

Facial Expression Classification through Covariance Matrix Correlations

Wilfred O. Odoyo, Beom-Joon Cho, *Member, KIMICS*

Abstract— This paper attempts to classify known facial expressions and to establish the correlations between two regions (eye + eyebrows and mouth) in identifying the six prototypic expressions. Covariance is used to describe region texture that captures facial features for classification. The texture captured exhibit the pattern observed during the execution of particular expressions. Feature matching is done by simple distance measure between the probe and the modeled representations of eye and mouth components. We target JAFFE database in this experiment to validate our claim. A high classification rate is observed from the mouth component and the correlation between the two (eye and mouth) components. Eye component exhibits a lower classification rate if used independently.

Keywords - covariance matrices; eigenvectors; facial component features

I. INTRODUCTION

INFORMATION extraction from images has occupied a larger part of research in computer vision field for a couple of decades now. The specific task involved is to detect and extract valuable features for the correct face recognition and /or facial expression classification. Emotions are varied and can only be conceptualized and understood in the context of cultures, people, situations and the motivation behind them. The vast and ambiguous emotional expressions attest to the difficult task of ascertaining without doubt the true intent of the expresser. What we humans can recognize outright are just but a few facial expressions listed as; anger, disgust, fear, happiness, sadness, and surprise. A neutral face could be considered in other specific problems. Ekman and Friesen [1] work explains the six prototypic expressions that we humans have the ability to generate and interpret. Lots of research works have been carried out on facial analysis focusing on salient feature extraction on both still pictures and videos. This has not been a trivial matter especially in a fast developing field and faced with the high demand to satisfy the ever developing technology market. Not to mention the hurdles involved including pose, illumination,

and occlusion among others, feature extraction is the key to successful classification. This task of feature extraction attracts various methods to deal with the problem. Features on a face can be extracted either geometrically or holistically (appearance based on texture).

Many methods that have been applied are based on Facial Action Coding System (FACS) [1], a system designed for human observers to describe changes in facial expression in terms of visually observable activation of facial muscles. There are 44 different Action Units (AUs) which form the basis of FACS. AUs are a contraction of particular facial muscles or a combination of many of those muscles to cause particular movements. Quite a number of feature extraction methods have been successfully applied by different researchers in this regard. Optical flow estimation [2, 3] is used mostly in motion pictures to extract information. Motion pictures tend to contain more information and reveal so much in terms of surface structure, segmentation, recognition, behavior and scene dynamic understanding among others. Others have used spatial feature analysis [4] which employs topology to examine the relationship existing between image component features and use them to describe real-world situation. The spatial arrangement of features like points, surfaces and lines are captured for analysis. Local feature analysis [5] has also been instrumental in feature extraction. Widely researched traditional methods of feature extraction and classification of facial expressions include Principal Component Analysis (PCA) [6], Linear Discriminant Analysis (LDA) [7], Hidden Markov Model (HMM) [8], Active Appearance / Shape Model (AAM /ASM) [9, 10]. Some of them have excelled and produced commendable results on specific cases of expression classification problems. High performance rates have been observed on particular expressions like anger, happiness, and surprise by most of the aforementioned methods.

We, in this paper are concerned with automatic extraction and classification of facial expressions from still images given a database. The mouth and eye components on the face are first detected, then we employ covariance matrix on these two detected and extracted regions to capture and represent the facial image features. Feature matching is done by Förstner distance measure using eye and mouth components. The minimum distance between the probe and the gallery matrices is considered a match. This enables the facial classification for a particular expression. Therefore, we focus on the distinct

Manuscript received July 15, 2011; revised August 30, 2011; accepted September 5, 2011.

Beom-Joon Cho (Corresponding author) is with the Department of Computer Engineering, Chosun University, Gwangju, 501-759, Korea (Email: bjcho@chosun.ac.kr)

Wilfred O. Odoyo is with the Department of Computer Engineering, Chosun University, Gwangju, 501-759, Korea (Email: wilfody2004@yahoo.com)

facial expressions as exhibited on JAFFE database. In section II we briefly describe detection method as part of image pre-processing. The covariance matrix method of feature representation is explained in section III. Experimental results that validate our method are shown in part IV. We conclude the paper in section V.

II. FACIAL FEATURE DETECTION

The first step of any semi automatic or fully automated face recognition system is to detect the target; the face itself. It is therefore demanding that the task of detecting and extracting the components within the face image be done with a lot of care and exactness. Through template matching [11], we can detect the locations and extract the major facial components for representation. These features extracted from the components encode the most critical information about Action Units movements that might have caused the facial expressions in question to occur. Intuitively, local methods which concentrate on each feature components are deemed to be better than global ones which base detection on the face as a whole and jointly model it.



Fig. 1. Images from JAFFE database belonging to an individual

Our detection method borrowed from [12] with a modification of Gabor filter instead of canny edge detection divides the face into two sub-regions of eye and mouth. Here the eye region includes both eyes and eyebrows. Facial geometry has been used for the division of the face into two with the intuition that the eye and eyebrows occupy the upper and larger part of the face. The mouth is generally assumed to be on the lower part. We have modified Gabor filter and used it to discriminate features based on texture within the regions detected. The edges have been outlined with a combination of mathematical morphological operations of erosion and dilation on the detected surfaces for a more accurate extraction of the target features. This operation is shown in figure 2 below.

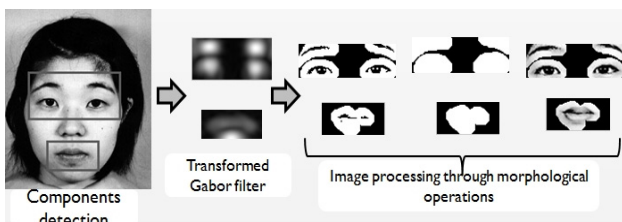


Fig. 2. Facial feature extraction process.

We are now ready with the particular feature regions for the next process. The computation of the covariance is done on the two extracted regions to form two sets of databases; the eye component database and the mouth component database.

III. COVARIANCE MATRICES REPRESENTATION

We consider this part as one of the main contributions of this paper. Even though covariance matrix [13, 14] has been widely used in various applications, we have used it here only within the extracted components of the eye and mouth regions as a descriptor. Furthermore, Covariance matrices have been applauded for the property to enhance robustness against factors like illumination changes, noise, and others which may hinder successful representation of feature points with the regions of interest. The image dimensionality is further reduced owing to the small size of the extracted regions. Covariance also doubles up as feature extraction technique due to the embodiment of both spatial and statistical properties, and used in the classification process as a measure of distance between two different variables. Given a d -dimensional feature points inside a particular region R , the region is represented by $\{z_i\}, i = 1, \dots, n$. The entire image can be represented by a $d \times d$ covariance matrix of feature points computed as;

$$Cov = \frac{1}{n-1} \sum_{i=1}^n (z_i - \mu)(z_i - \mu)^T \quad (1)$$

where μ is the mean value of z_i computed as;

$$\mu = \frac{1}{n} \sum_{i=1}^n z_i \quad (2)$$

Cov is the matrix form that contains and represents the facial component features carried forward for classification purposes. The features captured by the matrix Cov could carry a combination of image texture, edge, and / or gradient for distinctive functions at matching stage.

In figure 3 above, μCov_{eye} , μCov_{mouth} , Cov_{eye} , and Cov_{mouth} are initials given to represent mean covariance of the eye, that of the mouth from the training set, the covariance of the extracted eye, and mouth of the input image, respectively.

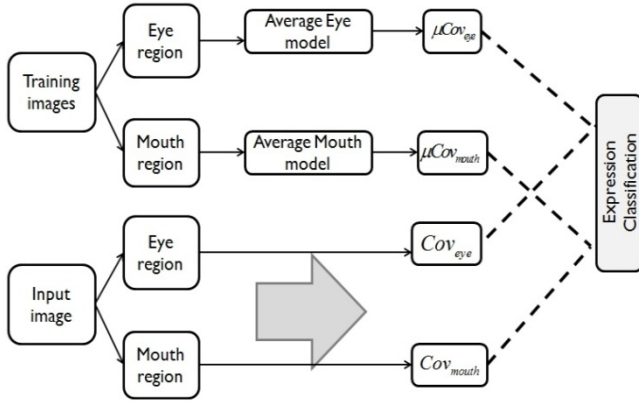


Fig. 3. The proposed expression recognition process

IV. MATCHING AND CLASSIFICATION

The process of matching can now be done by computing the distance between the covariance matrices of the individual components from the input image against that of the trained mean covariance matrices of the six prototypic facial expressions. The model expressions covariances have been trained offline and are stored in the database as feature descriptors. The distance achieved with the minimum value between the probe and the gallery matrices is considered a match. This enables a face to be classified as a particular expression. Having applied it successfully in face recognition project, Förstner distance measure has the ability to compute distances between feature points and is employed here for discrimination purposes. The formula explained in detailed in [14], is stated as below. $Cov1$ and $Cov2$ will represent covariance matrix of the first and second image, respectively. The first and second are used interchangeably depending on which component of the face execution is currently carried out. For instance, when computing the distance between the eye components then $Cov1$ could represent the probe while $Cov2$ to be taken as one of the gallery representation. The same applies to the mouth components.

$$\rho(Cov1, Cov2) = \sqrt{\sum_{i=1}^d \ln^2 \lambda_i(Cov1, Cov2)} \quad (3)$$

where $\lambda_i(Cov1, Cov2)$ are the generalized eigenvalues of $Cov1$ and $Cov2$. The eigenvalues are computed from

$$\lambda_i Cov1 \mathbf{x}_i - Cov2 \mathbf{x}_i = 0 \quad i = 1 \dots d \quad (4)$$

while $\mathbf{x}_i \neq 0$ are the generalized eigenvectors. The distance measure ρ satisfies the metric axioms for positive definite symmetric matrices $Cov1$ and $Cov2$ of

positivity, symmetry, and triangle inequality

In finding the best match using facial components of eye and mouth, some false rejection and acceptance are experienced. Where such mismatch occur and false recognition is the outcome, both the eye and mouth components are used conjunctively to ascertain the true class of the expression. This is where correlation coefficient is calculated between components of the same class and components of the face in question are compared in order to classify the expression to the nearest match.

Correlation: The two parameters of covariance and correlation are related in the sense that they both indicate the extent to which two or more variables co-vary. We scale covariance in order to obtain correlation. This scaling will cue us towards knowing how two different components co-vary with each other. Standard Deviation (SD) σ , is of great importance as the magnitude of the covariance depends on it. Taking a sample pair of two components X_i and X_j , their correlation can be denoted as either $Corr(X_i, X_j)$ or $P_{i,j}$. The correlation is defined by the following formula.

$$P_{i,j} = \frac{Cov(X_i, X_j)}{\sigma_i \sigma_j} \quad (5)$$

where σ_i and σ_j are the standard deviations of X_i and X_j .

V. EXPERIMENTAL RESULTS

In order to evaluate the performance of our method, JAFFE [15] publicly available database has been used. The database contains 213 images of facial expressions. Because each individual has three to four shots of the same expression on different frames, we selected 60 images to train for the database and the rest used as test images. Each expression was trained by at least 10 selected images excluding the neutral face images. 7 x 7 average covariance matrix is computed which represents each class of the six expressions.

Figure4 is the full process at the classification level where each component extracted from the input image are compared with their counterparts as registered in the database. Where distinct classification to an exact expression class by use of components fails, correlation measure is invoked. A mismatch even after the two experiments are done is considered to be a failed classification.

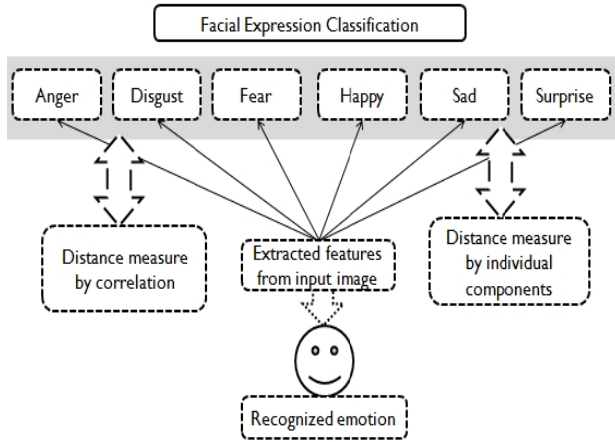


Fig. 4. shows the classification stage.

Distance measure by individual components computes the distance between eye / mouth components from the input image and the corresponding average belonging to different facial expressions. Our experiments show that the mouth components had higher classification rate than that experienced with the eye component. Below are graphs that show this fact. Figure 5a is the happy and surprise mouth component tested with input images from other expressions. Figure 5b is the same but with the eye component. Inter-class expression similarities between eye components are so close to call according to our experiments.

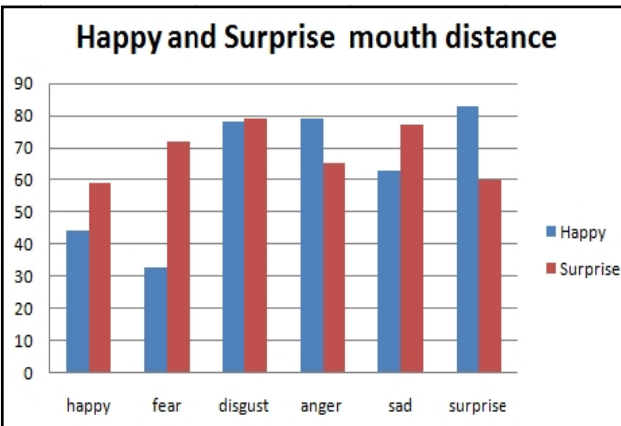


Fig. 5a. shows the difference in magnitude from the happy and surprise average (model) covariance of the mouth.

The distance measure by correlation is derived from the average correlations within a particular expression class compared with the average correlations from other classes. This is shown by the graph in figure 6 below. The fear and disgust can be seen to be very similar and can easily be classified either way.

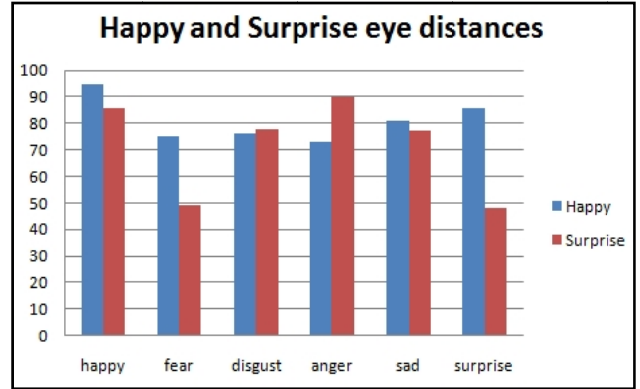


Fig. 5b. shows the difference in magnitude from the happy and surprise average (model) covariance of the eye.

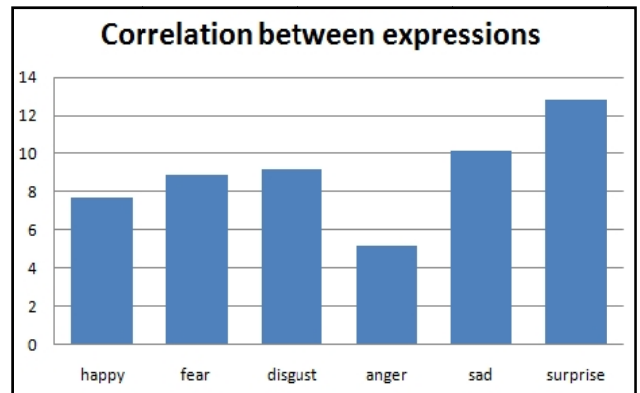


Fig. 6. is the correlation among the six expressions

To compare our method with other successful face classification methods, we used the rate of recognition in this regard. The methods using the location of features and filter matrices have been studied and are popular with successful applications as in AAM(Active Appearance Models) [16] and GFF(Gabor Feature Face) [17] . Only JAFFE database face images were used.

The performance evaluations of above methods are show in Figure 7. The recognition results for AAM, GFF and our method are 91.7, 97.4 and 97.9%, respectively.

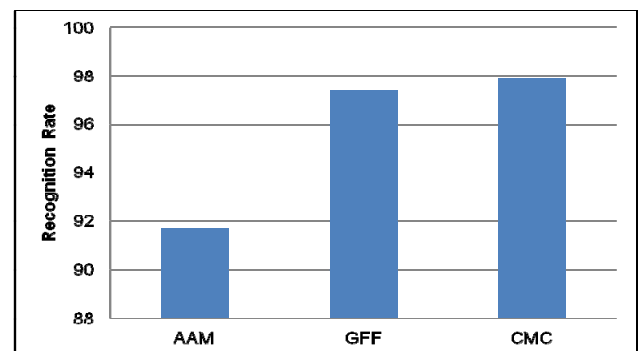


Fig. 7. shows the recognition rate of AAM, GFF and CMC

Our method, which uses the Covariance Matrix Correlations to extract facial component features, presents better results. Since facial expression has the main variation factors such as eyes and mouth, the proposed method uses their correlation of covariance. The better performance of our method is proved more robust against the change of facial expression than AAM. Figure 7 shows that the Gabor-filter based method is also good for recognition of facial expression. The better performance of our approach benefits in image orientation and image conditions and few parameters.

VI. CONCLUSIONS

We have successfully computed the covariance matrices of facial components of eye and mouth regions for similarity measurement between them. The method used extracted the vital feature components from the face and represented them in covariance matrix form. The displacement of eye, eyebrows, and mouth features on different expression poses is seen to significantly discriminate among various expressions. From these activities on the upper and lower face is the deduction that emotional states of an individual can be correlated with these visual features derived from eye and mouth regions. A simple but powerful Förstner distance measure has been used in mapping the similarity or dissimilarity between input data and the gallery images. The use of specific target regions of the eye and mouth improved the recognition rate significantly.

REFERENCES

- [1] P. Ekman and W.V. Friesen, *Facial Action Coding System (FACS)*, Consulting Psychologist Press, Inc., 1978.
- [2] Y. Yacoob and L. S. Davis, "Recognizing Human Facial Expressions from Long Image Sequences Using Optical Flow", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 636-642, June 1996.
- [3] P. Michael, R. E. Kaliouby, "Real Time Facial Expression Recognition in Video using Support Vector Machines", *ICMI 2003*, pp.258-264, 2003.
- [4] J. Zhang, Y. Yan, and M. Lades, "Face Recognition: Eigenface, Elastic Matching, and Neural Nets," *Proc. IEEE*, vol. 85, no. 9, pp. 1423 – 1435, 1997.
- [5] P. S. Penev and J. J. Atick, "Local Feature Analysis: A General Statistical theory for Object Representation," *Network: Computation in Neural Systems*, vol. 7, no. 3, pp. 477-500, 1996.
- [6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 19, pp. 743-720, 1997.
- [7] X. W. Chen and T. Huang, "Facial expression recognition: a clustering-based approach," *Pattern Recognition Letter*, vol. 24, pp.1295-1302, 2003.
- [8] Y. Zhu, L. C. De Silve, C. C. Ko, "Using Moment Invariants and HMM in Facial Expression Recognition", *Pattern Recognition Letters*, vol. 23, pp. 83-91, 2002.
- [9] G.J. Edwards, C.J. Taylor, T.F. Cootes, "Interpreting face images using active appearance models", *3rd IEEE Inter. Conf. of Automatic Face and Gesture Recognition*, vol. 3, pp. 30, 1998.
- [10] A. U. Batur and M.H. Hayes, "Adaptive Active Appearance Models", *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1707-1721, November 2005.
- [11] Haizhou Ai, L. Liang and G. Xu, "Face Detection based on Template Matching and Support Vector Machines", *Proceedings of International Conference on Image Processing*, pp. 1006-1009, 2001.
- [12] Byung-sung Lee, Junchul Chun, Peom Park, "Classification of Facial Expression Using SVM for Emotion Care Service System," *2008 Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, pp.8-12, 2008.
- [13] O. Tuzel, F. Porikli, and P. Meer, "Region Covariance: A fast Descriptor for Detection and Classification", *European Conference on Computer Vision (ECCV, 2006TR2005-111)*, May 2006.
- [14] W. Forstner and B. Moonen, "A metric for covariance matrices", *Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University*, 1999.
- [15] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor Wavelets", *Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara japan, IEEE Computer Society*, pp. 200-205, 1998.
- [16] W. O. Odoyo, G. B. Lee and J. J. Park and B. J. Cho, "Facial component features for facial expression identification using Active Appearance Model," *IEEE Africon 2009*, vol. 9, pp. 1~5, 2009.
- [17] W. O. Odoyo and B. J. Cho, "Global Feature Extraction and Recognition from Matrices of Gabor Feature Faces," *International Journal of Maritime Information and Communication Sciences*, vol. 9, no. 2, pp. 207-211, April 2011.



Wilfred O. Odoyo Received B.S and M.S degrees in Computer Engineering from Chosun University in 2005 and 2007, respectively. He is currently working toward the Ph.D. degree in the department of Computer Engineering at Chosun University. His research interests include Pattern Recognition, Image Segmentation, Wavelets, and Computer Vision.



Beon-Joon Cho Received B.S. and M.S. degrees in Electrical Engineering from Chosun University in 1980 and 1982, respectively. He received his Ph.D degrees in electrical Engineering from Hanyang University in 1988 and Electrical Engineering & Computer Science from KAIST in 2004. He is currently a professor in the Department of Computer Engineering at Chosun University. His research interests include the Pattern Recognition, Neural Network, Character Recognition, and Computer Vision.