

Digital Information Facial Recognition Based on PCA and Its Improved Algorithm

Hai-feng Zhu
School of Electronics and Information
Nantong University
Nantong City
Jiangsu Province
China 226019
bauhauscg@163.com



ABSTRACT: Face recognition (PCA algorithm) based on principal component analysis method is a classic algorithm of digital information face recognition. This paper first analyzes the PCA algorithm and its basic principles. And in view of the shortage that it is affected by its subject facial expression changes and other factors, it designs the improved PCA algorithm, namely adding the image enhancement processing method, such as the local mean and standard deviation, to the original algorithm to increase the robustness of human face illumination and facial expression change during recognition. In addition, verify it on ORL face database and the results show that the improved PCA algorithm can get a higher resolution of the human face and up to 81.59% accuracy rate. The average growth of time-consuming is 0.378 seconds compared to traditional algorithms, and it is a fruitful improved algorithm.

Categories and Subject Descriptors:

I.2.10 [Vision and Scene Understanding]: Video Analysis;
I.4.10 [Image Representation]

General Terms:

Video Frame Processing, Content Processing

Keywords: Digital Information, Face Recognition, Principal Component Analysis, Image Enhancement Processing Method

Received: 1 June 2013, Revised 14 July 2013, Accepted 19 July 2013

1. Introduction

The application of face recognition system tends to direct facilitation, simple operation, efficient and accurate identification. It is a authentication method actually the

most widely used [1]. Face recognition based on principal component analysis method (PCA algorithm) is not only an important concept in the field of face recognition, but also a classic digital information face recognition algorithms.

In 1990, Sirovich and Kirby and so on used transform (PCA core idea) to achieve face reconstruction [2]; in 1991, PCA algorithm proposed by Turk and Pentland became truly a landmark work in facial recognition. However, the global features extracted by the traditional PCA algorithm is greatly influenced by the lighting conditions and facial expression changes and other factors, so the recognition effect is not very good.

At present, on the basis of the method, 2DPCA algorithm and the improved PCA algorithm of kernel-based PCA algorithm have emerged. Liu Chao (2011) uses dimensional PCA algorithm to extract the sample image's feature in the hybrid algorithm based on improved PCA face recognition and got the corresponding image feature matrix. And then he used the PCA algorithm and its dimensionality reduction to acquire dimension that reduced sample image feature data in order to improve effects of the recognition performance of face recognition system [3].

This paper on the basis of reference to the above results, digs up the the basic principles of PCA algorithm and makes a reliable processing for face digital signal to design an improved PCA algorithm. That is by the image enhancement processing method increasing the local mean and standard deviation to improve the robustness of light and face expression changes when facial recognition, so as to improve the accuracy and efficiency of face recognition.

2. The PCA algorithm

Principal component analysis (PCA) is a typical statistical analysis method. That is to parse out the main factors from the multivariate things, and project the high-dimensional data onto a lower dimensional space. Simplify complex issues can still reveal the essence of things.

This PCA algorithm face recognition technology involved in the paper takes the stability and individual differences of biometric as the basis, and through digital information processing to achieve real-time and accurate authentication purpose. PCA is based on the Karhunen-Loeve transform (the transformation). The conversion is essentially to establish a new coordinate system that take the spindle of an object along feature vector for the rotational transform. The transform relieves correlation between each component of the original data vectors. Then it is possible to remove those coordinates with fewer information as to minimize the dimension feature space.

The high-dimensional image space after transformation obtains a set of new orthogonal basis. Retain the significant orthogonal group, which can be converted into the low-dimensional linear space. If we assume that people face has separability in these low-dimensional linear space of projection, these projectors can be used as feature vector for identification, which is the basic idea of eigenface methods.

3. Face Recognition of PCA Algorithm

The application of complete PCA face recognition consists of several steps: face image preprocessing; reading face database, training forms subspace; project the training images and test images onto the subspace obtained in the previous step; select a function of distance to identify. The process is described in detail below:

3.1 Read face database

After normalizing face library, take a certain number of images every one selects from the library to form training set and the rest constitute the test set. Assumed that the normalized image is constituted by columns connected to form the dimensional vector. It can be regarded as a point in dimensional space, which can convert a low-dimensional sub-space to represent the image.

3.2 Calculate the changing generation matrix

Covariance matrix of all training samples (the following three equivalence) are:

$$C_A = \left(\sum_{k=1}^M x_k \cdot x_k^T \right) / M - m_x \cdot m_x^T \quad (1)$$

$$C_A = (A \cdot A^T) / M \quad (2)$$

$$C_A = \left[\sum_{i=1}^M (x_i - m_x)(x_i - m_x)^T \right] / M \quad (3)$$

$A = (\phi_1, \phi_2, \dots, \phi_M)$, $\phi_i = x_i - m_x$, m_x is the average of person's

face, M is the number of training the face. Covariance matrix C_A is a $N \times N$ matrix, N is the dimension of x_i .

In order to facilitate the calculation of eigenvalues and eigenvectors, generally adopt the second formula. According to $K - L$ transform theory, the new coordinate system is composed by the eigenvector components corresponding to nonzero eigenvalues in matrix $A \cdot A^T$. Directly seek the eigenvalues of the $N \times N$ matrix C_A and orthonormal eigenvectors, which is very difficult to get. According to the singular value decomposition principle, solve the eigenvalues and eigenvectors of $A^T \cdot A$ to obtain the eigenvalues and eigenvectors of $A \cdot A^T$.

After calculation, obtain all non-zero eigenvalues $[\lambda_0, \lambda_1, \dots, \lambda_{r-1}]$ of C_A (descending order, $1 \leq r \leq M$) and the corresponding unit orthogonal feature vectors $[u_0, u_1, \dots, u_{r-1}]$, and the feature space $U = [u_0, u_1, \dots, u_{r-1}] \in \mathcal{R}^{N \times r}$ can be obtained. So the projection coefficients of an image can be calculated in the feature space (it also can be understood as the coordinate of X in space of U):

$$Y = U^T * \mathcal{R}^{r \times 1}$$

(3) Identify

Take advantage of the formula $Y = U^T * \mathcal{R}^{r \times 1}$. At first, project all the training images, and then do the same for the test image. Use the discriminant function to identify the projector coefficients. Identification result of PCA algorithm is

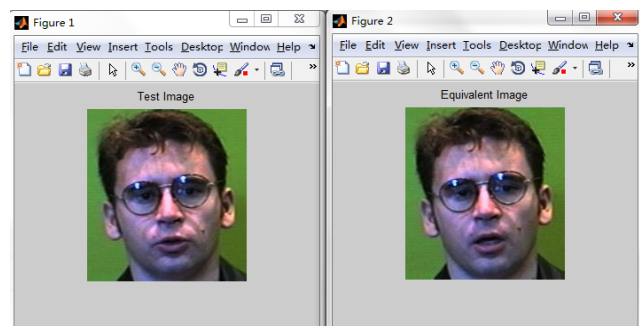


Figure 1. Identification Sample of PCA Algorithm

4. Analysis and Improvement of PCA Algorithm

4.1 Analysis

There are some problems in PCA algorithm. Such as:

A. From the mathematical foundations of view, $K - L$ transforms for feature extraction and compression at the premise of the minimum mean square error, which indicates that PCA algorithm is the best representation of feature extraction method. But if want to get good recognition results, the facial features extracted should have the best discrimination. And in the process of PCA feature extraction, do not use the class information, that means the method is unsupervised.

B. PCA algorithm conducts the linear transformation, which is based on the whole image to form a very large matrix. Although you can use indirect eigenvalue and eigenvector method to reduce the amount of computation, during the train or identification of sample, the high dimensional vector and matrix operations still occur frequently, which limits the application of this method in real-life.

C. The face image is two-dimensional distribution. the correlation between each pixel of the images has a close relationship with the distance between the pixels. But in the PCA method, the vector is used to express the image without using the distance information to weaken the ability of the human face of the method.

4.2 Improvement

The traditional PCA method extracts the global features, so it is affected by lighting conditions, the facial expression changes and other factors and the recognition effect is not very good. The traditional PCA algorithm fuses the image enhancement based on local mean and standard deviation. Before feature extraction, the improved PCA algorithm can effectively reduce the impact of uneven illumination on face recognition to expand the application conditions of PCA algorithm. Introduce image enhancement approach based on local mean and standard deviation in detail below.

Suppose an image and its gray level is $[0, L - 1]$. r represents the discrete random variables in the gray level of the image, $p(r)$ is the probability of occurrence of gray level. Then the global mean of the entire image can be expressed as:

$$E_g = \sum_{i=0}^{L-1} r_i p(r_i)$$

So the global contrast of the image, that the variance of the whole image is:

$$\sigma_g^2 = \sum_{i=0}^{L-1} (r_i - E_g)^2 p(r_i)$$

Since the brightness of an image can be measured by the mean of the image, and the contrast can measured by variance, compare the overall mean, local mean, the overall contrast and local contrast to enhance the dark areas with relatively low contrast in the image to be processed, which will not change the image area that has been relatively bright.

Assume a point $Q = (i, j)$ as the center of the image $M \times M$ to be processed and the neighborhood is $S_{(i,j)}$, so the mean of the neighborhood, that the local mean can be expressed as:

$$E_s = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^M x(i, j)$$

Here $x(i, j)$ is the gray-scale of the image to be processed. Local variance can be expressed as:

$$\sigma_s^2 = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M (x(i, j) - E_s)^2$$

The specific programs of image enhancement processing based on local mean and standard deviation are as follows:

(1) Make sure the darker areas in the image. If $E_s < k_0 E_g$, k_0 is a normal number that less than 1, it indicates that the region is darker area in the image, which needs to be further strengthened.

(2) Determine areas of low contrast in the image. If the contrast of an area in the image is too low, then the region can be identified without detail and does not need to be enhanced. Therefore, the region of low contrast to be enhanced in the image can be assumed: $k_1 \sigma_g < \sigma_s < k_2 \sigma_g$, $k_1 < k_2$, k_1 and k_2 are the normal number that less than 1.

(3) Make the gray amplification and the contrast stretching treatment of identified areas.

Based on the above scheme, the algorithm of image enhancement processing based on local mean and standard deviation can be expressed as:

$$f(i, j) = \begin{cases} \lambda x(i, j) + \beta x(i, j)^\gamma, & E_s < k_0, k_1 \sigma_g < \sigma_s < k_2 \sigma_g \\ x(i, j) & , other \end{cases}$$

Here k_0, k_1 and k_2 are the normal number that less than 1, E_s and σ_s are the local mean and standard deviation, E_g and σ_g are the global mean and standard deviation, λ as gray scale amplification factor, β and γ for contrast stretching coefficients. From the above equation, the algorithm is through local mean and standard values to determine the area need to be enhanced in the image (ie, low contrast and low gray area), and it would not make enhancement processing for the area that do not require enhancement.

Combining the PCA algorithm and image enhancement processing algorithm based on local mean and standard deviation can be very good to highlight the more important parts in the face image (such as eyes, nose, mouth, etc.). In the facial feature extraction process, it is easy to extract more discriminative facial features to improve recognition rate, which eliminates the effect of illumination factors on face recognition to a large extent.

5. Experimental data and results Analysis

The software Matlab r2012a used in the experiment of this paper adopts face database, which is used for the international generic ORL face database.

Matlab platform is one of the most widely used technology application platforms. It has a strong scientific computing capability, and it is the basic tool in fields of algorithm, application development and computer-aided design and analysis and others. It is also the first choice of platform in these areas.

ORL face database is produced by AT & T Labs of the British University, which includes 40 categories of faces. Each category of face has 10 images and the total is 400 images. Each piece of primitive face image contains 256 gray levels. The part of the face sample image in ORL database is shown in Figure 2:



Figure 2. The Part of Face Samples in ORL



Figure 3. ORL Average Face Obtained from Traditional PCA Algorithm



Figure 4. ORL Average Face Obtained from Improved PCA Algorithm



Figure 5. ORL featured Face Obtained from Traditional PCA Algorithm



Figure 6. ORL featured Face Obtained from Improved PCA Algorithm

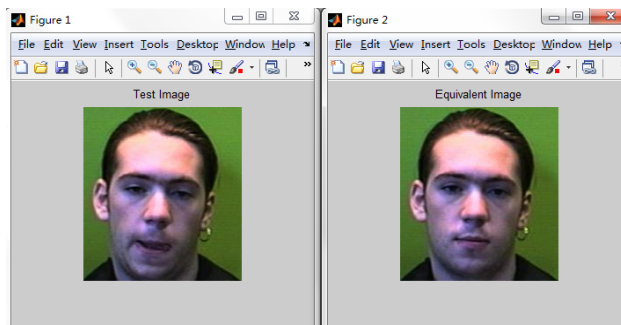


Figure 7. Identify Examples of Improved PCA Algorithm

Recognition time Number of training samples	5	6	7	8
Traditional PCA Algorithm	1.932	2.325	2.613	3.270
Improved PCA Algorithm	2.194	2.624	3.006	3.827

Table 1. Face Recognition Rates of Traditional PCA Algorithm and Improved PCA Algorithm in ORL

Recognition rate % Number of training samples	5	6	7	8
Traditional PCA Algorithm	70.21	72.57	75.68	75.35
Improved PCA Algorithm	72.30	73.18	77.26	81.59

Table 2. Face Recognition Time of Traditional PCA Algorithm and Improved PCA Algorithm in ORL

On the sharpness of the image, face obtained from the improved PCA algorithm has a big improvement. Compared the average faces in Figure 3 and 4, characterized face in Figure 5 and 6 in ORL face database comparison, the improved PCA algorithm get the clearer average face and featured face. From the data in Table 1, it can be seen that the improved PCA has a more obvious improvement than the traditional PCA algorithm on the recognition rate. However, from the data in Table 2, it gets that the improved PCA algorithm calculate relatively slowly on the speed of recognition. Figure 7 is the result that the improved PCA algorithm recognizes faces in the general gallery.

6. Conclusion

Face recognition technology has been widely used in business and everyday life, such as forensic detection, video surveillance, document authentication and access control. In the development of the face recognition technology, PCA algorithm is the first recognition method based on the overall characteristics or global visual feature. Its novel idea develops the research cogitation for face recognition technology and opens up its new field. PCA algorithm has a solid mathematical foundation as well as good operability, which is the new cornerstone of face recognition algorithm.

This paper conducts an in-depth research and analysis for the traditional PCA face recognition algorithm, and designs the improved PCA algorithm based on the algorithm. Namely combine the local mean and standard deviation of image enhancement processing in the PCA algorithm to increase the robustness of human face illumination and facial expressions changes during recognition. Compared with the traditional PCA algorithm, the improved PCA algorithm further expands application conditions of the PCA algorithm.

However, since PCA is a statistical method, training samples and species in face database the number will affect the performance of the identification system. And in some cases, it is impossible to have a sufficient number of species and training samples. Therefore, PCA algorithm application is also severely restricted.

Meanwhile, the improved PCA algorithm still has some issues to think and improve, such as:

(1) Start from complex information contained in the face organ, how to utilize structural information, grayscale information and geometry information contained in the face in the feature stage of the process of the face recognition to extract the facial features that has more power and facilitate the identification. Thereby it improves the accuracy and real-time of face recognition system.

(2) There are many other methods of face recognition in the field of face recognition. How to unit these identification methods together to play the advantages of each method, promotes the rapid development of face recognition technology.

(3) Experiments performed in this article are carried out on the face database. Through improving the preprocessing of algorithm, it extracts more effective recognition performance to improve the recognition distance function so as to extend the scope of algorithm.

Report of the main program :

1. Create Database function

```
function T = Create Database (Train Database Path)
Train Files = dir (Train Database Path); % Training set path
```

```
Train_Number = 0; % Initial number of training images
for i = 1 : size (TrainFiles,1) % Directory reserves except the
picture itself (.\.Thnmb.s.db)
If not (strcmp (Train Files (i) . name, '.') | strcmp (Train Files (i) .
name, '..'))
strcmp (TrainFiles(i).name, 'Thumbs.db'))
Train_Number = Train_Number + 1; % pictures of
statistical training set,
end
end
T = [];
for i = 1 : Train_Number % to every picture
str = strcat ('\ ', int2str (i), '.jpg'); % String concatenation,
get\i.jpg
str = strcat (Train Database Path, str); % String
concatenation to get the full picture of each one training path0
img = imread (str); % read into the training images
img = rgb2gray (img); % convert to gray scale
[irow icol] = size (img); % get the size of picture
temp = reshape (img, irow*icol, 1); % transform the two-
dimensional image into a one-dimensional vector
T = [T temp]; % take information of each image as the row of
T
end
2. Eigenface Core function
function [m, A, Eigenfaces] = Eigenface Core(T)
m = mean (T, 2); % average picture / line averaging (seek the
average of each pair of images corresponding pixel)  $m = (1/P) * \sum (T_j)$  ( $j = 1 : P$ )
Train_Number = size (T, 2); % the number of columns
% calculate each image to the variance of the picture mean A
= [];
for i = 1 : Train_Number % for each column
temp = double (T (:, i)) - m; % difference between each one
chart and the mean
A = [A temp]; % covariance matrix
end
% Dimensionality reduction
L = A' * A; % L is the covariance matrix  $C = A * A'$  transpose.
[V D] = eig (L);
L_eig_vec = []; % feature vector for  $i = 1 : \text{size}(V, 2)$  % for every
feature vector if ( $D(i, i) > 1$ ) % the feature value is greater
than 1
L_eig_vec = [L_eig_vec V (:, i)]; % concentrated
corresponding eigenvectors end
end
Eigenfaces = A * L_eig_vec;
Example function
clear all
clc
close all
Train Database Path = uigetdir ('D:\pca algorithm is used for
the face recognition \PCA_based Face Recognition System', ...
'Select training database path' );
Test Database Path = uigetdir ('D:\pca algorithm is used for
the face recognition \PCA_based Face Recognition System', ...
'Select test database path');
prompt = { 'Enter test image name (a number between 1 to 10)
:' };
dlg_title = 'Input of PCA-Based Face Recognition System';
```

```

num_lines= 1;
def = { '1' };
TestImage = inputdlg (prompt, dlg_title, num_lines, def);
TestImage = strcat (Test Database Path, '\', char (Test Image),
'.jpg');
im = imread (Test Image);
T = Create Database (Train Database Path);
[m, A, Eigenfaces] = Eigenface Core (T);
OutputName = Recognition (Test Image, m, A, Eigenfaces);
Selected Image = strcat (Train Database Path, '\', Output Name);
Selected Image = imread (Selected Image);
imshow (im)
title ('Test Image');
figure, imshow (Selected Image);
title ('Equivalent Image');
str = strcat ('Matched image is : ', OutputName);
disp (str)
4.Recognition function
function OutputName = Recognition (TestImage, m, A,
Eigenfaces)
Projected Images = []; % shine on the image
Train_Number = size (Eigenfaces,2);
for i = 1: Train_Number% for every training feature
temp = Eigenfaces'*A (:, i);
Projected Images = [Projected Images temp]; % to get
L_eig_vec;
end

```

```

Input Image = imread (Test Image); % read into the test image
temp = rgb2gray (Input Image); % get one of one-dimensional
to deal with
[irow icol] = size (temp); % test the size of the image
In Image = reshape (temp, irow*icol, 1); % transform to the
one-dimensional
Difference = double (In Image) - m; % L_eig_vec'
Projected Test Image = Eigenfaces'*Difference; % test the
feature vector of the image
Euc_dist = [];
for i = 1 : Train_Number % for every row
q = Projected Images(:, i); % bring out the training image
temp = ( norm( ProjectedTestImage - q) ) ^2; % Euclidean
distance
Euc_dist = [Euc_dist temp]; %
end
[Euc_dist_min, Recognized_index] = min (Euc_dist); % get an
index of image with the smallest difference
Output Name = strcat (int2str (Recognized_index), '.jpg'); %
get the name of the document

```

References

- [1] Xie Lixin, Mou Hui, Wang Huan, Liu Mingxia. (2009). Based on computer vision face detection and recognition summary. *The Third National Software Testing Conference and Mobile Computing, Grid, Intelligent Senior Forum Proceedings*.
- [2] Zhou Lin. (2011). Based on nonlinear partial least squares feature extraction method. *Nanjing University of Science Thesis*.

- [4] Wu Tiande, Dai Zaiping. (2011). Improved block 2DPCA Face Recognition. *Communication Technology*.
- [5] Tan Ziyong, Liang Jing. (2011). Based on PCA +2 DPCA face recognition method analysis. *Jishou University*.
- [6] Qi Xingmin, Liu Guanmei. (2007). The comparative study based on PCA face recognition method. *Modern Electronic Technology*.
- [3] Liu Chao. (2011). Based on the improved PCA face recognition hybrid algorithm. *Taiyuan University Graduate Thesis*.