A Simulated Annealing and 2DPCA Based Method for Face Recognition

Zhijie Xu, Laisheng Wang, Liming Yang
College of Science, China Agricultural University, Beijing, China
E-mail: zhijiexu@163.com

Abstract

In this paper we address the problem of face recognition based on two-dimensional principal component analysis (2DPCA). The similarity measure plays an important role in pattern recognition. However, with reference to the 2DPCA based method for face recognition, studies on similarity measures are quite few. We propose a new method to identify the similarity measure by simulated annealing (SA), which is called SA similarity measure. Experimental results on two famous face databases show that the proposed method outperforms the state of the art methods in terms of recognition accuracy.

1. Introduction

Face recognition is an important subfield of pattern recognition, and it is one of the essential person identification methods since it is contactless. During the past two decades, face recognition has attracted much effort of researchers from research fields such as image processing, computer vision and pattern recognition. To date, a lot of approaches have been proposed for face recognition, the prominent method among which are the subspace method. The well-known Eigenfaces [1] and Fisherfaces [2] are some typical examples of this kind of method. In the Eigenfaces method, PCA is used for projecting the image space to a low dimensional feature space. The PCA based methods for face recognition have been developed since then. Recently, Yang and Yang [3] and Yang et al. [4] proposed two-dimensional PCA (2DPCA). Different from PCA, 2DPCA is founded upon 2D image matrix rather than 1D vectors. An image covariance matrix can be constructed directly by using the original image matrices. And a feature matrix for each training face sample can be computed easily. It has been proved that 2DPCA can give a better performance than PCA since the covariance matrix in 2DPCA can be computed more accurately and easily than that in PCA [5]. Reference [6] showed that 2DPCA is equivalent to a special case of the block-based PCA. Specially, the blocks are the row directional lines of the images. Inspired by 2DPCA, some other methods based on 2D image matrix were proposed, such as 2DLDA, 2D-DLDA, (2D)2DLDA and two dimensional Laplacianfaces method [7]–[11].

In 2D image matrix based methods for face recognition, much attention is paid to the feature extraction, while studies on the pattern classification aspect are quiet few. Yang and Yang [3] used the Frobenius distance measure in the 2DPCA-based method for face recognition. The Frobenius distance measure is a metric derived from the Frobenius matrix norm. Yang distance measure was proposed in [4], in which the sum of the Euclidean distance between feature vectors in feature matrix was used as the similarity measure. Zuo et al. [12] proposed an assembled matrix distance metric to measure the distance between two feature matrices, which is the generalized Yang distance. The above measures for 2DPCA are based on distance between feature vectors in feature matrix was used as the similarity measure. Zuo et al. [12] proposed an assembled matrix distance metric to measure the distance between two feature matrices, which is the generalized Yang distance. The above measures for 2DPCA are based on distance between feature vectors. Different from that, Meng and Zhang [13] proposed a matrix volume based classification measure called the volume measure, which is compatible with high-dimensional geometry theory. Their experimental results indicated that the method can outperform the typical 2DPCA-based method and PCA in face recognition. Nevertheless, the similarity measures mentioned above are fixed and can not adapt to the special data set. Since 2DPCA does not contain the discriminative information among different classes, these similarity measures have a drawback in that they do not have the ability to reject the influence of undesirable changes in a face pattern, such as changes due to illumination conditions, poses or expressions.

In this paper, we develop a simulated annealing and 2DPCA based method for face recognition. In
the method, the simulated annealing algorithm is used for learning a set of weights, which is used for computing the similarity measure between two face images. The learnt similarity measure adapts to the special face data set. Therefore, the proposed classification measure can reject the influence of undesirable changes in a face pattern to some extent. In this way, we can obtain a better recognition performance.

The remainder of this paper is organized as follows. In Section 2, the 2DPCA-based face recognition method, the previous works on similarity measures and the simulated annealing algorithm are introduced briefly. And the idea of using simulated annealing for learning a distance measure is described in Section 3. In Section 4, experimental results are presented and discussed. Finally, conclusions are presented in section 5.

2. Related works

2.1. Review of 2DPCA based face recognition

Given a training face set \( \{ X_1, X_2, \ldots, X_N \} \), 2DPCA uses all training samples to compute the following image total covariance matrix \( C \).

\[
C = E[(X - E(X))(X - E(X))^T] = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^T (X_i - \bar{X}),
\]

(1)

where \( X_i \) is the \( i \)th training sample matrix with size \( h \times w \), \( \bar{X} \) is the mean matrix of all training samples, and \( N \) is the number of training samples. The main idea of 2DPCA is to select some projection vectors that maximize the total scatter of the projected samples, which can be denoted as

\[
J(w) = tr(S_w),
\]

(2)

where \( S_w \) is the covariance matrix of the projected vectors of training images, and \( tr(S_w) \) denotes the trace of \( S_w \). The covariance matrix \( S_w \) can be written as

\[
S_w = E[(y - E(y))(y - E(y))^T] = E[(X - E(X))w](X - E(X))^Tw).
\]

(3)

From (1), (2) and (3), we can get

\[
J(w) = w^TCw.
\]

(4)

Factually, the optimal projection vectors \( w_1, w_2, \ldots, w_d \) are the orthonormal eigenvectors of the image total covariance matrix \( C \) corresponding to the first \( d \) largest eigenvalues. Thus a feature matrix \( Y_i = [y_{i1}, y_{i2}, \ldots, y_{id}] \) for each training sample can be computed by \( y_{ik} = X_i w_k \), where \( k = 1, 2, \ldots, d \).

Also, we can get a feature matrix \( Y_i = [y_{i1}, y_{i2}, \ldots, y_{id}] \) for each testing sample. Then, the nearest neighbor classifier with some similarity measure can be used for classification.

2.2. Previous work on similarity measures

The state of the art similarity measures used in 2DPCA-based method consist of the following measures.

Frobenius distance \([3]\):

\[
d_F(Y_i, Y_j) = \left( \sum_{k=1}^{d} \sum_{l=1}^{m} (y_{ik} - y_{jl})^2 \right)^{1/2},
\]

(5)

Yang distance \([4]\):

\[
d_Y(Y_i, Y_j) = \sum_{k=1}^{d} \| y_{ik} - y_{jk} \|_2
\]

(6)

Volume measure \([13]\):

\[
d_V(Y_i, Y_j) = \sqrt{\det(Y_i - Y_j)^T(Y_i - Y_j)}
\]

(7)

These similarity measures mentioned above are fixed and can not adapt to the special data set. We will present a new similarity measure based on simulated annealing in section 3.

2.3. The simulated annealing algorithm

The simulated annealing algorithm is a heuristic search procedure for optimization problems. Based on the physical sciences, it was initially proposed by Metropolis in 1953 \([14]-[16]\). This method is essentially a random local search technique. The search starts from an initial feasible solution. Each solution has a specific cost value. A small change in one or a combination of some variables can generate a neighboring solution with a different cost value. In simulated annealing, the neighboring solution is generated randomly. If the cost value of the candidate solution is lower than that of the current solution, a move to the candidate solution is made. While if the candidate does not improve the current solution, the neighbor solution worse than the current solution is accepted according to a probability function. The acceptance probability is determined by a control parameter (temperature) which decreases during the SA procedure. When the parameter approach to zero, the method accepts the worse solution hardly. This make the SA has an ability to avoid becoming trapped at local minima and find a global optimal solution. The algorithm is described in Algorithm 1.
3. Using simulated annealing for learning a similarity measure

Assume a training face set containing $N$ face samples is given as $\{X_1, X_2, \cdots, X_N\}$, where $X_i$ is the $i$th training sample, which is a $m \times n$ matrix. The feature matrix for $X_i$ obtained by 2DPCA is denoted as $Y_i = [y_{i1}, y_{i2}, \cdots, y_{id}]^T$, where $y_{ik}, (k = 1, \cdots, d)$ is a $m$ dimensional vector.

Given the feature matrices $Y_i$ and $Y_j$ of face matrices $X_i$ and $X_j$, our aim is to learn a similarity function between them.

Generally speaking, $\|y_{ik} - y_{jk}\|_2$, $k = 1, \cdots, d$ carries some information for discriminating face $X_i$ and face $X_j$. Yang et al. uses the sum of all $\|y_{ik} - y_{jk}\|_2$, $k = 1, \cdots, d$ as the distance of $Y_i$ and $Y_j$, which is presented in (6). However, the contribution of each $\|y_{ik} - y_{jk}\|_2$, $k = 1, \cdots, d$ to the classification performance are different, so it is not an optimal choice to use the sum of all $\|y_{ik} - y_{jk}\|_2$, $k = 1, \cdots, d$ as the distance measure. We will learn an optimal similarity measure using the simulated annealing algorithm.

Given a group of weights $w_k, k = 1, \cdots, d$, we compute the following similarity measure.

$$d(Y_i, Y_j) = \sum_{k=1}^{d} w_k \|y_{ik} - y_{jk}\|_2 / \sum_{k=1}^{d} w_k.$$  (8)

In the simulated annealing algorithm, the training error of the nearest neighbor classifier using (8) as similarity measure is employed as the cost function. During the procedure of annealing, the training error is minimized. And a group of weights is learnt. We linearly combine $\|y_{ik} - y_{jk}\|_2$, $k = 1, \cdots, d$ according to the learnt weights into the optimal similarity measure. The algorithm is presented detailedly in Algorithm 2.

Algorithm 1. Simulated Annealing

Begin initialize $T_{\text{max}}, l_{\text{max}}, k_{\text{max}}, d(T), x_0, N(x)$

$T = T_{\text{max}}$, $x = x_0$

Do for $l = 1, \cdots, l_{\text{max}}$

Do for $k = 1, \cdots, k_{\text{max}}$

1. Select a candidate solution randomly from the neighborhood of $x$

$x' := N(x)$

2. Compute cost value of $x$ and $x'$, and compare them

$\Delta c = c(x') - c(x)$

3. If $\Delta c < 0$

then $x := x'$

else if $\exp(-\Delta c / T) > \text{rand}$

then $x := x'$

end if

end do

$T = T(l)$

end do

Output $x$

Algorithm 2. Simulated Annealing for Learning a Similarity Measure

Begin initialize $T_{\text{max}}, l_{\text{max}}, k_{\text{max}}, d(T), w_0, N(x)$,

$T = T_{\text{max}}, w = w_0$

where $w = (w_1, \cdots, w_d)^T$, and $w_0$ is the randomly selected $d$ dimensional initial feasible solution.

Do for $l = 1, \cdots, l_{\text{max}}$

Do for $k = 1, \cdots, k_{\text{max}}$

1. Select a candidate solution randomly from the neighborhood of $w$, we define the neighborhood as changing one random elements of $w$

$w' := N(w)$

2. Compute cost value of $w$ and $w'$, which is the training error of the nearest neighbor classifiers using (8) as similarity measure, and compare them

$\Delta c = c(w') - c(w)$

3. If $\Delta c < 0$

then $w := w'$

else if $\exp(-\Delta c / T) > \text{rand}$

then $w := w'$

end if

end do

$T = T(l)$

end do

Output $w$
We denote the learnt weight by simulated annealing as
\[ w^* = (w_1^*, \cdots, w_d^*)^T. \tag{9} \]

Using the learnt \( w^* \), the following similarity measure can be computed:
\[ d_{SA}(Y_i, Y_j) = \sum_{k=1}^{d} w_k^* \left\| y_{ik} - y_{jk} \right\|_2 / \sum_{k=1}^{d} w_k^*. \tag{10} \]

Given the testing feature matrix \( Y_i \) for each testing face sample and the feature matrix \( Y_j \) for each training face sample after the transformation by 2DPCA, a nearest neighbor classifier based on the similarity measure (10) can be depicted for the 2DPCA based face recognition as follows.

\[ c = \arg \min_i d_{SA}(Y_j, Y_i) \]
\[ = \arg \min_i \sum_{k=1}^{d} w_k^* \left\| y_{ik} - y_{jk} \right\|_2 / \sum_{k=1}^{d} w_k^*, \tag{11} \]

where \( c \in [1, 2, \cdots, N] \), and \( Y_i \) is classified into the class where \( Y_c \) belongs to.

4. Experiments

To evaluate the proposed algorithm, we use two well-known face databases, the ORL face database [17] and the YALE face database [18] for experiments.

4.1. Experimental results on the ORL face database

The ORL database contains 400 images of 40 distinct subjects, with 10 different images of each of these individuals. For some subject, the images were taken at different times, varying in different light conditions, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). The resolution of each image is 112×92 pixels. In our experiments, the first, third, fifth, seventh and ninth images of each person are used for training, while the remaining five images are used for testing. Fig. 1 shows some images used in the simulation.

We run four kinds of 2DPCA based methods (Frobenius distance, Yang distance, volume measure, SA similarity measure) on this data set, and Fig. 2 plots the recognition rates according to a different number of projection vectors when each method is applied. The number of projection vectors and recognition rates in the best case are shown in Table 1.

The table shows a fact that the maximal number of projection vector usually does not give a best recognition rate. We use the log temperature decreasing function [19] in our experiments, which is shown in Fig. 3. The parameters used in our experiments are shown in Table 2. How the value of cost function changes according to the number of iteration is shown in Fig. 4. The plot shows the simulated annealing algorithm can obtain the global optimal solution after about 600 iterations.

From Fig. 2, we can see that the performance of the 2DPCA based methods using SA similarity measure is better than the methods using Frobenius distance measure and Yang distance measure. When the number of projection vectors exceeds two, the 2DPCA based methods using boosted similarity measure provides better recognition rates than the
method using Volume measure.

The YALE face database includes 11 different images of each of 15 individuals, and the images vary in different light conditions and facial expressions. All images are grayscale and we cropped and normalized them to a resolution of 230×200 pixels in our experiments. Our experiments were performed using the first 6 image samples per subject for training, and the remaining images for testing. Fig. 5 shows the images of the first individual used in the simulation. The classification results are shown in Fig. 6. The number of projection vectors and recognition rates in the best case are shown in Table 3.

4.2. Experimental results on the YALE face database

The table shows the SA similarity measure outperforms others in the best case. From Fig. 6, we notice that the 2DPCA based methods using SA similarity measure gives better performance than the methods using Frobenius distance measure and Yang distance measure. When the number of projection vectors exceeds five, the 2DPCA based methods using SA similarity measure outperforms the method using Volume measure. Fig. 6 also indicates that generally the performance of the 2DPCA based method using Yang distance measure is better than that using Frobenius distance measure, and the Volume measure outperforms both of them.

Table 2. The parameters used in our experiments

<table>
<thead>
<tr>
<th>$T_{max}$</th>
<th>$l_{max}$</th>
<th>$k_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>1000</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3. The best recognition rates and the corresponding number of projection vectors on the Yale face database

<table>
<thead>
<tr>
<th>Measures</th>
<th>Number of Projection vectors</th>
<th>Recognition rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frobenius distance</td>
<td>4, 10</td>
<td>82.0</td>
</tr>
<tr>
<td>Yang distance</td>
<td>9, 10</td>
<td>90.0</td>
</tr>
<tr>
<td>Volume measure</td>
<td>4</td>
<td>94.0</td>
</tr>
<tr>
<td>SA similarity measure</td>
<td>9</td>
<td>96.0</td>
</tr>
</tbody>
</table>
5. Conclusion

In this paper we proposed a 2DPCA based method using a new similarity measure learnt by the simulated annealing algorithm for face recognition. The main area of novelty is that it is presented how the simulated annealing algorithm can be used for learning a similarity measure in 2DPCA based method for face recognition. Different from Frobenius distance measure, Yang distance measure and Volume measure, the SA is learnt on the training samples, which minimized the training error. Thus the similarity measure proposed in this paper adapts to the special face data set and can reject the influence of undesirable changes in a face pattern to some extent. The effectiveness of our method is demonstrated by simulation experiments, which shows that our method outperforms the state of the art methods in terms of recognition accuracy. However, this method still has drawbacks on the time complexity. It is an intriguing question to reduce the running time of the method. And the influence of the parameters in the simulated annealing is to be addressed in future.

REFERENCES