Open access • Proceedings Article • DOI:10.1117/12.386642

Real Time Face Recognition using Eigenfaces — Source link

Raphael Cendrillon, Brian C. Lovell

Institutions: University of Queensland

Published on: 30 May 2000 - Visual Communications and Image Processing

Topics: Three-dimensional face recognition, Eigenface and Facial recognition system

Real-Time Face Recognition using Eigenfaces

Raphael Cendrillon[a] and Brian C. Lovell[b]

[a] Department of Computer Science and Electrical Engineering
The University of Queensland, Brisbane, Australia, 4072
lovell@csee.uq.edu.au
[b]CRC for Sensor Signal and Information Processing
Department of Computer Science and Electrical Engineering
The University of Queensland, Brisbane, Australia, 4072
cendrill@csee.uq.edu.au

ABSTRACT

In recent years considerable progress has been made in the area of face recognition. Through the development of techniques like eigenfaces, computers can now compete favourably with humans in many face recognition tasks, particularly those in which large databases of faces must be searched. Whilst these methods perform extremely well under constrained conditions, the problem of face recognition under gross variations in expression, view, and lighting remains largely unsolved. This paper details the design of a real-time face recognition system aimed at operating in less constrained environments. The system is capable of single scale recognition with an accuracy of 94% at 2 frames-per-second. A description of the system's performance and the issues and problems faced during it's development is given.

Keywords: face recognition, real time, eigenface, computer vision, temporal filtering

1. INTRODUCTION

Automated face recognition systems with the ability to detect and recognize faces in crowds have many potential applications. In 1998 alone the Australian government spent $247 million on law enforcement activities. A substantial amount of this money is consumed by surveillance and general monitoring of public areas. Handing these tasks over to an automated system would present substantial savings in both expenditure and manpower. For example, a system could monitor a large number of closed circuit TVs looking for known criminals, drug offenders, and other interesting persons, and then notifying the authorities when one is located. Other applications could be airport surveillance, access control for computers and buildings, added security for automatic teller transactions, and improved human-computer interaction.

To this end, many face recognition techniques have been proposed. These include the use of elastic grid matching,[1] correlation,[2] and eigenfaces.[3] Of these techniques eigenface-based techniques appear to have had the most success, consistently winning the annual international competition based on the Ferret database.

2. EIGENFACES

Many techniques applied to the automated face recognition problem make arbitrary decisions on which facial characteristics are actually important for recognition. For example, correlation-based techniques assume that all pixels of a facial image are equally important, when this is actually not the case at all. Another difficulty with other techniques is that they presume some significance of certain facial characteristics over others with no grounds for this presumption.

A good example of this is the use of distance measures between facial *key-points*, where it is assumed that faces can be uniquely identified by the distance between certain facial features. Whether this is true or not is debatable, but it is certainly true that these are not the only features of importance to face recognition, nor are they necessarily the most significant. The scaling or normalization of facial features according to their relative importance in face recognition is the basic premise behind the eigenfaces technique.

Send correspondence to Brian Lovell

In *Visual Communications and Image Processing 2000*, King N. Ngan, Thomas Sikora, Ming-Ting Sun, Editors, Proceedings of SPIE Vol. 4067 (2000) • 0277-786X/00/$15.00

The eigenface technique attempts to capture the underlying variations between facial images in an orthogonal set of basis vectors refered to as eigenfaces. The eigenfaces are thus the image vectors that map the most significant variations between faces, commonly referred to as principle components. Under the assumption that faces form a simply connected region in image space, we can represent any face as a linear combination of eigenfaces. Each face can thus be represented by a weight vector, which contains the proportions of each eigenface needed to construct that face as in Figure 1. By comparing the weight vector of an unknown face to a database of known faces, the closest match can be determined.

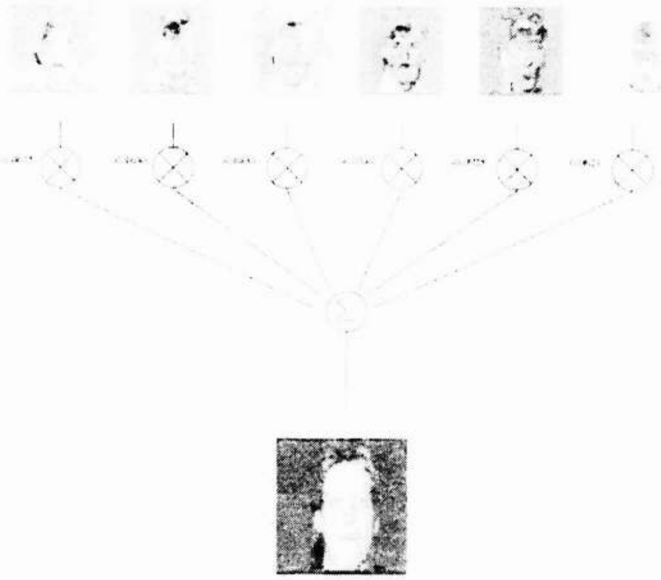

Figure 1. Reconstruction of a face using principal components.

Unlike the correlation based technique, eigenfaces are robust against noise, poor lighting conditions, and partial occlusion.[4] This coupled with their robustness against variations in scale and rotation yields an advantage over techniques based on distance between facial key-points. Eigenfaces require no training and so are not subject to many of the learning problems associated with neural networks. Eigenface-based systems have also been shown to maintain accuracy even with large scale databases (on the order of 1000 faces). Finally eigenfaces have been shown to run in real-time speeds on low-end workstations. This means they are suitable for a real-time system based on PC hardware with little or no hardware acceleration.

3. CALCULATION OF EIGENFACES FROM IMAGE DATABASE

So how do we go about finding this optimal co-ordinate system, or optimal basis for facial images? Components of this optimal basis will be orthogonal and will maximize the variance in the set of facial images as we discussed previously. So we need some way of evaluating the variance in the facial image set along a given basis. This is precisely what eigenvectors achieve.

Given a matrix C the eigenvectors u and eigenvalues λ of C satisfy

$$Cu = \lambda u \tag{1}$$

The eigenvectors are orthogonal and normalized hence

$$u_i^T u_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{2}$$

Let Γ_k represent the column vector of face k obtained through lexographical ordering of $I_k(x, y)$. Here face k is any face in the face set. Now let us define ϕ_k as the mean normalized column vector for face k. This means that

$$\phi_k = \Gamma_k - \psi \tag{3}$$

where

$$\psi = \frac{1}{M}\sum_{k=1}^{M}\Gamma_i \tag{4}$$

Now let C be the covariance matrix of the mean normalized faces.[4]

$$C = \frac{1}{M}\sum_{k=1}^{M}\phi_k\phi_k^T \tag{5}$$

Note that M is the number of facial images in our representative set. These facial images help to characterize the sub-space formed by faces within image space. This sub-space will henceforth be referred to as *face-space*. From (1)

$$\begin{aligned} Cu_i &= \lambda_i u_i && (6)\\ u_i^T C u_i &= u_i^T \lambda_i u_i \\ &= \lambda_i u_i^T u_i && (7) \end{aligned}$$

now since $u_i^T u_i = 1$

$$\begin{aligned} u_i^T C u_i &= \lambda_i && (8)\\ \lambda_i &= \frac{1}{M} u_i^T \sum_{k=1}^{M} \phi_k \phi_k^T u_i \\ &= \frac{1}{M}\sum_{k=1}^{M} u_i^T \phi_k \phi_k^T u_i \\ &= \frac{1}{M}\sum_{k=1}^{M} (u_i \phi_k^T)^T (u_i \phi_k^T) \\ &= \frac{1}{M}\sum_{k=1}^{M} (u_i \phi_k^T)^2 \\ &= \frac{1}{M}\sum_{k=1}^{M} (u_i \Gamma_k^T - mean(u_i \Gamma_k^T))^2 \\ &= \frac{1}{M}\sum_{k=1}^{M} var(u_i \Gamma_k^T) && (9) \end{aligned}$$

Thus eigenvalue i represents the variance of the representative facial image set along the axis described by eigenvector i. So by selecting the eigenvectors with the largest eigenvalues as our basis, we are selecting the dimensions which can express the greatest variance in facial images or the dominant modes of face-space. Using this co-ordinate system, a face can be reasonably reconstructed with as few as 6 co-ordinates. This means that a 128x128 pixel face which previously took 16,384 bytes to represent in image space now requires only 6 bytes. Once again, this reduction in dimensionality makes the problem of face recognition much simpler since we concern ourselves only with the relevant and most discriminatory attributes of the face.

4. SYSTEM ARCHITECTURE

The architecture of the system is depicted in Figure 2. The face recognition system developed comprises five major processing modules which are:

- Temporal Filters: Segments regions of movement from the incoming video feed or scene. It is predicted that faces will be located within these regions.
- Face Location: Locates the face within the regions of movement detected by the temporal filters. The face is then segmented and passed on to the Feature Location module.
- Feature Location: Determines eye locations in the segmented face. These locations are used by the Scale and Rotation Normalization module to compensate for variations in scale and rotation.
- Scale and Rotation Normalization: Compensates for variations in scale and rotation.
- Lighting Normalization: Accepts the geometrically normalized face produced by the Scale and Rotation Normalization module and compensates for variations in lighting conditions. These include global lighting changes as well as non-uniform gradients.
- Face Recognition: Projects the normalized face in to face-space to determine it's identity. A certainty measure is also produced.

The system was developed on a PC with Pentium II-400 MHz processor with 128 MB RAM, an Asus TNT video capture card, and a Flexcam video camera. The frame rate of the face recognition system on this platform was 2 FPS (frames per second).

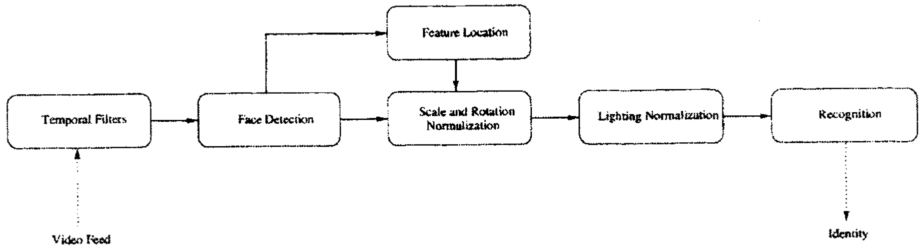

Figure 2. System architecture.

4.1. Temporal filters

Faces are non-static objects. Even when we sit still our face is constantly moving as a result of breathing, variations in expression, speaking etc. Temporal filters allow us to determine regions of movement within a scene. By searching for faces only within these regions, we can reduce false-alarms and increase overall computational efficiency.

Simplistic methods for locating regions of movement include background subtraction and frame differencing. These tend to be problematic however as uniform changes in ambient lighting result in decreased performance. One technique which can be used to compensate for this uses several frames to locate regions of movement.[5] By determining the weighted sum of several frames, this technique decreases sensitivity to lighting variations.

A temporal filter of this type was implemented in the face recognition system. It operates by convolving several frames in the temporal dimension with a Gaussian second derivative $\frac{\partial^2}{\partial t^2}G(t)$.[6]

If we consider global, large scale variations in lighting, they will tend to occupy the lower end of the frequency spectrum. As a result, high-pass filtering will tend to remove sensitivity to changes in ambient lighting. This is

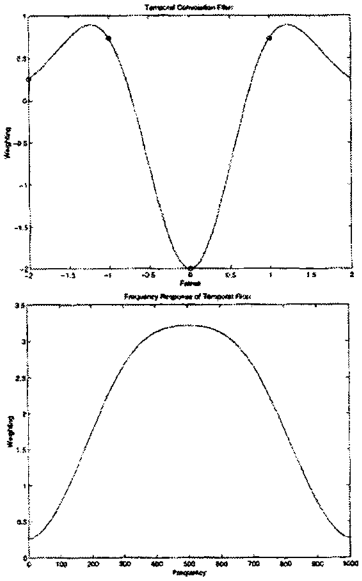

Figure 3. Temporal convolution filter and its frequency response.

essentially what convolving with the temporal filter achieves. As can be seen from Figure 3, the temporal filter does indeed act like a high-pass filter.

Once temporal filtering has taken place, zero-crossings in the resultant matrix are located.[6] This is achieved through convolution with the horizontal and vertical step functions to locate horizontal and vertical zero-crossings. The zero-crossing maps are then threshold to generate binary maps, and logically OR'ed together.

$$vertical\,step\,function = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

$$horizontal\,step\,function = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$$

Zero-crossings give an indication of regions of movement. Once these have been found the next task is to estimate the person's location. This is currently implemented with a simplistic technique based on the assumption that only one person will be in the scene at any time. This technique locates the vertical and horizontal centroids of the zero-crossings, and their corresponding standard deviations. The head is then estimated to lie within 2 horizontal and 1 vertical standard deviation of the centroids.

Although this method reduced sensitivity to lighting changes, it's performance dropped when there was little movement in the scene. As a result, the algorithm was modified to include a weighting factor which is inversely proportional to the number of detected zero-crossings. This weighting has been implemented by specifying a horizontal and vertical span default which is proportional to both the standard deviation and the number of zero-crossings. It was observed that the system performed well when a proportionality to the square root of the number of detected zero-crossings was used.

The horizontal and vertical spans are thus defined as

$$\begin{aligned} horizontal\,span &= \frac{\sigma_{horizontal}}{\sqrt{N}} \\ vertical\,span &= \frac{\sigma_{vertical}}{\sqrt{N}} \end{aligned} \tag{10}$$

The temporal filter now works as follows. When there is a large amount of movement in the scene, the system postulates multiple individuals, and hence become more selective in its detected regions of movement, decreasing the horizontal and vertical spans. On the other hand, when there is little movement within the scene, the temporal filter becomes more accepting in it's detected regions of movement and the horizontal and vertical spans are increased.

The idea of using this form of adaptive temporal filtering has not been encountered in literature and is considered to be a novel approach. This technique could be expanded to the tracking of multiple individuals through the use of clustering and Kalman filters.

The whole temporal filtering process is illustrated in Figure 4. The first image contains the result of temporal convolution; the second image contains the result of zero-crossing detection and the third image contains the predicted head location. Once the predicted head location is found, this information is passed to the Face Location module.

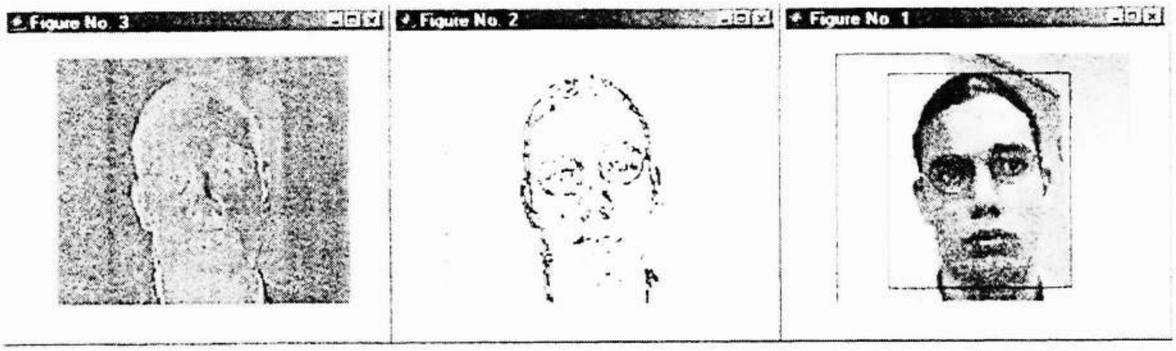

Figure 4. Temporal filtering.

4.2. Face location

Face location and segmentation was accomplished by two means. First moving regions of the scene were located and postulated as areas that may contain a face. Temporal filters were responsible for locating moving regions prior to face location to increase efficiency and robustness. The co-ordinates of these regions were then passed to a face detection algorithm that found the most probable location of a face through a 'face-similarity' measure based on projection into face space.

Thus Eigenfaces can also be applied to *detect* the location of faces using the face-space reconstruction error.[3] Whilst this technique experienced much success at single-scale face location, multi-scale location was more problematic. As a result of the higher noise sensitivity experienced at smaller scales, the face-space reconstruction error tends to be larger at smaller scales. This caused our system to consistently recognize faces at the larger scales whether they occurred or not. It is proposed that the use of scale-based eigenspaces will overcome this problem, however our system is currently operational only in single-scale mode.

4.3. Face Normalization

Once located, faces need to be normalized for rotation, scale and lighting prior to recognition. If this does not occur considerable performance degradation will be experienced as a result in the increased dimensionality of the problem. Our system used eigenfeatures[3] to locate the left and right eyes of the face. These points were then used to normalize the image for variations in scale and rotation.

Facial images were also normalized for lighting by setting the mean to zero and the L2-norm to unity. More advanced lighting techniques were evaluated including Sobell filters and non-linear block processing, but these were found to actually decrease performance (although the reason is not known at this stage).

4.4. Face recognition

Face recognition was performed using the Mahalanobis distance and eigenfaces. An offline recognition rate of 95.5% was recorded. Furthermore, it was noted that recognition performance degraded significantly (less than 70%) for rotations of more than 10 degrees and size reductions less than 88% of the original image size.

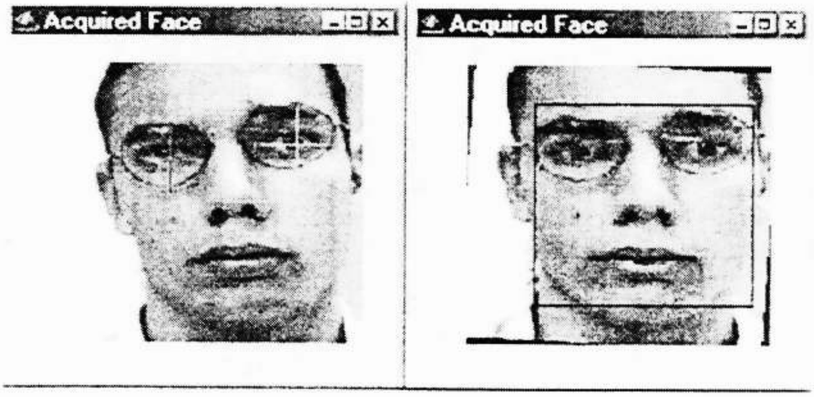

Figure 5. Scale and rotation normalization using facial keypoints.

5. INTEGRATED SYSTEM PERFORMANCE

The integrated system was evaluated in real-time over 100 trials with 5 individuals. Recognition error was found to be 94% and the integrated system runs at 2 frames-per-second on a Pentium II with 400 MHz clock frequency. Operation of the system as a face detector is illustrated in Figure 6.

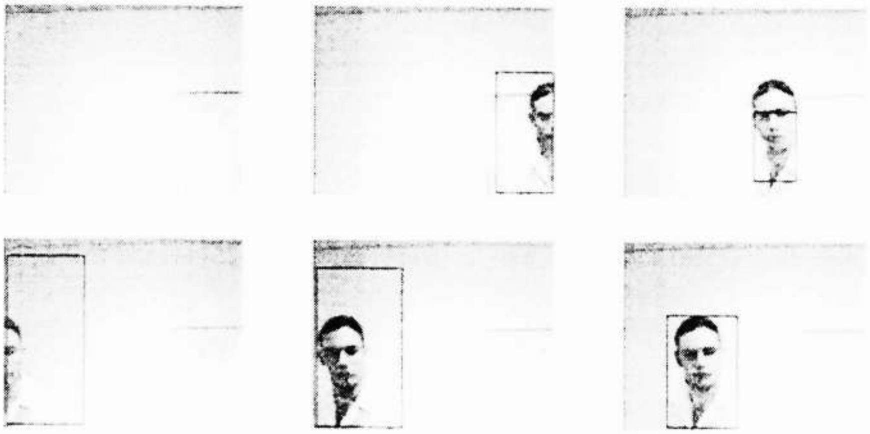

Figure 6. Temporal filters and face location module used for tracking.

6. CONCLUSIONS AND FUTURE WORK

A face recognition system with a near real-time frame rate and high recognition accuracy has been developed. The biggest issue currently facing our system is scale sensitivity. While system performance is high in single scale operation, performance drops considerably as the subject moves out of scale. Multi-scale operation is essential for applications in unconstrained environments and hence should be the focus of future work.

REFERENCES

1. M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, and R. Wurtz, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on Computing* **43**(3), pp. 300 311, 1993.
2. M. Kosugi, "Human-face search and location in a scene by multi-pyramid architecture for personal identification," *Systems and Computers in Japan* **26**(6), pp. 27 38, 1995.

3. A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," Tech. Rep. 245, Perceptual Computing Section, Media Laboratory, MIT, 1998.
4. M. Turk and A. Pentland, "Eigenfaces for recognition," *journal of Cognitive Neuroscience* **3**, pp. 71–86, March 1991.
5. D. Reisfeld and Y. Yeshurun, "Preprocessing of face images: Detection of features and pose normalization," *Computer Vision and Image Understanding* **71**, pp. 413–430, September 1997.
6. S. McKenna and S. Gong, "Non-intrusive person identification for access control by visual tracking and face recognition," in *Proceedings of the First International Conference on Audio and Video-Based Biometric Person Identification*, pp. 187–189, Springer-Verlag, 1997.