Facial Recognition with Singular Value Decomposition

Guoliang Zeng
Phone: (480) 727-1905; Fax: (480) 727-1723; E-mail: gzeng@asu.edu
Motorola DSP Laboratory, Division of Computing Studies
Arizona State University Polytechnic Campus
7001 East Williams Field Rd., ISTB3, Mesa, AZ 85212

ABSTRACT:
This paper implements a real-time system to recognize faces. The approach is essentially to apply the concepts of vector space and subspace to face recognition. The set of known faces with $m \times n$ pixels forms a subspace, called "face space", of the "image space" containing all images with $m \times n$ pixels. This face space best defines the variation of the known faces. The basis of the face space is defined by the singular-vectors of the set of known faces. These singular-vectors do not necessarily correspond to the distinct features like ears, eyes and noses. The projection of a new image onto this face space is then compared to the available projections of known faces to identify the person. Since the dimension of face subspace is much less than the whole image space, it is much easier to compare projections than origin images pixel by pixel. Based on the above idea, a Singular Value Decomposition (SVD) approach is implemented in this paper. The framework provides our system the ability to learn to recognize new faces in a real-time and automatic manner.

KEYWORDS:
Image processing, Face recognition, Singular value decomposition.

I. INTRODUCTION:

Over the last ten years or so, face recognition has become an active area of research in computer vision, neuroscience, and psychology. It is the general opinion that advances in computer vision research will provide useful insights to neuroscientists and psychologists into how human brain works, and vice versa.

Some of the applications of the facial recognition technology include:

- Security
- Computer-human interaction
- System (key) access based on face/voice recognition.
- Tracking people, either spatially with a large network of cameras or temporally by monitoring the same camera over time.
- Locating people in large image data.

Several approaches to face recognition have been proposed for the 2-dimensional facial recognition. Each has its own advantages and limitations. Much of the work has focused on detecting individual features such as eyes, nose, mouth, and head outline, and defining a face model by the position, size, and relationships among these features [1][2]. The face recognition strategies have modeled and classified faces based on normalized distances and ratios among feature points. Such approaches have proven difficult to extend to multiple views, and have often been quite fragile. Like Principal Component Analysis (PCA), our SVD approach treats a set of known faces as vectors in a subspace, called "face space", spanned by a small group of "base-faces" [3]. Recognition is performed by projecting a new image onto the face space, and then classifying the face by comparing its coordinates (position) in face space with the coordinates (positions) of known faces. But the SVD approach has better numerical properties than PCA.

II. PRINCIPALES:

K. Elleithy (ed.), Advances and Innovations in Systems, Computing Sciences and Software Engineering, 145–148.
© 2007 *Springer*.

Assume each face image has $m \times n = M$ pixels, and is represented as an $M \times 1$ column vector $\mathbf{f}_i$. Then, a 'training set' S with N face images of known individuals forms an $M \times N$ matrix:

$$S = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_N] \tag{1}$$

The mean image $\bar{\mathbf{f}}$ of set S, is given by

$$\bar{\mathbf{f}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{f}_i \tag{2}$$

Subtracting $\bar{\mathbf{f}}$ from the original faces gives

$$\mathbf{a}_i = \mathbf{f}_i - \bar{\mathbf{f}},,, i = 1,2,...N \tag{3}$$

This gives another $M \times N$ matrix A:

$$A = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N] \tag{4}$$

Assume the rank of A is r, and $r \le N << M$. It can be proved that A has the following Single Value Decomposition (SVD):

$$A = U\Sigma V^T \tag{5}$$

Here, Σ is an $M \times N$ diagonal matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{r+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \sigma_N \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{6}$$

For i = 1, 2, …, N, σ_i are called Singular Values (SV) of matrix A. It can be proved that

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0, \text{ and}$$
$$\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_N = 0. \tag{7}$$

Matrix V is an $N \times N$ orthogonal matrix:

$$V = [\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_r, \mathbf{v}_{r+1}, ..., \mathbf{v}_N] \tag{8}$$

That is the column vectors $\mathbf{v}_i$, for i = 1, 2, …, N, form an orthonormal set:

$$\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij} = \begin{cases} 1,,, i = j \\ 0,,, i \ne j \end{cases} \tag{9}$$

Matrix U is an $M \times M$ orthogonal matrix:

$$U = [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_r, \mathbf{u}_{r+1}, ..., \mathbf{u}_M] \tag{10}$$

That is the column vectors $\mathbf{u}_i$, for i = 1, 2, …, M, also form an orthonormal set:

$$\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij} = \begin{cases} 1,,, i = j \\ 0,,, i \ne j \end{cases} \tag{11}$$

The $\mathbf{v}_i$'s and $\mathbf{u}_i$'s are called right and left singular-vectors of A.

From (5),

$$AV = U\Sigma \tag{5a}$$

We get

$$A\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i,,, i = 1,2,..., r \\ 0,,,,,,, i = r+1,..., N \end{cases} \tag{12}$$

It can be proved that $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r\}$ form an orthonormal basis for $R(A)$, the range (column) subspace of matrix A. Since matrix A is formed from a training set S with N face images, $R(A)$ is called a 'face subspace' in the 'image space' of $m \times n$ pixels. Each $\mathbf{u}_i$, $i = 1,2,..., r$, can be called a 'base-face'.

Let $\mathbf{x}$ (= $[x_1, x_2, ..., x_r]^T$) be the coordinates (position) of any $m \times n$ face image $\mathbf{f}$ in the face subspace. Then it is the scalar projection of $\mathbf{f} - \bar{\mathbf{f}}$ onto the base-faces:

$$\mathbf{x} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]^T (\mathbf{f} - \bar{\mathbf{f}}) \tag{13}$$

This coordinate vector $\mathbf{x}$ is used to find which of the training faces best describes the

face **f**. That is to find some training face $\mathbf{f}_i$, $i = 1,2,...,N$, that minimizes the distance:

$$\varepsilon_i = \left\|\mathbf{x} - \mathbf{x}_i\right\|_2 = [(\mathbf{x} - \mathbf{x}_i)^T(\mathbf{x} - \mathbf{x}_i)]^{1/2} \quad (14)$$

where $\mathbf{x}_i$ is the coordinate vector of $\mathbf{f}_i$, which is the scalar projection of $\mathbf{f}_i - \bar{\mathbf{f}}$ onto the base-faces:

$$\mathbf{x}_i = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]^T(\mathbf{f}_i - \bar{\mathbf{f}}) \quad (15)$$

A face **f** is classified as face $\mathbf{f}_i$ when the minimum ε_i is less than some predefined threshold ε_0. Otherwise the face **f** is classified as "unknown face".

If **f** is <u>not</u> a face, its distance to the face subspace will be greater than 0. Since the vector projection of $\mathbf{f} - \bar{\mathbf{f}}$ onto the face space is given by

$$\mathbf{f}_p = [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_r]\mathbf{x} \quad (16)$$

where **x** is given in (13).

The distance of **f** to the face space is the distance between $\mathbf{f} - \bar{\mathbf{f}}$ and the projection $\mathbf{f}_p$ onto the face space:

$$\varepsilon_f = \left\|(\mathbf{f} - \bar{\mathbf{f}}) - \mathbf{f}_p\right\|_2 = [(\mathbf{f} - \bar{\mathbf{f}} - \mathbf{f}_p)^T(\mathbf{f} - \bar{\mathbf{f}} - \mathbf{f}_p)]^{1/2} \quad (17)$$

If ε_f is greater than some predefined threshold ε_1, then **f** is not a face image.

Notes:

1). In practice, a smaller number of base-faces than *r* is sufficient for identification, because accurate reconstruction is not a requirement. This smaller number of significant base-faces is chosen as those with the largest associated singular values.

2). In the training set, each individual could have more than one face images with different angles, expressions, and so on. In this case, we can use the average of them for the identification.

III. STEPS:

1. Obtain a training set *S* with *N* face images of known individuals. An example of *S* is shown in Fig. 1.

2. Compute the mean face $\bar{\mathbf{f}}$ of set *S* by (2). The mean face of the set in Fig. 1 is shown in Fig. 2. Form matrix *A* in (4) with the computed $\bar{\mathbf{f}}$.

3. Calculate the SVD of *A* as shown in (5). The result base-faces are shown in Fig. 3.

4. For each known individual, compute the coordinate vector $\mathbf{x}_i$ from (15).

5. Choose a threshold ε_1 that defines the maximum allowable distance from face space. Determine a threshold ε_0 that defines the maximum allowable distance from any known face in the training set *S*.

6. For a new input image **f** to be identified, calculate its coordinate vector **x** from (13), the vector projection $\mathbf{f}_p$, the distance ε_f to the face space from (17). If $\varepsilon_f > \varepsilon_1$ the input image is not a face.

7. If $\varepsilon_f < \varepsilon_1$, compute the distance ε_i to each known individual from (14). If all $\varepsilon_i > \varepsilon_0$, the input image may be classified as unknown face, and optionally used to begin a new individual face. If $\varepsilon_f < \varepsilon_1$, and some $\varepsilon_i < \varepsilon_0$, classify the input image as the known individual associated with the minimum ε_i ($\mathbf{x}_i$) in (14), and this image may optionally added to the original training set. Steps 1-5 may be repeated. This can update the system with more instances of known faces.

IV. CONCLUSIONS:

We tested our system with the following data:

- Image Size: M = 92 x 112 = 10,304.
- Number of known individuals: N = 20 -40.
- Number of face images per person: C = 1-5.
- Different Conditions: All frontal and slight tilt of the head, different facial expressions.

Essentially, a face image is of M (say 10,000) dimension. But the rank r of matrix A is less than or equals N (say 40). For most applications, a smaller number of base-faces than r is sufficient for identification. In this way, the amount of computation is greatly reduced. The Golub-Reinsch Algorithm is used for computing the singular value decomposition of a matrix. This can be done offline as part of the training. On machines with a 1.2 GHz clock rate, the recognition will take a time much less than 33 msec, a frame rate for real-time system.

The SVD approach is robust, simple, and easy and fast to implement which works well in a constrained environment. It provides a practical solution to the recognition problem. Instead of searching a large database of faces, it is better to give a small set of likely matches. By using base-faces, this small set of likely matches for given images can be easily obtained.

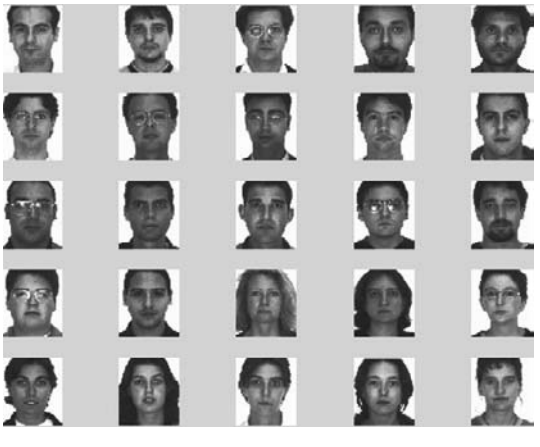

Fig. 1 Training set S

Fig. 2 Mean face $\bar{\mathbf{f}}$

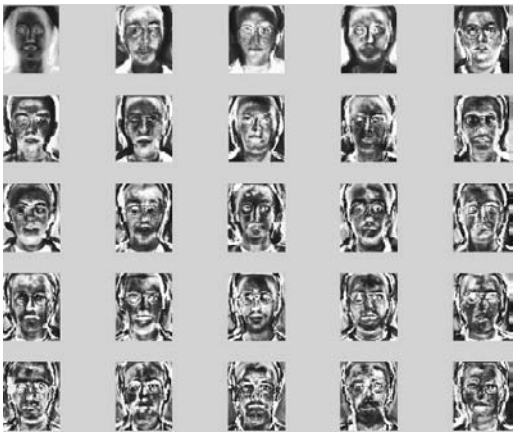

Fig. 3 Base-faces $\mathbf{u}_i$

REFERENCES:

[1] T. Kanade, “Picture Processing System by Computer Complex and Recognition of Human Faces,” Department of Information Science, Kyoto University, Nov.1973.

[2] A. L. Yuille, D. S. Cohen, and P. W. Hallinan, “Feature Extraction from Faces Using Deformable Templates,” *Proc. CVPR*, San Diego, CA June 1989.

[3] Matthew A. Turk, Pentland P. Alex(1991), “Face Recognition using Eigenface method”, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.586-591, 1991.