Face Recognition by Combining Wavelet Transform and k-Nearest Neighbor*

Yugang Jiang, Ping Guo

(Department of Computer Science, Beijing Normal University, Beijing 100875, China)

Abstract: A novel technique for face recognition is presented in this paper. Wavelet transform and k-nearest neighbor rule are combined in this technique. Wavelet transform is adopted to obtain the low pass images of the original face images in order to reduce the impact of facial expressions. When the scales of the head of an individual in sample images are different, wavelet transform is employed to extract the position of facial features. Then a new method is adopted to construct feature vectors using the grey values of face images. Finally, k-nearest neighbor rule is employed to complete the recognition process. Experimental results show that the proposed technique is efficient at computation and robust to the facial expressions.

Key words: Face Recognition; k-nearest Neighbor Rule; Wavelet Transform; Principle Component Analysis

1. Introduction

Face recognition is a hot research topic in the fields of pattern recognition and computer vision, which can be directly used in many applications, such as verification of credit card, security access control, etc. Generally, face recognition methods can be divided into two main categories[1]. One is facial feature based approaches. The features such as eyes, nose and mouth are first located and then various feature extraction methods can be adopted to construct feature vectors of these facial features. Finally, traditional pattern recognition methods like a neural network can be employed to recognize the feature vectors. The other approach takes a holistic view of the recognition problem. Holistic feature extraction of face images is adopted in this approach. The most famous technique in the second approach is Principle Component Analysis (PCA)[2].

The results of face recognition can be easily affected by some variations in the face images, such as lighting, expression and pose. Wavelet transform provides a powerful and versatile framework for image processing. It is widely used in the fields of image de-noising, compression, fusion, etc. The changes of expressions in the sample images of an individual result in the differences of higher frequency band of the images[3]. That is, the lower frequency band is relatively stable to the higher band under different expressions. By using wavelet transform, the low pass image can be extracted. Moreover, the size of the low pass image can be reduced to 1/16 of the original size when 2-level wavelet decomposition is adopted[4], which can improve the computational efficiency of the feature extraction process. In this paper, we have adopted wavelet transform to obtain the low pass images of the original face images. When the scales of the head in sample images are different, wavelet transform is also employed to extract facial features.

For face images, it is very difficult to collect a large quantity of samples of one person, namely, the number of variables (pixels) is much higher than that of the samples. It is said that the data is severely ill-posed and neural network is not appropriate for this problem.

*Acknowledgements: This work is supported by the National Natural Science Foundation of China (No.60275002).

Yugang Jiang, male, M. S. postgraduate of Department of Computer Science; main research fields: pattern recognition and spectra analysis; E-mail: yugangjiang@sohu.com.

Ping Guo, male, IEEE senior member, professor; main research fields: neural network, image process, software reliability engineering, pattern recognition and spectra analysis; E-mail: pguo@ieee.org.
Therefore, \( k \)-Nearest Neighbor (\( k \)-NN) rule is adopted to complete the recognition process.

2. A Brief Review of Wavelet Transform and \( k \)-NN

Wavelet transform is a technique for analyzing signals. The basic functions of wavelet transform are obtained from a single prototype wavelet (or mother wavelet) \( \psi(t) \) by translation and dilation:

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)
\]

where \( a \) and \( b \) are both real numbers quantifying the scaling and translation operations. In this paper, we only consider discrete orthonormal wavelet transform, where the time-frequency plane is discretized according to a Logarithm sampling. By substitution of \( a=2^m \) and \( b=n\times2^m \), the basic functions become

\[
\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t-n)
\]

The one-dimensional discrete wavelet transform and reconstruction process of a signal \( x(t) \) is defined as:

\[
W(m,n) = \langle x(t),\psi_{m,n}(t) \rangle \quad (3)
\]

\[
x(t) = \sum_{m,n} W(m,n)\overline{\psi}_{m,n}(t) \quad (4)
\]

The two-dimensional wavelet transform is performed by applying one-dimensional wavelet transform to the rows and columns of two-dimensional data. The 1-level wavelet decomposition of an image results in an approximation image and three detail images in horizontal, vertical and diagonal directions respectively. The approximation image is used for the next level of decomposition, as shown in Fig. 1.

3. Experiments

3.1 Pre-processing

The face images used in the experiments are chosen from the FERET face database. It consists of 16 persons, with each person’s face appearing in 5 images, and comprises 80 images altogether. The images of an individual are obtained under different lighting, expression and pose. All of the images are 384×256 in size. Fig. 2 shows 5 images of an individual. In order to suppress the lighting’ impact on the recognition results, dynamic stretching is adopted to preprocess the images.

Dynamic stretching could suppress the lighting differences by making optimal use of the gray scale
available on the display. Changing the range of values in an image is accomplished by the redistribution of gray values of the input image histogram. Therefore, the output image histogram occupies a full gray level band. The transformation is given as follows:

\[
G' = \frac{G - G_{\text{min}}}{G_{\text{max}} - G_{\text{min}}} \times 255
\]  

(5)

where \(G_{\text{min}}\) and \(G_{\text{max}}\) are the minimal and the maximal gray level values in the histogram of the input image respectively.

### 3.2 Feature Extraction and Recognition

Firstly, 2-level wavelet decomposition is calculated for all of the images. Then the low pass images at the second level are extracted. According to Mallat algorithm, the size of the low pass image can be reduced to 1/16 of the original size. Here we adopt two methods to construct the feature vectors:

- Reduce the dimension of the low pass images using PCA;
- The low pass images are divided into several parts with \(m \times n\) in size. Then the mean of grey values in each part is calculated and the feature vectors are constructed using the mean values.

The bootstrap technique\(^6\) is applied in our experiments, and 3 images of an individual are selected as training samples randomly. The Correct Recognition Rate (CRR) is calculated using other samples. Each experiment is repeated 20 times and the mean and standard deviation (Std) of the CRR are reported. \(k\)-NN is employed to recognize the feature vectors obtained by the two feature extraction methods. The results are shown in Table 1.

Radial Basis Function (RBF) neural network has been widely used in the pattern recognition problems. In the experiments we adopt it to recognize the images. In order to obtain the input patterns of the neural network, PCA is employed to reduce the dimension of the feature vectors to 3. The sample images of an individual are divided into two groups: 3 for training the network and 2 for calculating the CRR. The CRR is shown in Table 1. Due to the limitation of the number of training samples, the recognition results are not satisfying.

<table>
<thead>
<tr>
<th>Table 1 Recognition Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition Method</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>CRR</td>
</tr>
<tr>
<td>Std</td>
</tr>
</tbody>
</table>

In Table 1, FE1 represents the first feature extraction method, FE2 represents the second one. From the above table, the CRR of 3-NN is the highest. Moreover, the standard deviation of the results obtained from 3-NN is the lowest, which indicates that the CRR of 3-NN is more stable than the others.

### 3.3 Extracting Facial Features

In Section 3.2, the feature vectors are directly extracted from the low pass images and the recognition results are very satisfying. However, when the scales of the head in sample images are different, as shown in Fig. 3, the recognition results will be aggravated. In this case, wavelet transform is adopted to extract the positions of facial features under this condition.

![Fig. 3 5 Images of an Individual with Different Scales of the Head](image)

Wavelet coefficients contain the horizontal and vertical information of images. The integral projection technique\(^7,8\) is adopted to extract the facial edges in wavelet domain. For an image \(I(m, n)\), where \(0 < m < M\), \(0 < n < N\), the horizontal integral projection \(H(n)\) and vertical integral projection \(V(m)\) are defined as:

\[
H(n) = \sum_{m=0}^{M} I(m, n), \quad 0 < n < N
\]  

(6)

\[
V(m) = \sum_{n=0}^{N} I(m, n), \quad 0 < m < M
\]  

(7)

The facial edges contain higher frequency than other areas do in an image, so the wavelet coefficients in these areas are larger than other ones. Through
calculating the vertical integral projection on the vertical detail image in 1-level wavelet decomposition and searching the two extreme local peaks, we can locate the left and right borders of the face, as shown in Fig. 4.

![Fig. 4 Locating Facial Edges](image)

Then the horizontal integral projection on the horizontal detail image is calculated. As shown in Fig. 4, there exist extreme local peaks near the positions of the top and bottom borders of the face. The top border of the image can be found by searching an extreme value in the top 1/4 part. Then the bottom border can be estimated by using the knowledge of common face shapes and the width between the left and right borders. Then the face images are cut out and resized to 48×48, as shown in Fig. 5.

![Fig. 5 Faces of an Individual after Being Cut out](image)

The dynamic stretching method is adopted to suppress the lighting differences. Then the methods introduced in section 3.2 are employed to construct the feature vectors. The recognition results using 3-NN are shown in Table 2.

<table>
<thead>
<tr>
<th>Recognition Method</th>
<th>CRR</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>FE1+3-NN</td>
<td>79%</td>
<td>0.0491</td>
</tr>
<tr>
<td>FE2+3-NN</td>
<td>84%</td>
<td>0.0342</td>
</tr>
</tbody>
</table>

The facial edges extracted by the integral projection technique are not very precise, but the obtained recognition results are acceptable for fast face recognition. The standard deviation of the results obtained from FE2+3-NN is low, which proves that it is better than the first method.

4. Conclusions

In this paper, we present a face recognition technique that combines wavelet transform and k-NN rule. Wavelet transform is adopted to extract the facial edges when the scales of the head in sample images are different. The CRR obtained by k-NN is better than that obtained by RBF neural network under small training sample conditions. This characteristic is very appropriate for face recognition where such conditions are typically encountered. The experimental results show that this technique is robust to facial expression and lighting, which demonstrates that it is a promising technique for face recognition.

References:

(Editors: Mei Huang, Nory, Ivan)