Facial Expression Recognition using Virtual Neutral Image Synthesis

Abhishek Sharma, Anamika Dubey
Electrical Engineering Department, Indian Institute of Technology Roorkee, India
{ana.iitr, abhisharayiya}@gmail.com

Abstract - **A novel feature extraction technique for expression recognition is proposed in this article. The proposed method exploits the properties of eigenvector decomposition in extracting the neutral and expression component of an expression containing test image. The methodology synthesizes a virtual neutral image from the given expression image of a person and then uses it obtain the difference image which captures the expression content of the test image. The proposed method eliminates the requirement of a registered neutral image of the same subject whose expression is to be recognized. Thereby, provides a way for application of difference-image methods in expression recognition to any unknown subject. The proposed methodology is extensively tested on Cohn-Kanade expression data set using three different classifiers: FLDA, Multilayer Perceptron and Support Vector Machine. The accuracy obtained by SVM is 92.4% which is comparable to state-of-art methods. Moreover, the required computation has also been decreased because the proposed method does not require the local features for classification.**

Keywords – Expression recognition, eigenvector decomposition, virtual neutral image synthesis, difference image.

1. Introduction

Facial expressions play an important role in human–machine interaction. Hence, Automatic recognition of facial expressions may act as a component of natural human machine interfaces [1]. In the recent years, numerous approaches to facial expression analysis have been proposed. These methods differ generally in facial feature extraction algorithm. Feature extraction methods can be categorized according to whether they focus on motion or deformation of faces and whether they consider local or holistic facial feature. Deformation extraction has been broadly classified in two categories: namely Image based and Model based. A significant amount of research work has been done in this area and a comprehensive survey pertaining to all the methods used so far in this field is given in [2].

Motion extraction is one of several different techniques for expression recognition. It is based on tracking motion features between expression containing images and corresponding neutral image. Literature is available on basically three motion tracking algorithms: dense optical flow, feature point tracking and difference-images. Inspite of excellent performance of dense optical flow methods in categorising facial actions as indicated in FACS (facial action coding system), these methods often result in extensive computational requirements and hence are practically inefficient. Similar is the case with feature point tracking algorithm [14 15]. Either they require lots of manual effort in point tracking, making the process cumbersome or they require a robust automated system for automated feature tracking which again makes the method computationally expensive and complicated.

Difference-image based methods are yet another popular way of extracting motion from expression images and have been widely used in the discipline of pattern recognition. For facial expression analysis, difference-images are generally synthesized by subtracting a given facial image from a previously registered reference image, containing a neutral face of the same subject. Holistic difference image based motion extraction was employed by many researchers [3, 4]. Choudhury and Pentland [5] used motion field histograms for the modelling of eye and eyebrow actions. However major drawback of “difference images” is the requirement of neutral image of the subject which makes this algorithm unsuitable for expression recognition from single image.

In this paper we propose a virtual neutral image based difference image synthesis, which eliminates the necessity of neutral image of the same person. We have used various classifiers: LDA (linear discriminate analysis) Multilayer perceptron, SVM (support vector machine) to classify extracted features and obtained a significant accuracy of 92.4% using SVM, which is comparable to state-of-art technology. The recognition accuracy itself represents the excellent expressiveness and representation quality of the extracted feature vector. The novel and the most important contribution of the proposed algorithm is the elimination of the requirement of a neutral image for obtaining the difference image.

Rest of the paper is organised as follows: Section 2 discusses the difference image based methods, Section 3 presents the details of proposed algorithm, Section 4 we outlines experimentation details and finally conclusion is drawn up in Section 5 followed up by a short note on future work in Section 6.

2. Difference Image Method

Use of difference images in automatic face recognition is a very old concept. It was first proposed by Baback Moghaddam et al [6], for face recognition by modelling Intrapersonal and Interpersonal subspace. This concept has also been used extensively in field of Facial Expression recognition. As already mentioned, difference images are a way to extract facial motion from neutral images to expression containing image, and are usually created by subtracting a given expression containing image from a previously registered neutral expression image. If **t** represents expression containing image and **n** is the neutral image of the same subject then difference image (**d**) is obtained by pixel-wise subtraction of two images.

$$\mathbf{d} = \mathbf{t} - \mathbf{n}$$

One of the major drawbacks of this algorithm is the requirement of registered neutral image of the test subject. So the approach fails if a neutral image of the same person is not available which makes its application impossible for expression recognition of an unknown person. In the next section the proposed methodology is discussed which eliminates this drawback efficiently.

3. Proposed Methodology

As already mentioned, the major drawback with difference image based algorithms is the necessity of reference image for synthesis of difference image. To meet this limitation we propose a novel method for Difference image (termed as *expression component*) synthesis which is completely subject independent and hence eliminates the requirement of registered image of the subject. Karhunen-Loeve transform [7] was applied and a neutral subspace was formulated by eigenvector decomposition of neutral face images containing the variations present in neutral face space. Whenever, any image is projected on this subspace it is expressed as a linear combination of eigenfaces corresponding to neutral face images. Hence, this subspace can be used to extract the neutral face information present in an expression containing image. The proposed methodology exploits this property of eigenvector decomposition for synthesis of a virtual neutral image of the subject whose image under some expression is given. Once, this virtual neutral image is obtained we can easily subtract it from the given expression containing image to get the difference image. Basically, projection on neutral subspace splits the test image into two components; one representing its neutral component and the other representing the expression component. Let us assume that the test image is represented by a vector **t** in the image space and it's neutral and expression components are to be extracted separately. To achieve this, the image **t** is projected on a subspace spanned by the eigenfaces corresponding to neutral face images which is termed as *neutral subspace.* Let us call the projection point to be **v** which is actually the neutral image component of **t** and is given by-

$$\mathbf{v} = a_1\,\mathbf{e}_1 + a_2\,\mathbf{e}_2 + \ldots\ldots. a_n\,\mathbf{e}_n \qquad (2)$$

where, $\mathbf{e}_i$'s are the eigenfaces corresponding to the neutral face images and they constitute the *neutral subspace.* And a_i's are the components along each eigenface, given by-

$$a_j = (\mathbf{t}\,.\,\mathbf{e}_j) \qquad \forall\ j \in 1.....n \qquad (3)$$

Let us define a vector **d** such that-

$$\mathbf{d} = \mathbf{t} - \mathbf{v} \qquad (4)$$

It implies that,

$$\mathbf{t} = \mathbf{d} + \mathbf{v} \qquad (5)$$

So, the test image **t** is now broken into two components **v** and **d**, the former represents the *neutral component* while latter contains the *expression component.* Fig. 1 depicts the idea of breaking down the test image into two components pictorially. The *expression component* **d** can now be used for expression recognition. The vector **d** is actually an image which contains the expression present in the test image **t** and it is termed as *expression component image.*

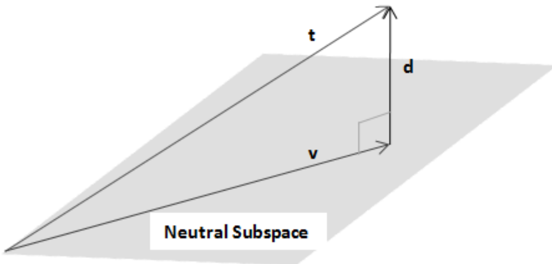

Fig 1: Demonstrating decomposition of test image into neutral and expression components.

The dimension of *expression component image* is too high (2500 in our case) for training the classifiers. So, PCA is used to project the *expression component image* onto a lower dimensional subspace. The required *expression component PCA subspace* is obtained by applying PCA to *expression component images.* The complete process of synthesizing *expression component image* and recognizing the expression present in the test image is depicted pictorially in the flow chart of Fig. 2.

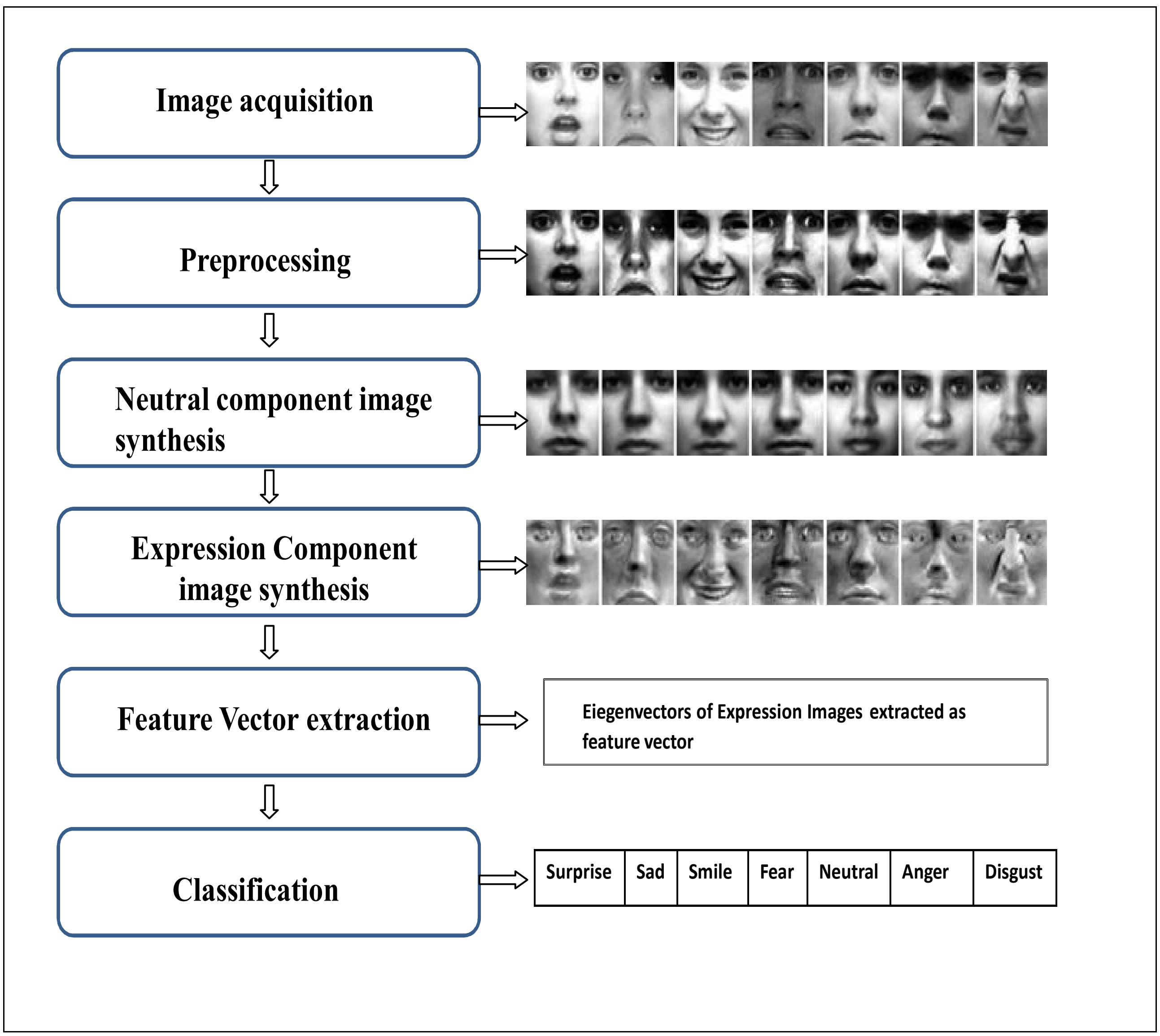

Fig 2 Flowchart for the proposed methodology

4. Experimentation

4.1. Database used

To assess the validity and efficiency of our algorithm we have used Cohn-Kanade AU-Coded Facial Expression Database [12]. The database consists of the images of 100 subjects containing 7 prototypic emotions (i.e., joy, surprise, anger, fear, disgust, neutral and sadness). In each emotion the image starts with neutral image and intensity of expression goes on increasing in small steps.

100 neutral images corresponding to different subjects were used to prepare *neutral subspace*, 100 images per expression i.e. 700 images (with varying expression intensity from neutral to extreme) were used to prepare training set. Training set was used to train the three classifiers: FLDA [9,11], MLP [10,11] and SVM [8,10]. The trained classifiers were tested on 1400 test images (200 images per expression). A detailed outline regarding the implementation of the three classifiers is presented next.

a. Fisher Linear Discriminant Analysis: As discussed earlier, *neutral subspace* is obtained by PCA based eigenvectors of neutral image dataset. Top 10 eigenvectors (based on experimentation) were selected and *expression component images* (corresponding to 7 expressions: happy, sad, surprise, fear, neutral, anger and disgust) were obtained by taking difference of training image and *neutral component image* corresponding to the training image. Leading 50 eigenvector according to their eigenvalues of *expression component images* were used as feature extractor. FLDA classifier was trained using 100 samples per expression and all 6 fisher-axes were retained. Those 6 fisher-axes (of 50×1 dimension) were used as Fisher feature extractor. And the matching is done on the basis of Euclidian distance similarity in Fisher-space. The classifier was tested on 1400 images (200 per expression) and an average accuracy of approximate 81.4% was recorded. Table 1 shows expression wise accuracy of FLDA classifier.

b. Multilayer Perceptron (MLP): 700 *expression component images* (100 images per expression) were obtained for training and all of them were reduced to 25 dimensional features vectors (based on experimental observation for best result) using PCA. A two layer perceptron with 30 hidden neurons (based on experimentation) and sigmoid transfer function has been used as classifier. The output layer contains 7 neurons with linear transfer function. Average recognition accuracy of approximately 88.7% was found using this classifier. The recognition accuracies across different expressions are shown in Table 1.

c. Support Vector Machine (SVM): For training purpose the same training dataset was used as in MLP. LibSVM toolbox [13] which is a freely available toolbox for Support Vector Machines has been used for training and testing purpose. Value of SVM parameters C and γ was determined empirically and optimum values were found to be 20,000 and 0.001 respectively. Average recognition accuracy of 92.4% was achieved which is better than MLP and FLDA classifier. Table 1 tabulates expression-wise recognition accuracy using multiclass SVM classifier along with the other two classifiers for comparison. SVM is essentially a binary classifier to use it for multi-class classification we have used one-versus-one strategy. In [14], 3D face model is used and an accuracy of around 95% is reported with variation in pose too. But, the use of 3 D model has significantly increased the computational and memory requirements. In [15], expression recognition accuracy of 92 % for 2D images is achieved using Gabor features and Enhanced Fisher Discriminator which is slightly less than our proposed method and also the additional computational overhead is incurred too. So, it can be deduced that proposed method is a good compromise between computational requirements and accuracy making it good for real time application.

Table 1: Recognition accuracy of Proposed Method

Expression	FLDA	MLP	SVM
Neutral	85.7%	90.3%	93.5%
Happy	79.2%	88.1%	92.3%
Sad	75.1%	86.7%	90.8%
Fear	82.4%	89.1%	91.9%
Surprise	90.6%	92.6%	96.1%
Disgust	82.1%	89.2%	92.3%
Anger	75.2%	85.4%	89.6%
Average	**81.4%**	**88.7%**	**92.4%**

5. Conclusion

In this article a new approach for difference image based facial expression recognition by making use of properties of eigenvector decomposition is proposed. The proposed methodology eliminates the requirement of neutral image of the subject by synthesizing a virtual neutral image. Hence, it provides a way for expression recognition from a given expression containing test image based on difference image concept. The difference image thus obtained is termed as *expression component image* which captures the information corresponding to the expression of the test image. The validity and efficiency of feature vector was tested using various classifiers (FLDA, MLP and SVM) and results obtained justify the pertinence of *expression component image* as feature vector. A comparative analysis pertaining to classification capability of various classifiers has also been conducted and it has been found that multiclass SVM outperform both MLP and FLDA. The best Recognition accuracy of approx. 92.4% has been reported using SVM classifier which is comparable to the state-of-art methods.

6. Future Work

In the future we are planning to extract the expression component by making use of patch-wise projection on *neutral subspace.* Application of Artificial Wavelet Networks would also be interesting owing to their local approximation capabilities.

7. References

[1] Claude C. Chibelushi, Fabrice Bourel, "Facial Expression Recognition: A Brief Tutorial Overview". *school of Computing, Staffordshire University, 2002.*

[2] B. Fasel, Juergen Luettin, "Automatic facial expression analysis: a survey", *J. Pattern Recognition Soc. v36 i1. 259-275.*

[3] G. Donato, S. Bartlett C. Hager, P. Ekman, J. Sejnowski, Classifying facial actions, *IEEE Trans. Pattern Anal. Mach. Intell. 21 (10) (1999) 974–989.*

[4] B. Fasel, J. Luettin, Recognition of asymmetric facial action unit activities and intensities, *Proceedings of the International Conference on Pattern Recognition (ICPR 2000), Barcelona, Spain, 2000.*

[5] T. Choudhury, A. Pentland, Motion field histograms for robust modelling of facial expressions, *Proceedings of the International Conference on Pattern Recognition (ICPR 2000), Barcelona, Spain, 2000.*

[6] B. Moghaddam, T. Jebara, A. Pentland, Bayesian Face Recognition, Pattern Recognition, Vol. 33, Issue 11, November 2000, pp. 1771-1782

[7] M. Turk, A. Pentland, "Eigenfaces for recognition," J. Cognitive Neurosciences, vol. 3, no. 1, 1991, pp. 71-96.

[8] V. N. Vapnik. Statistical learning theory. *John Wiley & Sons, New York, 1998.*

[9] Richard O. Duda, Peter E. Hart and David G. Stork, Pattern Classification, New York: John Wiley and Sons.

[10] Simon Heykins, Neural Networks: A comprehensive foundation, published by Pearson education (Singapore) Pte. Ltd

[11] Christopher M. Bishop, Pattern Recognition and Machine learning, Springer Science + business media, LLC

[12] Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, Grenoble, France, 46-53.

[13] Chih-Chung Chang, Chih-Jen Lin , LIBSVM – A library for Support vector machines,2001. Software available at http://www.csie.npu.edu.tw/~cjlin/libsvm

[14] Lukasz Zalewski, Shaogang Gong, "Synthesis and Recognition of Facial Expressions in Virtual 3D Views," pp.493, Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.

[15] Doo-Soo Lee, Yang-Bok Lee, Soo-Mi Choi, Yong-Guk Kim and Moon-Hyun Kim. 3-D Facial Expression Recognition-Synthesis on PDA Incorporating Emotional Timing. Advances in Multimedia Information Processing - PCM 2004, Pages 569-576, 2004.