Facial Expression Recognition Using Principal Components Analysis

Jonathan DeRuiter and Yusheng Zhu

DePaul University

Abstract – This paper examines the use of principal components analysis (PCA) for the purpose of facial expression recognition. After preprocessing and face detection, PCA was used to extract features. These features were then used in various classification models. Results showed good performance when compared with other methods.

Keywords -- Facial expression recognition; Facial expressions; Principal Component Analysis; Eigenfaces

I. INTRODUCTION

Facial expression is an important mode of nonverbal communication among people. According to Mehrabian [1] amongst the human communication, facial expressions comprises 55% of the message transmitted in comparison to the 7% of the communication information conveyed by linguistic language and 38% by paralanguage. Essentially, it [2] has been revealed that facial expressions are movements of the numerous muscles supplied by the facial nerve that are attached to and move the facial skin. Since Facial expression recognition is an easy task for humans there must be some pattern on the human face that determines certain facial expression which can be used for machine recognition. Meanwhile, psychological studies [3] has identified happy, sad, angry, surprise, disgust, fear, and contempt as the seven most basic expressions that are universal across all cultures.

In this paper, by applying principal component analysis on face images, we have focused our study on classification based on these seven standard expressions. Eigenfaces have been previously used in face recognition, and in this paper, they have been extended to facial expression recognition as dimensionality reduction and feature extraction techniques to achieve a robust and accurate result on two parallel databases.

Facial expression recognition has many significant applications in real world. With facial expression recognition people can test the impact of any content, product or service that is supposed to elicit emotional arousal and facial responses - physical objects such as food probes or packages, videos, and images, sounds, odors, tactile stimuli, etc. Particularly involuntary expressions, as well as a subtle widening of the eyelids, are of key interest as they are considered to reflect changes in the emotional state triggered by actual external stimuli or mental images.

II. RELATED WORK

In recent years, the research of developing automatic facial expression recognition systems has attracted a lot of attention from many different fields. Face detection, feature extraction, and expression classification are the main components that each paper differs from each other. The approaches to facial expression recognition can be roughly divided into two classes: geometrical feature-based approaches and appearance-based approaches. [4]The geometrical feature-based approaches rely on the geometric facial features which represent the shapes and locations of facial components such as eyebrows, eyes, nose, mouth. As for the appearance-based approaches, the whole-face or specific regions in a face image are used for the feature extraction via eigenfaces, fisherfaces or some kinds of filters.

Kotsia and Pitas [5] presented an SVM classification model for facial expression recognition on CK database using geometric displacement of Candide nodes, tested on the whole DB achieved a facial expression recognition accuracy of 99.7%

Zhang and Tjondronegoro [6] presented patch-based Gabor feature extraction from the automatically cropped images, in the form of patches. They matched the patches of the input image with the trained images by comparing the distance metrics and classification carried out by four different kernel SVMs. The results were seen for two databases, obtaining correct recognition rate of 92.93% for JAFFE database and 94.8% for CK database.

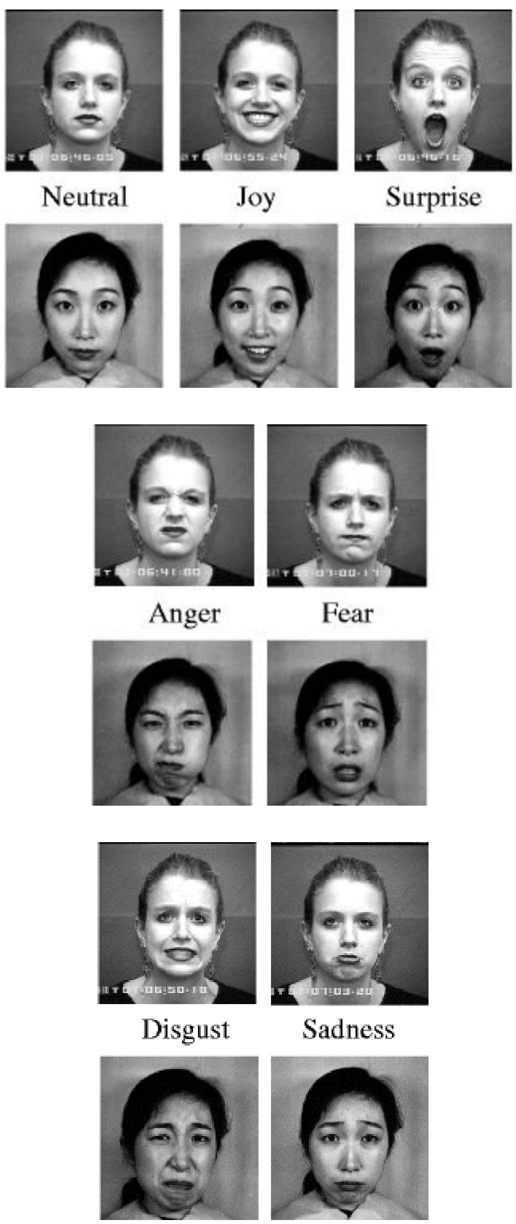
Gosavi and Khot [7] also implemented PCA on JAFFE database, with 70 test images using euclidean distance based matching classifier. They obtained average accuracy of 91.63% and precision of 72.82%. Among the expressions, anger at 94.41% has the highest accuracy of various facial expressions and sad has the highest recognition rate at 71.43%.

III. METHOD

The data used for this analysis comes from two commonly used data sets in facial expression recognition research. One is the Extended Cohn-Kanade Dataset (CK+)[8]. The dataset consists of 123 subjects making the following facial expressions: neutral, sadness, surprise, happiness, fear, anger, and disgust. The dataset contains 593 image sequences, 327 of which have emotion labels. The images are mostly grayscale, with a few color, and have a resolution of 640 by 490. The subjects include both male and female and have a range of various ethnic backgrounds. An example of each expression is shown in the top row of each collection of faces in Figure 1.

The other dataset used was the Japanese Female Facial Expressions (JAFFE) dataset [9]. This dataset includes 213 images of 10 Japanese females making the following expressions: neutral, sadness, surprise, happiness, fear, anger, and disgust. The images are all in grayscale, contain expression labels, and have a resolution of 256 by 256. An example of each expression is shown in the bottom row of each collection of faces in Figure 1.

Figure 1: CK+ and JAFFE Dataset Examples



Given the raw data, the first step is preprocessing the images prior to using principal components analysis. There are eight steps that were completed in preprocessing.

1. First, all images were converted to grayscale. As noted previously, some images from the CK+ dataset are in color. As this method is based on pixel intensities, it is important to have all images in grayscale.

2. Next, face detection had to be performed to identify where the face was in the frame. As seen in figure 1, the faces are in different parts of the image. In order for this method to be effective, it is important that all faces are aligned in the center of the image. The Viola-Jones algorithm was used for facial detection which is based on four stages: Haar feature selection, creating an integral image, AdaBoost training, and cascading classifiers [11].

3. Using the pixels identified by face detection, the images are then cropped so that all faces are centered in the middle of the image.

4. Next, an oval mask is applied to the image to try to isolate just the face and remove hair and background information as this has little relation to facial expressions. Examples of the oval mask can be seen in figure 2 and 3.

5. Next, histogram equalization is used both for contrast enhancement and so that all images have similar pixel distributions. This may also help with differences in both lighting and skin tone.

6. Each image is then converted from an image matrix to a column vector.

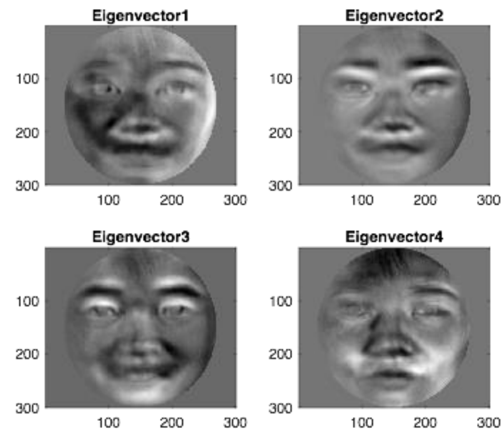
7. Finally, each image in its column vector form is appended to a matrix containing all of the images for a given dataset. In this final image, each column represents a different image.

8. Calculate the "mean face" by taking the mean of all pixel values across all images. This is used later when extracting features for classification. An example of the mean face from JAFFE is shown below in Figure 2.

Figure 2: Mean Face

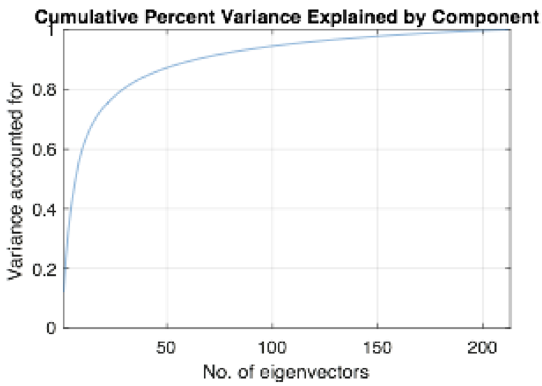
Following preprocessing, PCA is performed to extract the eigenvectors. When using PCA on image data that has been organized as described during preprocessing, each eigenvector can be turned back into an image, which are commonly referred to as "eigenfaces". The first 4 eigenvectors (eigenfaces) are shown below in Figure 3.

Figure 3: Eigenfaces



After performing PCA, eigenvectors are arranged by the variance explained. For example, the first eigenvector explains the most variance in the dataset, followed by the second eigenvector. Figure 4 shows the cumulative variance explained by the eigenvectors for the JAFFE dataset. In the plot, you can easily see the diminishing variance explained by each eigenvector as the eigenvectors increase. For this analysis, we selected enough eigenvectors to account for approximately 90% of the variance the datasets. For the CK+ dataset this was 80 eigenvectors and for the JAFFE dataset, it was 60 eigenvectors.

Figure 4: Variance Explained by Eigenvector



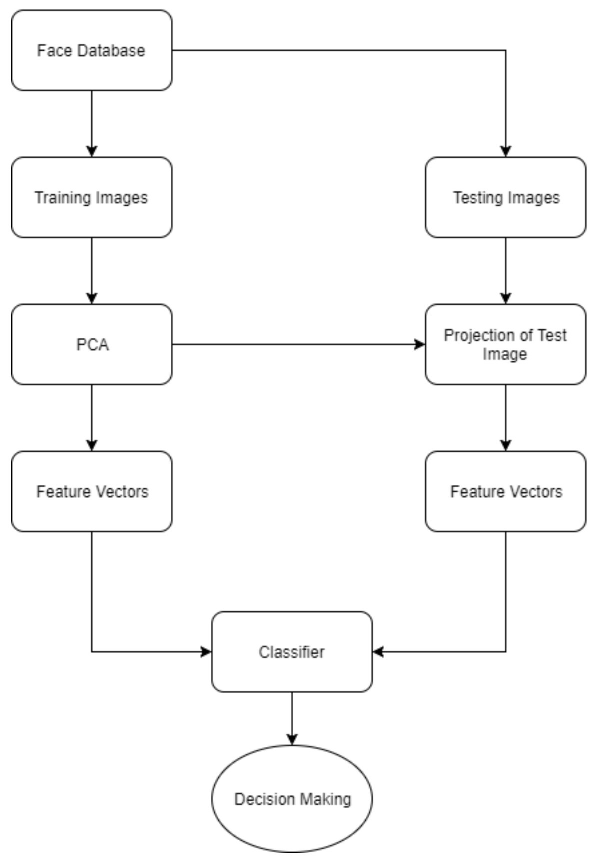
After selecting the number of eigenvector to use, the dataset is then projected onto this subspace, creating feature vectors for each image. This produces a dataset in which each image has a number corresponding to each of the eigenvectors as well as a label for the emotion being shown. This data can then be used for classification modeling.

For the classification modeling, k-fold cross-validation was used rather than a training and testing set. As the datasets are not very large, a standard training and testing set was not appropriate, especially as the CK+ dataset had as few as 17 records for one of the classes. 10 folds were selected k-fold cross-validation, as suggested by Hastie et. al in Elements of Statistical Modeling [12].

Various classification models were used including linear discriminant analysis (LDA), random forest (RF), and support vector machines (SVMs) using a linear kernel. The reason for trying multiple model types is that it is unclear whether more linear or nonlinear approaches will work best for detecting the patterns. With all model types, parameter tuning was performed to achieve the best results.

Figure 5 shows the general process flow for our approach.

Figure 5: Method Flowchart



IV. RESULTS

Table 1 shows the accuracy results for each dataset for each model. With both datasets, the linear discriminant analysis model performed the best, although with the JAFFE the LDA model only outperformed the random forest model by 0.01%.

Table 1: Accuracy Results

Model	CK+	JAFFE
LDA	86%	86%
SVM	79%	84%
RF	67%	86%

Tables 2 and 3 show two different accuracy measures, true positive rate and recall, by the specific expression.

In table 2 we can see that the CK+ dataset did not have an equal number of expressions. The expressions where the count was the lowest also have some of the lowest accuracy measures. For example, there were only 17 images with the contempt expression and the model only had a true positive rate for this expression of 58%. Similarly, there were not as many images with the fear and sadness expressions and the true positive rates and precision measures were low. This indicates that there may not have been enough data to accurately find patterns associated with these emotions.

Table 2: CK+ Results

Expression	Count	True Positive Rate	Precision
Angry	45	76%	71%
Contempt	17	58%	59%
Disgust	58	91%	98%
Fear	25	76%	76%
Happy	68	97%	96%
Sadness	26	62%	64%
Surprise	82	96%	95%

In table 3 on the other hand, we can see that the JAFFE dataset is much more balanced. Here the lower accuracy numbers cannot be attributed to a lower image count. In this case, PCA may not be the best method for extracting features for these expressions, or there may not have been enough data to accurately detect the patterns.

Table 3: JAFFE Results

Expression	Count	True Positive Rate	Precision
Angry	30	90%	87%
Disgust	29	72%	84%
Fear	32	81%	84%
Happy	31	90%	88%
Neutral	30	100%	86%
Sadness	31	87%	84%
Surprise	30	83%	93%

Table 4 shows the accuracy of the method presented in this paper with the other methods outlined in the Related Works section. As seen below, using PCA as a feature extraction method performed worse than these other results, but similarly.

Table 4: Accuracy Comparison to Other Methods

Feature Extraction	Classifier	CK/CK+	JAFFE
Geometric displacement of Candide nodes	SVM	99.70%	-
Gabor features	SVM and MLP	94.80%	92.93%
PCA - Our Method	LDA	87.00%	86.00%

V. CONCLUSION

In this paper, the technique of Principal Component Analysis for facial expression recognition has been thoroughly studied and implemented. The eigenfaces approach provides a practical solution that is well studied for the problem of face recognition. This method is fast, reliable, but only works well in a constrained environment.

Experimental results show that PCA based methods provide good representations and obtain a decent degree of accuracy for facial expression recognition. This is mainly because principal components have proven the capability to provide significant features and reduce the input size of the images. However, the method does not achieve as high of accuracy results as other methods for feature extraction.

The algorithm proposed in this paper has the advantage of a good recognition rate, simple calculations, and quick speed; however, certain issues of robustness to changes in lighting, head orientation, occlusion and the other outliers remain to be a concern. Also, PCA is a global method, and more accurate ways can be achieved by looking at local facial regions. Additionally, the trade-offs between the number of eigenfaces necessary for unambiguous classification are a matter of concern.

REFERENCES

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

[2] Alan J. Fridlund (1994). *Human facial expression* (1 ed.). San Diego: Academic Press. ISBN 978-0-12-267630-7.

[3] Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 211-234). New York: Guilford Press.

[4]. Y. l. Tian, T. Kanade, and J. F. Cohn, “Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity,” in Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 229-234, 2002.

[5] Kotsia, I. and Pitas, I. (2007). Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines. IEEE Transactions on Image Processing, 16(1), pp.172-187.

[6] L. Zhang and D. Tjondronegoro, “Facial Expression Recognition Using Facial Movement Features”, IEEE Transactions on Affective Computing, vol. 2, no. 4, pp. 219-229, October-December 2011.

[7] Gosavi, A. and Khot, S. (2014). Emotion recognition using Principal Component

Analysis with Singular Value Decomposition. 2014 International Conference on Electronics and Communication Systems (ICECS).

[8] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression," in 3rd IEEE Workshop on CVPR for Human Communicative Behavior Analysis, 2010

[9] M. J. Lyons, M. Kamachi and J. Gyoba, "Japanese Female Facial Expressions (JAFFE)," Database of digital images, 1997

[10] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991, hard copy

[11] Viola, P. and Jone, M. "Rapid Object Detection usinga Boosted Cascade of Simple Features", Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference.

[12] Hastie, T., Tibshirani, R., and Friedman, J., "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", New York: Springer, 2001