FACIAL EXPRESSION SYNTHESIS

A.Z. Kouzani

School of Engineering, Flinders University of SA, GPO Box 2100, Adelaide, SA 5001, Australia
egazk@flinders.edu.au

ABSTRACT

A new method is presented in this paper for synthesis of an expressionless face image from a face image containing an arbitrary known expression. The proposed method enhances the performance of the existing principal components analysis which can be trained using a limited set of expressionless face images. This performance enhancement is achieved through the extraction of the basis set of the example face images in a global-local-based fashion.

1. INTRODUCTION

A face recognition system should be able to cope with the numerous variations that can potentially exist between different images of the same face. One important source of image variations is the facial expressions that people show when they are in a specific emotion situation. The appearance of a happy, sad, or surprised person could be different to that of a normal person. If this difference is significant, automatic face recognition systems might fail to recognise that person. In order to cope with facial expressions, there are three methods which can be used with existing face recognition systems.

In the first method, the known face database employed must contain images with all possible facial expressions for each known person. However, most of the current face database collections, such as the FBI criminal face database, contain expressionless faces.

In the second method, a recognition strategy is developed to obtain invariance to facial expression variations. However, this method can only achieve invariance to low intensity emotions. Moreover, this method cannot be used in conjunction with other methods to recognise a face image containing a combination of more than one image variations.

In the third method, the input face image is compensated for facial expression and the compensated image is presented to the recognition system. This method can be used in conjunction with other methods to develop a face recognition system that can handle multiple forms of image variations.

A facial expression compensation system should contain two stages: expression classification and expression synthesis. In the first stage, an input face image containing an arbitrary expression is classified into one of several possible facial expression classes. In the second stage, an expressionless face image is generated from the input face image. This paper deals specifically with the synthesis stage and the reader is referred to the following publications for information on the classification stage [1, 2, 3].

It is reported in [3, 4, 5] that the use of local image information for classifying facial expression allows higher success rates than would be obtained with global image information techniques, such as *Principal Components Analysis* (PCA). However, there are two issues associated with local information based methods which need to be dealt with: the selection of an appropriate local image region size, and the selection of the location of the image regions. No attempt has been reported in the literature to address these two issues. In this paper a new technique *Quadtree Principal Components Analysis* (QPCA) is proposed which implements a global-local decomposition of the input face image so as to address these two issues. The QPCA offers a significant improvement in performance over that of the PCA for approximating face images using a limited set of examples. In addition, a method is proposed, based on the QPCA, for synthesising an expressionless face image from a face image containing a known expression.

This paper is organised as follows. In Section 2, the existing methods are reviewed. In Section 3, the QPCA is proposed and the corresponding experimental results are given. The expression synthesis system and its associated experimental results are presented in Section 4. In Section 5, the system discussions are given. Finally, concluding remarks are presented in Section 6.

2. REVIEW OF EXISTING METHODS

Generally, the compensation for facial expression can be performed using two main methods: *3D model-based* and *2D image-based*. These methods are explained in the following:

3D Model-Based Methods: In these methods, facial synthesis is performed using an explicit physical model of the face. A good example of these methods is the work of Essa and Pentland [2] who extended a detailed anatomical and physical model of the face, and applied the model to both recognition and synthesis of facial expression. However, Essa-Pentland's synthesised images are not realistic, and it is doubtful whether their computationally intensive synthesis approach is practical without the use of a dedicated hardware.

0-7803-5467-2/99/ $ 10.00

2D Image-Based Methods: These methods have been applied to the synthesis of rigid and non-rigid face transitions [6, 7, 8]. The methods exploit prior knowledge of prototypical faces from example images, and work by building flexible representations of them using a linear combination of labelled examples.

The work of Beymer et al. [6] is a popular image-based technique to synthesise facial expression. Beymer et al. cast the expression interpolation approach in a learning-by-example framework where each example image is associated with a position in a high-level, multi-dimensional parameter space denoting expression. Beymer et al. further presented a system for synthesising novel images of a particular face from just one image of the face. The idea is to use a collection of example images of another person as a prototype for representing generic face transformations. The image of the input face is first brought into correspondence with the closest prototype image, then the new images are synthesised from just one input face.

An alternative image-based method which can be used for generating novel facial expression is based on the concept of linear object classes as proposed by Vetter and Poggio [9]. For linear object classes, the new image of an object from a class is analysed in terms of the shape and texture vectors of prototype objects. This requires correspondence to be established between all feature points of the prototype images. It also requires correspondence between the new image and one of the prototypes of the same pose. Although Vetter-Poggio's method was not applied to facial expression synthesis, it has the potential to be used for this purpose.

The image-based methods have their own advantages and disadvantages. The main advantage of the Beymer et al.'s method is that only two example images of another person are required as a prototype for generating an expressionless face image from just one input face. The disadvantage of this method is that the generated expressionless face image cannot represent the real face image because if, for instance, the person's eyes are closed in the input image, there is no way to produce an image with open eyes using only local image deformation.

On the other hand, the advantage of Vetter-Poggio's method is that the generated image is not directly built from the content of the input image, but from a linear combination of a collection of example images. Therefore, if a sufficient number of example faces are used, the method will produce a much better approximation of the expressionless face image. The main limitation of this method is the existence of linear object classes and the completeness of the available examples.

3. QUADTREE PRINCIPAL COMPONENTS ANALYSIS

Given a linear space such as the face space, one can choose among different sets of basis vectors that will span the same space. The basis set used by Vetter and Poggio was the original set of images themselves. However, other potential basis sets can be used as well. One popular method for choosing the basis set is to apply PCA to the example set. Since the basis images found by the PCA are orthogonal, the reconstruction of new images can be easily performed. This orthogonality produces a more stable set of linear coefficients [10] which may produce a better approximation of a different face image. However, a large number of example faces are still required to accurately model all human faces.

To enhances the PCA's performance in modelling a face image using a limited number of example face images, the QPCA is proposed for extraction of the basis set. The reason for the performance enhancement is that the QPCA computes the basis set both globally and locally from the example set whereas the PCA only obtains the basis set globally.

Algorithm 1 *(Quadtree Principal Components Analysis) This 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.*

2. *The principal components of the distribution of faces images in Training Set 0 (the first level of the quadtree partition) are calculated. The basis vectors (images) are named $\{\boldsymbol{b}_1^0, \boldsymbol{b}_2^0, \ldots, \boldsymbol{b}_N^0\}$. (see Figure 1). The face images are aligned.*

   Any n-dimensional input face image $\boldsymbol{I}$ can be projected onto the basis images by the following operation:

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

   where $\boldsymbol{\mu}$ is the average face. This describes a set of point-by-point image multiplications and summations. The weights w_k describe the contribution of each basis image to the input face image representation. These weights can be used for reconstruction of the input face image. The reconstruction operation is implemented by:

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

   where $\hat{\boldsymbol{I}}$ represents the approximation of $\boldsymbol{I}$ by a global combination of the face images of the training set. Figure 2 illustrates the approximation of two face images using a global combination of the ten face images displayed in Figure 1. As can be seen from Figure 2, the first input image (a) belongs to the training set, hence the approximated face (b) looks very similar to the original face. The error image (c) which is obtained by:

   $$\boldsymbol{e} = \boldsymbol{I} - \hat{\boldsymbol{I}} \qquad (3)$$

appears dark which means that it contains pixel values which are close to zero or 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 properly represent the face space and that many more face images should be added to the training set. Penev and Atick [5] 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 image.

3. *For each face image $I_k, (k = 1, \ldots, N)$ in Training Set 0, the following operations are performed:*
   - (a) *The image 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 I_k and the basis images.*
   - (c) *$\hat{I}_k$ is reconstructed.*
   - (d) *The error image e_k is calculated (as shown in Figure 3)*
   - (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 (see Figure 4).*
4. *The PCA is performed on the face images of Training Set 1 to obtain the associated basis images $\{b_1^1, b_2^1, \ldots, b_N^1\}$. (See Figure 5).*
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. *The algorithm continuously returns to step 5 unless the images sizes in the new training set are less than four pixels.*

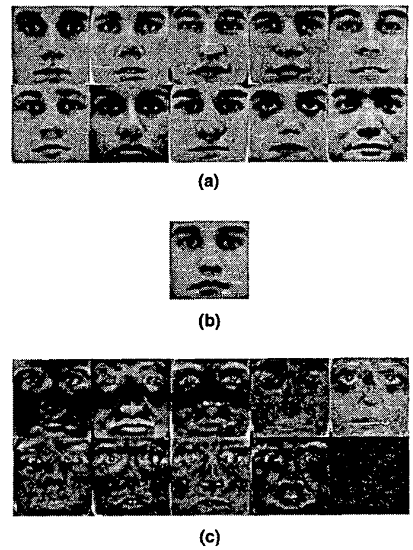

Figure 1: Principal components of ten face images. (a) Training set. (b) Average face. (c) Basis images.

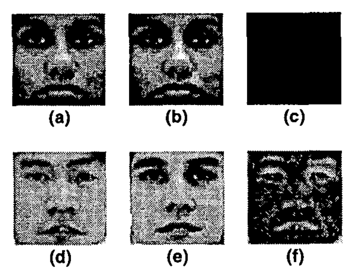

Figure 2: Approximation of two face images using global combination of ten face images of the training set. (a) Input image 1. (b) Approximated image 1. (c) Error image 1. (d) Input image 2. (e) Approximated image 2. (f) Error image 2.

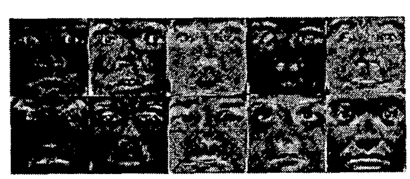

Figure 3: Ten error images calculated for the training set of ten face images shown in Figure 1(a).

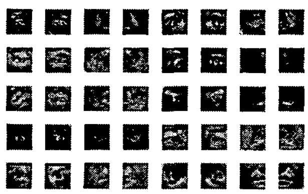

Figure 4: Training Set 1 constructed from the training set of ten face images shown in Figure 1(a).

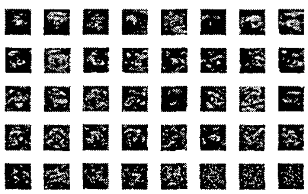

Figure 5: Basis images obtained from performing the PCA on Training Set 1 shown in Figure 4.

Two experiments are carried out to visualise the results of approximating three different face images using the PCA and the QPCA techniques and to determine how their approximation capabilities vary with the size of the training set. In these experiments two training sets are used. Training Set 1, which is used in Experiment 1, consists of 10 front-view face images. Training Set 2, which is used in Experiment 2, contains 50 front-view face images. Three front-view face images are also selected as input images (see Figure 6 (first row)). These images are neither in Training Set 1 nor Training Set 2. The experiments are performed as follows:

Experiment 1: In this experiment, both the PCA and the QPCA are first trained on the images of Training Set 1 and the associated basis images are obtained. Then the three input face images are separately projected onto the basis images and the projection weights are calculated. Finally, two sets of three images are reconstructed, one set using the calculated weights and the basis images of the PCA (see Figure 6 (second row)), and the other set using the calculated weights and the basis images of the QPCA (see Figure 6 (fourth row)).

Experiment 2: The same procedure is repeated for Experiment 2 as used in Experiment 1, except that the larger Training Set 2 is now employed. The PCA reconstructed images are shown in Figure 6 (third row), while those for the QPCA are shown in Figure 6 (fifth row).

As can be seen from Figure 6, the images reconstructed by the PCA which is trained on 10 example images, do not resemble the original images. The images reconstructed by the QPCA which is trained on the same number of example images, however, contain some resemblance to the original images. Utilisation of a larger training set of 50 images improves the performance of both the PCA and the QPCA as shown in Figure 6. The important points found from these images are: (i) the QPCA outperforms the PCA for both experiments, and (ii) the QPCA performs better if trained on a larger training set.

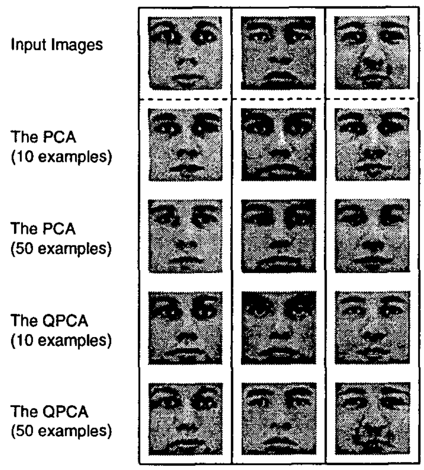

Figure 6: Approximating three front-view face images using the PCA and the QPCA (see text).

4. PROPOSED EXPRESSION SYNTHESIS SYSTEM

The QPCA is employed to synthesise expressionless face image from face images containing known facial expression. The synthesis process is as follows:

Algorithm 2 *(Expression Synthesis) This process is performed in two steps. These steps are as follows:*
Initialisation: *In the initialisation step the following operations are performed:*

1. *One training set is acquired for each facial expression.*
2. *An image is manually selected from each training set and is named the reference image of the set. Although this selection is an arbitrary choice, the principle employed is that the face should be located in the centre of the image.*
3. *All images from each training set are aligned based on the associated reference image using pixel-based correspondence.*
4. *The QPCA is applied to each training set to generate the basis images.*

Synthesis: In the synthesis step the following operations are performed:

1. *A set of weights are calculated based on the input image and the basis image by projecting the input image onto each basis image of the training set representing a similar facial expression as that of the input image.*

2. *An expressionless face image is synthesised using the calculated weights from Step 1 and the basis image of the training set representing no facial expression.*

An experiment is carried out to demonstrate the performance of the above procedure. A database of 30 Caucasian subjects containing 15 males and 15 females with both surprised and neutral aligned images are used to build two training sets: *surprised* and *neutral.* The initialisation, described in Algorithm 2, is performed on both training sets. A face image of a person not included in either the training set and containing a surprised expression is projected onto the surprised face space and the projection weights are obtained. Then, an image is synthesised using the calculated weights obtained from the surprised face space projection and the basis image of the neutral training set. The original and synthesised images are shown in Figure 7. Although, the quality of the synthesised image is not satisfactory, the synthesised image does closely resembles the same person's face.

5. DISCUSSIONS

The proposed synthesis procedure is based on 2D image-based method and can generate artificial images of an object when only a single image is given. This approach may, however, suffer if the set of available expression examples is not complete. Since the QPCA method tries to eliminate the difference between the input image and its projection onto the example space by performing the quadtree decomposition procedure, increasing the number of example images will improve the approximation of the input image. This is due to the fact that a larger number of example images will provide the QPCA with more global and local information to synthesise a face image. Clearly, the use of 30 example faces is not sufficient to accurately model all human faces. Nevertheless, the application of the method to a small example set of human faces does provide some promising preliminary results.

6. CONCLUSIONS

To synthesise an expressionless face image from a face image containing an arbitrary known expression, the new QPCA method is proposed. The QPCA implements a global-local quadtree decomposition of the input face image in order to minimise the difference between the input image and its projection onto the example space. The QPCA produces a higher quality approximation of the input image than the PCA, using only a limited number of example images.

(a) (b)

Figure 7: Synthesis of an expressionless face image.

7. ACKNOWLEDGEMENTS

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

8. 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] I.A. Essa and A. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Tran. on PAMI*, 19(7):757–763, 1997.

[3] C. Padgett and G. Cottrell. Representing face images for emotion classification. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Inf. Proc. Systems*, volume 9. MIT Press, 1997.

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

[5] P. Penev and J. Atick. Local feature analysis: a general statistical theory for object representation. *Network: Comp. in Neural Sys.*, 7(3):447–500, 1996.

[6] D. Beymer, A. Shashua, and T. Poggio. Example based image analysis and synthesis. Technical Report 1431, MIT AI Lab., 1993.

[7] M.J. Jones and T. Poggio. Model-based matching by linear combinations of prototypes. Technical Report 1583, MIT AI Lab., November 1996.

[8] T. Vetter. Synthesis of novel views from a single face image. Technical Report 26, Max-Plank Inst., Germany, February 1996.

[9] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. Technical Report 1531, MIT AI Lab., 1995.

[10] D. Beymer. Vectorizing face images by interleaving shape and texture computations. Technical Report 1537, MIT AI Lab., September 1995.