AN INTERPOLATION METHOD FOR THE RECONSTRUCTION AND RECOGNITION OF FACE IMAGES

N. C. Nguyen and J. Peraire
Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 20139, USA

Keywords: Face reconstruction, face recognition, best points interpolation method, principal component analysis.

Abstract: An interpolation method is presented for the reconstruction and recognition of human face images. Basic ingredients include an optimal basis set defining a low-dimensional face space and a set of interpolation points capturing the most relevant characteristics of known faces. The interpolation points are chosen as pixels of the pixel grid so as to best interpolate the set of known face images. These points are then used in a least-squares interpolation procedure to determine interpolant components of a face image very inexpensively, thereby providing efficient reconstruction of faces. In addition, the method allows a fully automatic computer system to be developed for the real-time recognition of faces. The advantages of this method are: (1) the computational cost of recognizing a new face is independent of the size of the pixel grid; and (2) it allows for the reconstruction and recognition of incomplete images.

1 INTRODUCTION

Image processing and recognition of human faces constitutes a very active area of research. The field has evolved rapidly and become one of the most successful applications of image analysis and computer vision partly because of availability of many powerful methods and partly because of its significant practical importance in many areas such as authenticity in security and defense systems, banking, human–machine interaction, image and multimedia processing, psychology, and neurology. Principal component analysis (PCA) or the Karhunen-Loève (KL) expansion is a well-established method for the representation (Sirovich and Kirby, 1987; Kirby and Sirovich, 1990; Everson and Sirovich, 1995) and recognition (Turk and Pentland, 1991) of human faces.

PCA approach (Kirby and Sirovich, 1990) for face representation consists of computing the "eigenfaces" of a set of known face images and approximating any particular face by a linear combination of the leading eigenfaces. For face recognition (Turk and Pentland, 1991), a new face is first projected onto the eigenface space and then classified according to the distances between its PCA coefficient vector and those representing the known faces. There are two drawbacks with this approach. First, PCA may not handle incomplete data well situations in which only partial information of an input image is available. Secondly, the computational cost per image classification depends on the size of the pixel grid.

This paper describes an interpolation method for the reconstruction and recognition of face images. The method was first introduced in (Nguyen et al., 2006) for the approximation of parametrized fields. The basic ingredient is a set of interpolation points capturing the most relevant features of known face images. The essential component is a least-squares interpolation procedure for the very rapid computation of the interpolant coefficient vector of any given input face. The interpolant coefficient vector is then used to determine which face in the face set, if any, best matches the input face. A significant advantage of the method is that the computational cost of recognizing a new face is *independent* of the size of the pixel grid, while achieving a recognition rate comparable to PCA. Moreover, the method allows the reconstruction and recognition of incomplete images.

The paper is organized as follows. In Section 2, we present an overview of PCA. In Section 3, we ex-

C. Nguyen N. and Peraire J. (2007).
AN INTERPOLATION METHOD FOR THE RECONSTRUCTION AND RECOGNITION OF FACE IMAGES.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 91-96

tend the best points interpolation method (BPIM) introduced in (Nguyen et al., 2006) and apply it to develop an automatic *real-time* face recognition system. In section 4, we test and compare our approach with PCA. Finally, in Section 5, we close the paper with some concluding remarks.

2 PRINCIPAL COMPONENT ANALYSIS

2.1 Eigenfaces

An ensemble of face images is denoted by $\mathcal{U}_K = \{u_i\}$, $1 \le i \le K$, where u_i represents an i-th mean-subtracted face and K represents the number of faces in the ensemble. It is assumed that after proper normalization and resizing to a fixed pixel grid Ξ of dimension N_1 by N_2, u_i can be considered as a vector in an N-dimensional image space, where $N = N_1 N_2$ is the number of pixels. PCA (Sirovich and Kirby, 1987; Kirby and Sirovich, 1990) constructs an optimal representation of the face ensemble in the sense that the average reconstruction error

$$\bar{\varepsilon}^* = \sum_{i=1}^{K} \left\| u_i - \sum_{j=1}^{k} (\phi_j^T u_i)\phi_j \right\|^2 , \quad (1)$$

is minimal for all $k \le K$. In the literature (Turk and Pentland, 1991), the basis vectors ϕ_j are referred as *eigenfaces* and the space spanned by them is known as the *face space*. The construction of the eigenfaces is as follows.

Let $\mathbf{U}$ be the $N \times K$ matrix whose columns are $[u_1, \ldots, u_K]$. It can be shown that the ϕ_i satisfy

$$\mathbf{A}\phi_i = \lambda_i \phi_i , \quad (2)$$

where the covariance matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \frac{1}{K}\mathbf{U}\mathbf{U}^T . \quad (3)$$

Here the eigenvalues are arranged such that $\lambda_1 \ge \ldots \ge \lambda_K$. Since the matrix $\mathbf{A}$ of size $N \times N$ is large, solving the above eigenvalue problem can be very expensive.

However, if $K < N$, there will be only K meaningful eigenvectors and we may express ϕ_i as

$$\phi_i = \sum_{j=1}^{K} \varphi_{ij} u_j . \quad (4)$$

Inserting (3) and (4) into (2), we immediately obtain

$$\mathbf{G}\varphi_i = \lambda_i \varphi_i , \quad (5)$$

where $\mathbf{G} = \frac{1}{K}\mathbf{U}^T\mathbf{U}$ is a symmetric positive-definite matrix of size K by K. The eigenvalue problem (5) can be solved for $\varphi_{ij}, 1 \le i, j \le K$, from which the eigenfaces ϕ_i are obtained.

Figure 1: Eigenfaces and the mean face. The mean face is on the top left and followed by 11 top eigenfaces, in order from left to right and top to bottom.

We present in Figure 1 the mean face and a few of the top eigenfaces for a training ensemble of 400 face images extracted from the AT&T database (see Section 4.1 for details).

2.2 Face Reconstruction

We briefly describe the reconstruction of face images using PCA and later compare the results with those obtained using our method. First, we project an input face u onto the face space $\Phi_k = \text{span}\{\phi_1, \ldots, \phi_k\}$ to obtain

$$u^* = \sum_{i=1}^{k} a_i \phi_i , \quad (6)$$

where for $i = 1, \ldots, k$,

$$a_i = \phi_i^T u . \quad (7)$$

We also define the associated error as

$$\varepsilon^* = \|u - u^*\| . \quad (8)$$

Note that the mean face of the ensemble $\mathcal{U}_K$ should be added to u^* to obtain the reconstructed image; and that if k is set equal to K, the reconstruction is exact for all members of the ensemble.

2.3 Face Recognition

We briefly describe the eigenface recognition procedure of Turk and Pentland (Turk and Pentland, 1991). To classify an input image, one first obtains PCA coefficients $a_i, 1 \le i \le k$, as described above. One then computes the Euclidean distances between its PCA coefficient vector $\mathbf{a} = [a_1, \ldots, a_k]^T$ and those representing each individual in the training ensemble. Depending on the smallest distance and the PCA reconstruction error ε^*, the image is classified as belonging to a familiar individual, as a new face, or a non-face image. Several variants of the above procedure are possible via the use of a different classifier such as the nearest-neighbor classifier and a different norm such as L_1 norm or Mahalanobis norm (Delac et al., 2005).

It is generally observed that the recognition performance is improved when using a larger k. Typically, the number of eigenfaces k required for face recognition varies from $O(10)$ to $O(10^2)$ and is *much* smaller than N. We note that classification of an input image requires the evaluation of PCA coefficients according to (7). The computational cost per image classification is thus at least $O(Nk)$. This cost depends linearly on N and is quite acceptable for a *small* number of input images. However, when classification of *many* images is performed at the same time, PCA approach appears increasingly intractable. Real-time recognition is thus excluded for large-scale applications. Other subspace methods such as independent component analysis (ICA) (Draper et al., 2003; Bartlett et al., 2002) and linear discriminant analysis (LDA) (Etemad and Chellappa, 1997; Lu et al., 2003) suffer from similar drawbacks.

3 BEST POINTS INTERPOLATION METHOD

In this section, we extend the best points interpolation method developed earlier in (Nguyen et al., 2006) to face reconstruction and recognition. The basic ingredients of the method are a stable interpolation procedure and a set of interpolation points.

3.1 Interpolation Procedure

Let us recall the pixel grid Ξ and the face space $\Phi_k = \mathrm{span}\{\phi_1, \ldots, \phi_k\}$. In this space, we shall seek an approximation of any input image u. Rather than performing the projection onto the face space for the best approximation, we pursue an interpolation as follows.

In particular, we aim to find an approximation $\tilde{u} \in \Phi_k$ of u via $m(\geq k)$ interpolation points $\{\mathbf{z}_j \in \Xi\}, 1 \leq j \leq m$, such that

$$\tilde{u} = \sum_{i=1}^{k} \tilde{a}_i \phi_i \tag{9}$$

where the coefficients $\tilde{a}_i$ are the solution of

$$\sum_{i=1}^{k} \phi_i(\mathbf{z}_j)\tilde{a}_i = u(\mathbf{z}_j), \quad j = 1, \ldots, m \;. \tag{10}$$

We define the associated error as

$$\tilde{\varepsilon} = \|u - \tilde{u}\| \;. \tag{11}$$

In general, the linear system (10) is over-determined because there are more equations than unknowns.

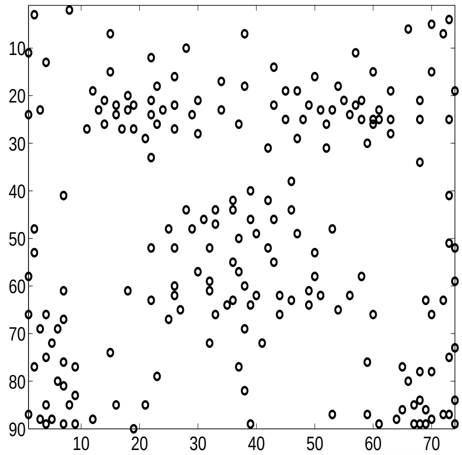

Figure 2: Distribution of the interpolation points on the pixel grid for $k = 100$ and $m = 200$.

Hence, the interpolant coefficient vector $\tilde{\mathbf{a}} = [\tilde{a}_1, \ldots, \tilde{a}_k]^T$ is determined from

$$\mathbf{C}^T\mathbf{C}\tilde{\mathbf{a}} = \mathbf{C}^T\mathbf{c} \;, \tag{12}$$

where $\mathbf{C} \in \mathbb{R}^{m\times k}$ with $C_{ji} = \phi_i(\mathbf{z}_j), 1 \leq i \leq k, 1 \leq j \leq m$ and $\mathbf{c} = [u(\mathbf{z}_1), \ldots, u(\mathbf{z}_m)]^T$. It thus follows that

$$\tilde{\mathbf{a}} = \mathbf{B}\mathbf{c} \;. \tag{13}$$

Here the matrix $\mathbf{B} = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T$ is precomputed and stored. Therefore, for any new face u, the cost of evaluating the interpolant coefficient vector $\tilde{\mathbf{a}}$ is only $O(mk)$ and becomes $O(k^2)$ when $m = O(k)$.

Obviously, the approximation quality depends crucially on the interpolation pixels $\{\mathbf{z}_j\}$. Therefore, it is extremely important to choose $\{\mathbf{z}_j\}$ so as to guarantee accurate and stable interpolation. For instance, Figure 2 shows the interpolation points for $k = 100$ and $m = 200$ obtained using our method described below. We see that the pixels are distributed somewhat symmetrically with respect to the symmetry line of the face and largely allocated around main locations of the face such as eyes, nose, mouth, and jaw.

3.2 Interpolation Points

We proceed by describing our approach for determining the interpolation points. The crucial observation is that much of the surface of a face is smooth with regular texture and that faces are similar in appearance and highly constrained; for example, the frontal view of a face is symmetric. Moreover, the value of a pixel is typically highly correlated with the values of the surrounding pixels. Therefore, a large number of pixels in the image space does not represent physically possible faces and only a small number of pixels may suffice to represent facial characteristics.

To begin, we introduce a set of images, $\mathcal{U}_K^* = \{u_\ell^*\}, 1 \le \ell \le K,$, where u_ℓ^* is the best approximation to u_ℓ. It thus follows that

$$u_\ell^* = \sum_{i=1}^{k} a_{\ell i} \phi_i, \quad (14)$$

where for $1 \le i \le k, 1 \le \ell \le K$,

$$a_{\ell i} = \phi_i^T u_\ell \ . \quad (15)$$

We then determine $\{\mathbf{z}_j\}, 1 \le j \le m$ as a minimizer of the following minimization

$$\min_{\mathbf{x}_1 \in \Xi, \ldots, \mathbf{x}_m \in \Xi} \sum_{\ell=1}^{K} \sum_{i=1}^{k} \left(a_{\ell i} - \tilde{a}_{\ell i}(\mathbf{x}_1, \ldots, \mathbf{x}_m)\right)^2 \quad (16)$$

$$\sum_{i=1}^{k} \phi_i(\mathbf{x}_j) \tilde{a}_{\ell i} = u_\ell(\mathbf{x}_j), \quad 1 \le j \le m, 1 \le \ell \le K.$$

We shall call the $\mathbf{z}_j$ as *best interpolation points*, because the points are optimal for the interpolation of the best approximations u_ℓ^*. We refer the reader to (Nguyen et al., 2006) for details on the solution procedure.

3.3 Application to Face Recognition

We apply the method to develop a fully automatic *real-time* face recognition system involving the generation stage and the recognition stage. The detailed implementation of the system is given below:

1. Determine the dimension of the face space k and then calculate $\phi_1, \ldots, \phi_k$.
2. Compute and store $\{\mathbf{z}_j\}$, $\mathbf{B} = \left(\mathbf{C}^T \mathbf{C}\right)^{-1} \mathbf{C}^T$. Recall that $C_{ji} = \phi_i(\mathbf{z}_j), 1 \le i \le k, 1 \le j \le m$.
3. For a "gallery" of images $\mathcal{V}_{K'} = \{v_i\}, 1 \le i \le K'$, compute $\tilde{\mathbf{a}}_i = \mathbf{B}[v_i(\mathbf{z}_1), \ldots, v_i(\mathbf{z}_m)]^T, 1 \le i \le K'$. (Note $\mathcal{V}_{K'}$ can be the same or different from $\mathcal{U}_K$).
4. For each new face to be classified u, calculate its interpolant coefficient vector $\tilde{\mathbf{a}}$ from (13) and find
$$i_{\min} = \arg \min_{1 \le i \le K'} \|\tilde{\mathbf{a}} - \tilde{\mathbf{a}}_i\| \ . \quad (17)$$
5. If $\|\tilde{\mathbf{a}} - \tilde{\mathbf{a}}_{i_{\min}}\|$ is less than a chosen threshold, the input image u is identified as the individual associated with the gallery image $i_{\min}$. Otherwise, the image is classified as a new individual.

The generation stage (steps 1–2) is computationally expensive, but only performed when the training set changes. However, the recognition stage (steps 4–5) is very inexpensive: the calculation of $\tilde{\mathbf{a}}$ takes $O(mk)$; and the nearest-neighbor search problem (17) which can be solved typically in $O(kK'^{0.25})$ (Andoni and Indyk, 2006). Hence, if K' is in order of $O(k^4)$ or less, the computational cost is only $O(k^2)$. This is often the case even for large-scale applications; for example, for a training database of 10^4 images, one would need more (or many more) than 10 eigenfaces to achieve acceptable recognition rates.

In summary, the operation count of the recognition stage is about $O(mk)$. The computational complexity of our system is thus *independent* of N. As mentioned earlier, the complexity of PCA-based algorithms is at least $O(Nk)$. Our approach leads to a computational reduction of N/m relative to PCA. Since m is typically much smaller than N, significant savings are expected. The savings per image classification certainly translate to real-time performance especially when many face images need to be classified simultaneously.

4 EXPERIMENTS

In practice, some applications of face recognition regard the recognition quality more importantly than the computational performance. Therefore, in order to be useful and gain acceptance, our approach must be tested and compared with existing approaches, particularly here with the PCA.

4.1 Face Database

The AT&T face database (Samaria and Harter, 1994) consists of 400 images of 40 individuals (10 images per individual). The images were taken at different times with variation in lighting, poses, and facial expressions, with and without glasses. The images were cropped and resized by us to a resolution of 74×90. We formed a training ensemble of 400 images by using 200 images of the database, 10 each of 20 different individuals, and including 200 mirror images of these images (Kirby and Sirovich, 1990).

The testing set contains the (200) remaining images of 20 individuals not belonging to the training ensemble. We further divide the testing set into the gallery of 20 individual faces and 180 probe images containing 9 views of every individual in the gallery. The recognition task is to match the probe images to the 20 gallery faces. The fact that the training and testing sets have no common individual serves to assess the performance of a face recognition system more critically — the ability to recognize new faces which are not part of the face space constructed from the training set.

Figure 3: The reconstruction results for a familiar face. The BPIM reconstructed images are placed at the top row for $k = 40, 80, 120, 160$ (from left to right) and $m = 2k$. The PCA reconstructed images are placed at the second row for $k = 40, 80, 120, 160$ (from left to right). The original face is shown on the right.

Figure 4: Reconstruction of a familiar face (bottom right) from a 10% mask (top right) with only the white pixels. The reconstructed images are shown at the top row for $k = 40, 80, 120, 160$ (from left to right) and $m = 2k$. The PCA reconstructed images are shown at the second row for $k = 40, 80, 120, 160$ (from left to right) with using all the pixels.

4.2 Results for Face Reconstruction

We first present in Figure 3 the reconstruction results for a face in the training ensemble. The BPIM produces reconstructions almost as well as PCA: most facial features captured by the PCA reconstructed images also appear in the BPIM reconstructed images. We underline the fact that the interpolation method requires less than 5% of the total number of pixels $N = 6660$, but delivers quite satisfactory results.

To illustrate the use of the interpolation approach for reconstructing a full image from a partial image, we consider a face (in the training set) shown at the bottom right and a mask shown at the top right in Figure 4. This is a relatively extreme mask that obscures 90% of the pixels in a random manner. Because the masked face may not have intensity values at all the best interpolation points, we need to define a new set of interpolation points. To this end, we keep the best interpolation points which coincide with some of the white pixels of the masked face and replace the remaining best pixels with the "nearest" white pixels. In Figure 4, the reconstructed images using those interpolation points are compared with the PCA reconstructed images utilitizing all the pixels. Although the interpolation procedure does not recover the original face exactly, the construction is visually close to the "best" reconstruction.

4.3 Results for Face Recognition

We apply the face recognition system developed in Section 3.4 to classify the probe images. We illustrate in Figure 5 the recognition accuracy as a function of k for the BPIM and PCA. As it may be expected, the BPIM yields smaller recognition rates than PCA. However, as k increases, the BPIM gives recognition rates which are quite comparable to those of PCA for large enough k: PCA achieves a recognition rate of 74.98%, while PBIM results in a recognition rate of 73.66% for $k = 80$. In many applications, the small accuracy loss of only 1.32% is paid off very well by the significant reduction of $6660/160 (> 40)$ in complexity. This is confirmed in Table 1 which shows the computational times for the BPIM and PCA. The values are normalized with respect to the time to recognize a face for $k = 10$ and $m = 20$ with the BPIM. Clearly, the BPIM is significantly faster than PCA. This important advantage is very useful to applications that requires a real-time recognition capability.

Table 1: Computational times (normalized with respect to the time to recognize a face for $k = 10$ and $m = 20$ with the BPIM) for the BPIM and PCA at different values of k.

k	BPIM	PCA
10	1.00	333.30
20	4.20	592.67
30	9.33	873.34
40	15.60	1107.66
50	26.12	1437.35
60	36.47	1708.02
70	47.93	1958.94
80	61.47	2293.73

Finally, in order to demonstrate the classification of incomplete images, we consider a random chosen mask of 10% pixels shown in Figure 6. Next to the mask, we show a few faces which are correctly recognized with using the interpolation procedure when their intensity values are available only at the white pixels of the mask. Note the interpolation points are chosen in the same way as before.

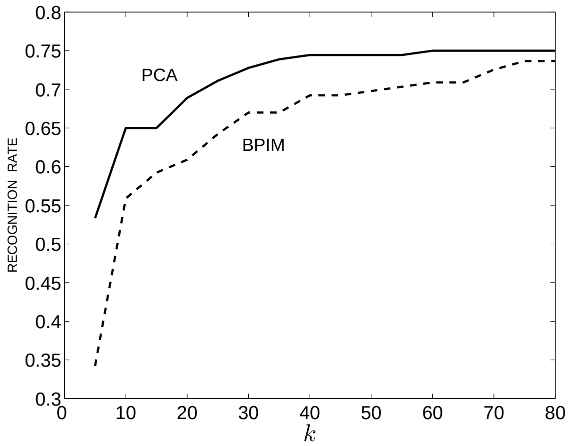

Figure 5: Recognition accuracy of PCA and BPIM with increasing the number of eigenfaces k. Note that the BPIM uses $m = 2k$ best interpolation points.

Figure 6: Recognition of incomplete face images. The 10% mask on the left is followed by a few faces which are correctly recognized with using the interpolation procedure.

5 CONCLUSION

We have presented an interpolation method for the reconstruction and recognition of face images. It is important to note that PCA uses full knowledge of the data in the reconstruction process. In contrast, our method uses only partial knowledge of the data. Therefore, the method is very useful to the restoration of a full image from a partial image. Based on the method, we have also developed a fully automatic real-time face recognition system. The system is shown to be able to recognize incomplete images. Moreover, the computational cost of recognizing a new face is only $O(mk)$, translating to a saving of N/m relative to PCA approach. Typically, since N is $O(10^4)$ and m is $O(10^2)$, this implies two orders of magnitude less expensive computationally than PCA. The significant reduction in time should enable us to tackle very large problems. Hence, it is imperative to test our system on a larger database such as the FERET database. We plan to pursue this direction in future research.

ACKNOWLEDGEMENTS

We would like to thank Professor A. T. Patera of MIT for his long-standing collaboration and many invaluable contributions to this work. The authors also thank AT&T Laboratories Cambridge for providing the ORL face database. This work was supported by the Singapore-MIT Alliance.

REFERENCES

Andoni, A. and Indyk, P. (2006). Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *Proceedings of the 47th IEEE Sym. on Foundations of Computer Science*, Berkeley, CA.

Bartlett, M. S., Movellan, J. R., and Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Trans. on Neural Networks*, 13:1450–1464.

Delac, K., Grgic, M., and Grgic, S. (2005). Independent comparative study of pca, ica, and lda on the feret data set. *International Journal of Imaging Systems and Technology*, 15(5):252–260.

Draper, B. A., Baek, K., Bartlett, M. S., and Beveridge, J. R. (2003). Recognizing faces with pca and ica. *Computer Vision and Image Understanding*, 91(1-2):115–137.

Etemad, K. and Chellappa, R. (1997). Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, 14(8):1724–1733.

Everson, R. and Sirovich, L. (1995). Karhunen-loeve procedure for gappy data. *Opt. Soc. Am. A*, 12(8):1657–1664.

Kirby, M. and Sirovich, L. (1990). Application of the karhunen-loève procedure for the characterization of human face. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:103–108.

Lu, J., Plataniotis, K. N., and Venetsanopoulos, A. N. (2003). Face recognition using lda-based algorithms. *IEEE Trans. on Neural Networks*, 14(1):195–200.

Nguyen, N. C., Patera, A. T., and Peraire, J. (2006). A best points interpolation method for efficient approximation of parametrized functions. *International Journal of Numerical Methods in Engineering*. Submitted.

Samaria, F. S. and Harter, A. C. (1994). Parameterisation of a stochastic model for human face identification. In *Proceedings of the 2nd IEEE workshop on Applications of Computer Vision*, pages 138–142, Sarasota, Florida.

Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4:519–524.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.