Quadtree Principal Component Analysis and its Application to Facial Expression Classification

A.Z. Kouzani, F. He, and K. Sammut
School of Informatics and Engineering
Flinders University of SA
GPO Box 2100, Adelaide, SA 5001, Australia
egazk, egfh, egks@flinders.edu.au

Abstract— **This paper presents a method called quadtree principal components analysis for facial expression classification. The quadtree principal components analysis is an image transformation that takes its name from the quadtree partition scheme on which it is based. The quadtree principal components analysis method implements a global-local decomposition of the input face image. This solves the problems associated with the existing principal components analysis and local principal components analysis methods when applied to facial expression classification.**

I. Introduction

People show facial expressions when they are in a specific emotion situation. The number of emotions blended together, the intensity of emotions, and the attempt to control the emotion by the person can all change the image of a face. If this change is significant, the face recognition system may fail to recognise the person. The reason is that known faces are expressionless.

It is reported in the literature [3][5][6] that the methods which classify facial expressions based on the local regions of the image have achieved higher success rates than those of the methods which classify facial expressions based on the entire image. The Principal Components Analysis (PCA) has been widely used for recognition tasks [2][4]. The PCA typically obtains basis images which are non-local from a training set. However, recent studies suggest that the image structure in local regions of an image may be important to the classification tasks. As a consequence, it has been proposed that decompositions which use a local basis image are preferable [3][5][6]. Gray et al. [3] have compared the Local Principal Components Analysis (LPCA) versus the PCA and the Independent Components Analysis (ICA)-based image decompositions on automatic visual lip-reading task. They have reported that image decompositions with local basis images outperform decompositions with global basis images. Similar results have also been obtained by Padgett and Cottrell [6] on a facial expression classification task using a neural network architecture. This supports the idea that the local basis may be a better approach for this classification task.

There are two issues associated with the LPCA-based image decomposition that need to be considered. These issues are: (i) selection of the size of local image regions, and (ii) selection of the location of image regions. A review of the existing literature shows that no attempt has been made for finding an appropriate size for local image regions. Different fixed-region sizes have been chosen by different methods. Whereas Padgett and Cotrell [6] have used 32×32 subimage patches, Gray et al. [3] and Bartlett [1] have chosen to use 12×12 and 15×15 subimage patches, respectively.

Regarding the selection of the image region locations, two different methods have been discussed in the literature: (i) random location and (ii) fixed location. In the random location method, image patches are selected from random locations within the image [1][3][6]. In the fixed location method, however, image patches are chosen from fixed preselected locations [1].

A method is proposed in this paper for classification of an input image into one of the possible facial expression classes. This method implements a global-local decomposition of the input face image, in which the above-stated issues associated with both the PCA and the LPCA need not be considered.

This paper is organised as follows. In Section II, the proposed quadtree principal components analysis method is presented. In Section III, experimental results are given. These results are then discussed in Section IV. Finally, concluding remarks are given in Section V.

0-7803-5731-0/99$10.00 ©1999 IEEE

II. Proposed Quadtree Principal Components Analysis

The Quadtree Principal Components Analysis (QPCA) is an image transformation that takes its name from the quadtree partition scheme on which it is based. A quadtree partition is a representation of an image as a tree in which each node, representing a square portion of the image, contains four subnodes that correspond to the four quadrants of the square. The root of the tree is the initial image. Figure 1 shows an example quadtree partition with four levels. The QPCA is explained below.

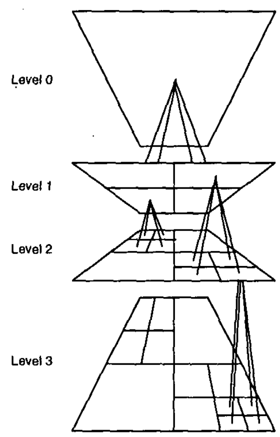

Fig. 1. Example of a quadtree partition with four levels.

Algorithm 1: (**Quadtree Principal Components Analysis**) The QPCA method consists of the following operations.

1. An initial set of N n-dimensional face images, $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_N\}$ (Training Set 0), is acquired. This forms the first level of the quadtree partition.
2. The principal components of the distribution of face images in Training Set 0 are calculated. The basis vectors (images) are named $\{\boldsymbol{b}_1^0, \boldsymbol{b}_2^0, \ldots, \boldsymbol{b}_N^0\}$. Any n-dimensional input face image $\boldsymbol{I}$ can be projected onto the basis images through the following operation:

$$w_k = \boldsymbol{b}_k^{0^T} (\boldsymbol{I} - \boldsymbol{\mu}), \quad (k = 1, \ldots, N), \tag{1}$$

where $\boldsymbol{\mu}$ is the average face. The weights w_k describe the contribution of each basis image to the input face image representation and can be used for reconstruction of the input face image. The reconstruction operation is implemented by

$$\hat{\boldsymbol{I}} = \boldsymbol{\mu} + \frac{1}{N} \sum_{i=1}^{N} w_i \boldsymbol{b}_i^0, \tag{2}$$

where $\hat{\boldsymbol{I}}$ represents the approximation of $\boldsymbol{I}$ by a global combination of the basis images. The error image can be obtained using

$$\boldsymbol{e} = \boldsymbol{I} - \hat{\boldsymbol{I}}. \tag{3}$$

3. For each face image $\boldsymbol{I}_k, (k = 1, \ldots, N)$ in Training Set 0, the following operations are performed.
   (a) The image $\boldsymbol{I}_k$ is omitted from Training Set 0 and the PCA is performed on the rest of the face images to obtain $(N-1)$ basis images.
   (b) The weights are calculated based on $\boldsymbol{I}_k$ and the basis images.
   (c) $\hat{\boldsymbol{I}}_k$ is reconstructed.
   (d) The error image $\boldsymbol{e}_k$ is calculated.
   (e) The error image is divided into four non-overlapping equally-sized subimages. These subimages are used to construct Training Set 1 (the second level of the quadtree partition).
4. The PCA is performed on the face images of Training Set 1 to obtain the associated basis images $\{\boldsymbol{b}_1^1, \boldsymbol{b}_2^1, \ldots, \boldsymbol{b}_N^1\}$.
5. For each image in the latest training set the following operations are performed.
   (a) One image is omitted from the training set and the PCA is performed on the rest of the images to obtain basis images.
   (b) The weights are calculated by projecting the omitted image onto the basis images.
   (c) The projected image is reconstructed.
   (d) The error image is calculated.
   (e) The error image is divided into four non-overlapping equally-sized subimages and a new training set is constructed.
6. The PCA is performed on the images of the new training set to obtain the associated basis images.
7. A jump to Step 5 is always performed unless the sizes of images in the new training set are less than four pixels.

As a demonstration, the average face and ten basis images which are obtained by performing the PCA on a training set of ten aligned face images, are displayed in Figure 2. Figure 3 illustrates the approximation of two face images using a global combination of the ten face images displayed in Figure 2. As can be seen from Figure 3, the first input image (a)

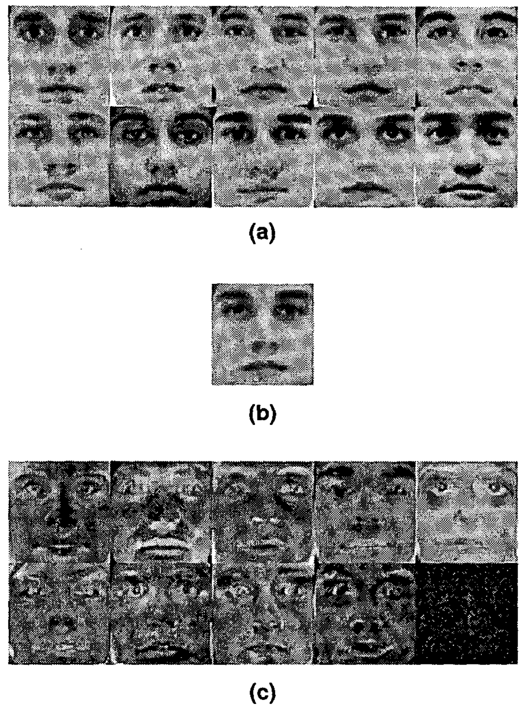

Fig. 2. Principal components of ten face images. (a) Training set. (b) Average face. (c) Basis images.

belongs to the training set, hence the approximated face (b) looks very similar to the original face. The error image (c) appears dark which means that it contains pixel values which are zero or close to zero. Small pixel values denote small errors in this image. The second input image (d) does not belong to the training set, hence the approximated face (e) does not look like the original image. The approximation error (f) is therefore higher than that of the first input image. This means that the principal components of ten face images cannot represent the face space well and many more face images should be added to the training set. Penev and Atick [7] estimated that a training set of more than 1000 face images (not aligned) was required in order to perform a reasonable approximation of any face images. Figure 4 displays ten error images calculated for the training set of the ten face images shown in Figure 2(a). Figure 5 shows Training Set 1 constructed from the training set of the ten face images shown in Figure 2(a). Figure 6 illustrates the basis images obtained from performing the PCA on Training Set 1.

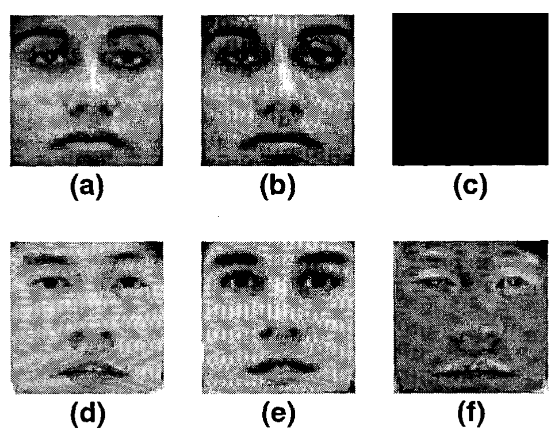

Fig. 3. Approximation of two face images using global combination of ten face images of the training set. (a) Input image 1. (b) Approximated image 1. (c) Error image 1. (d) Input image 2. (e) Approximated image 2. (f) Error image 2.

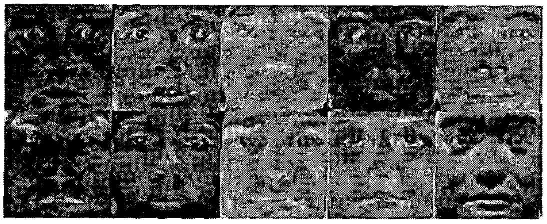

Fig. 4. Ten error images calculated for the training set of the ten face images shown in Figure 2(a).

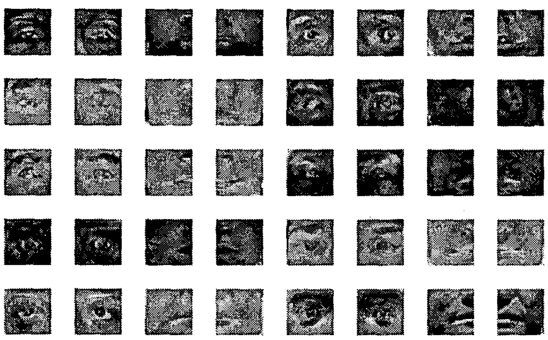

Fig. 5. Training Set 1 constructed from the training set of the ten face images shown in Figure 2(a).

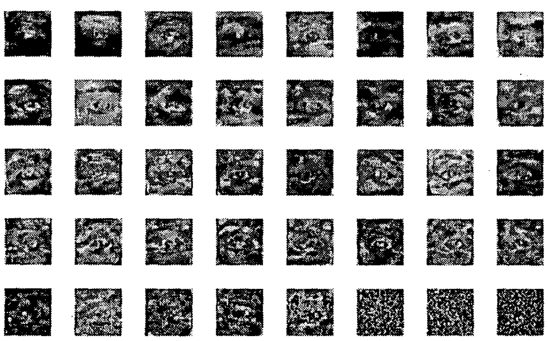

Fig. 6. Basis images obtained from performing the PCA on Training Set 1 shown in Figure 5.

III. Experimental Results

The QPCA together with three existing counterparts are implemented in order to compare their relative performances for expression classification. The existing implemented methods are the PCA, the LPCA Random, and the LPCA Fixed. In the LPCA Random and the LPCA Fixed, the size of subimage patches are set to 15×15 pixels. Experiments are carried out on the following two test sets of face images.

Test Set 1 contains 765 front-view face images of 15 individuals synthesised from a selected set of images from the Yale face database. Figure 7 displays a sample set of facial expressions. From this database, 90 images of 15 subjects, containing the 6 facial expressions, are selected. For each combination of two images of a subject, 3 extra images with different facial expressions are synthesised using the image morphing technique. Therefore, the total number of images per subject will increase from 6 to 51, representing 51 different facial expressions.

Test Set 2 contains 320 face images of 20 individuals selected from the second CMU face database. Each image contains one of four different facial expressions: neutral, angry, happy, and sad. There are 16 images per subject, four front-view, four left-view, four right-view, and four up-view images.

The face area is manually extracted and automatically aligned. In Test Set 1, the faces are grouped based on the expressions they represent. Six groups are built and a reference image is selected for each group. In Test Set 2, the face images are grouped based on their poses and expressions. Sixteen groups are formed and a reference image is selected for each group.

The methods under examination are trained and tested using a leave-one-out cross-validation procedure which makes maximal use of the available data for training. In this procedure, all of the images of that subject are reserved for testing. This procedure is repeated for each of the 15 subjects in Test Set 1 and 20 subjects in Test Set 2. A simple nearest-neighbour classifier is used for classification. The classification results are displayed in Table I.

IV. Discussions

According to Table I, the best performance is obtained using the QPCA which achieves 91.1% and 73.7% correct classification for Test Set 1 and Test Set 2, respectively. The two LPCA methods outperform the PCA.

The results discussed here are based on the experi-

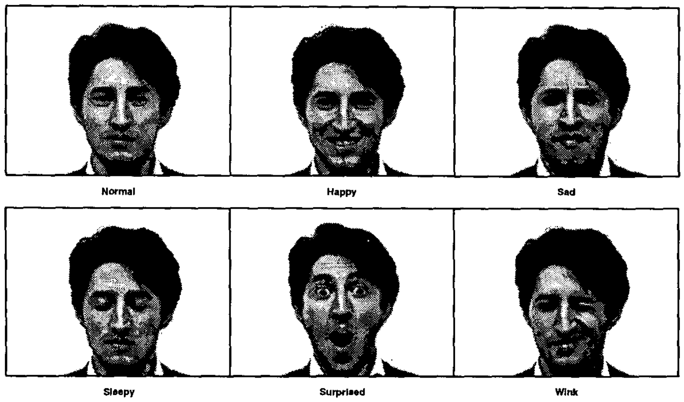

Fig. 7. Sample face images of a subject with six different facial expressions from the Yale face database.

TABLE I
CLASSIFICATION OF FACIAL EXPRESSIONS FOR TEST SETS 1-2.

Method	Test Set	Correct Classification	Classification Rate
PCA	1	586	76.6%
	2	180	56.2%
LPCA Random	1	603	78.8%
	2	198	61.8%
LPCA Fixed	1	638	83.3%
	2	209	65.3%
QPCA	1	697	91.1%
	2	236	73.7%

ments performed on only two small test sets containing a total of 410 face images. The face images do not contain all the possible facial actions. A larger collection of test sets would allow a more reliable evaluation of the different existing methods.

V. Conclusions

A novel quadtree principal components analysis method (QPCA) is proposed for performing facial expression classification. The QPCA implements a global-local decomposition of the input face image. This solves the problems associated with the existing PCA and LPCA methods described in Section I. Experimental studies suggest that the QPCA outperforms the other methods.

Acknowledgements

The first author would like to thank DEETYA, Australia, and FUSA, for providing him with scholarships to undertake this research.

References

[1] M.S. Bartlett, *Face Image Analysis by Unsupervised Learning and Redundancy Reduction*, PhD thesis, Cognitive Science and Psychology, University of California, San Diego, 1998.

[2] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey", *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, May 1995.

[3] M. Gray, J. Movellan, and T. Sejnowski, "A comparison of local versus global image decomposition for visual speechreading", in *Proc. 4th Joint Symposium on Neural Computation*, 1997, pp. 92–98.

[4] A.Z. Kouzani, *Invariant Face Recognition*, PhD thesis, School of Informatics and Engineering, Flinders University of SA, Adelaide, South Australia, Australia, 1999.

[5] B. Olshausen and D.J. Field, "Natural image statistics and efficient coding", *Network: Computation in Neural Systems*, vol. 7, pp. 333–339, 1996.

[6] C. Padgett and G. Cottrell, "Representing face images for emotion classification", in *Advances in Neural Information Processing Systems*, M. Mozer, M. Jordan, and T. Petsche, Eds., vol. 9. MIT Press, Cambridge, MA, 1997.

[7] P. Penev and J. Atick, "Local feature analysis: a general statistical theory for object representation", *Network: Computational in Neural Systems*, vol. 7, no. 3, pp. 447–500, 1996.