu, O.G.B. Garrod, and Schyns P.G. (2012) Perception-Driven Facial Expression Synthesis, Computer & Graphics, Vol. 36, Issue 3, May 2012, pp152-162.

[22] M. Valstar, M. Pantic, Fully automatic recognition of the temporal phases of facial actions, IEEE Trans. Syst. Man Cybern., Part B 42 (2012) 28–43

[23] M. Pantic, I. Patras, Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences, IEEE Trans. Syst. Man Cybern., Part B 36 (2006) 433–449.

[24] Y. Tong, W. Liao, Q. Ji, Facial action unit recognition by exploiting their dynamic and semantic relationships, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 1683–1699.

[25] J. J. Wong, & S.Y. Cho. (2009). A local experts organization model with application to face emotion recognition. Expert Systems with Applications, 36, 804–819

[26] Yu, H., Liu, H. (2014) Regression-Based Facial Expression Optimization, , IEEE Transactions on Human-Machine Systems, Vol 44 Issue 3, pp 386-394;.

[27] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In A. Leonardis, H. Bischof, and A. Pinz, editors, Computer Vision ECCV 2006, volume 3952 of Lecture Notes in Computer Science, pages 589–600. Springer Berlin / Heidelberg, 2006.

[28] Forstner, W., Moonen, B.: A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University (1999)

[29] X. Yu; Junzhou Huang; Shaoting Zhang; Wang Yan; Metaxas, D.N., "Pose-Free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model," Computer Vision (ICCV), 2013 IEEE International Conference on , vol., no., pp.1944,1951, 1-8 Dec. 2013

[30] P., Lucey, Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kande Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression. Paper presented at the Third IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010).

[31] V.Vapnik. The Nature of Statistical Learning Theory. Statistics for Engineering and Information Science. Springer Verlag, Berlin, 2000.

[32] B. Jiang, M. F. Valstar, B. Martinez, M. Pantic. A Dynamic Appearance Descriptor Approach to Facial Actions Temporal Modelling, IEEE Transactions on Cybernetics. 44(2): pp. 161 - 174, 2014.