A New Face Recognition Method Based on Fast Least Squares Support Vector Machine

Rui Kong, Bing Zhang

College of Electrical and Information
Jinan University
Zhuhai, Guangdong, China,
tkongrui@jnu.edu.cn;tzhangbing@jnu.edu.cn

Abstract

In this paper we propose a new method for face recognition by combining Independent Component Analysis (ICA) and Support Vector Machine (SVM). Firstly we extract face features by using Informax algorithm. We then implement face recognition using FLS-SVM algorithm. We compare classical method of Principal Component Analysis (PCA) and our new method using ORL face database. The experiment results show the performance of our method is significantly superior to that of PCA-based method. The recognition speed of our method is faster than that of classical SVM algorithms.

Keywords: Fast Least Squares Support Vector Machine, Independent Component Analysis, Principal Component Analysis

1 Introduction

Face recognition is an important branch of Pattern recognition. Face recognition technology can be used in wide range of application such as identity authentication, access control, and surveillance.

The variations between the images of the same face are larger due to change of illumination, expression, and pose. The methods of features extracting and classify is very important to improve correct recognition ratio.

This paper proposes a new method of face recognition. The method combines the algorithms of ICA and SVM.

ICA[1] is a new technology of statistic signal processing since 1990s. ICA can extract the independent features of face images.
SVMs[2,3] have been recently proposed by Vapnik and his co-worker as a very effective method for general pattern classification. For example, given a set of sample points belonging to two classes, SVM can find the hyperplane that separates the largest possible fraction of sample points of the same class on the same side, while maximizing the margin between two classes.

In the paper, we adopt a Fast Least Squares SVM classification algorithm[4] for face recognition. The algorithm directly selects the sample points whose support values is bigger into next stage training to find classification hyperplane. So, training sample points is reduced, the speed of training is improve. To validate the effective of our method, ORL face images data is used for test. Meanwhile, we compare the correct ratio of our method and PCA method[5]. The experiment results show the correct ratio of our method is higher than that of PCA method, the classification speed of our method is faster than that of classic SVM.

2 Feature extracting of face images by ica

ICA is a technique for extracting statistically independent signals from a mixture signals. The basic idea of ICA is to represent a set of random variables using basis function, where the components are statistically independent as possible. In model of linear image combination, each face image is linearly combined from n basis image and independent factor codes. Suppose, each face image is denoted by \( x \) vector, basis image is denoted as \( A = (a_1, \cdots, a_n) \), independent factor codes is denoted by \( s \) vector, ICA linear model of a face image is as follows:

\[ x = As \]  

Composed and decomposed model of ICA face images is below:

![Fig.1 Composed and decomposed model of ICA face images](image)

By using ICA to extract the features of face images, as a matter of fact, is to find a group of filter matrix \( W \), linearly transform training collection of the mixture face images \( X = (x_1, \cdots, x_{200}) \), so as to the filtered face images as possible as independent:

\[ u = WX \]  

Here, \( u \) is a estimator of resource \( s \), the features extracted.

Each face image is represented as below, here \( u = (u_1, \cdots, u_n) \) are estimated independent factors
Rui Kong and Bing Zhang / Physics Procedia 22 (2011) 616 – 621
(features) each other. \( A = (a_1, \cdots, a_n) \) are basis image computed.

\[
\begin{align*}
\begin{array}{c}
\hspace{1cm} \hspace{1cm} \\
\hspace{1cm} \hspace{1cm} \\
\hspace{1cm} \hspace{1cm} \\
\end{array}
\end{align*}
\]

Fig.2 Each face image is linearly consist of basis images and factors

### 3 Fast Least Squares Svm Classification Algorithm

The classical SVM approach solves a convex quadratic programming problem\(^{[2,3]}\) to find the classification hyperplane. Transforming inequality type constraints instead of equality type constraints, J.A.K.Suykens and Vandewalle propose Least Squares SVM classification algorithm\(^{[6]}\). The algorithm finds optimization classification hyperplane by solving a set of linear equations instead of convex quadratic programming. But, Least Squares SVM classification algorithm loses sparsity of solution, if we can find those training sample points whose value are bigger, we can only use them for training, so the number of training sample points are reduced, the speed of algorithm can be improved. Based on reference \([6]\), we propose a new fast Least Squares SVM classification algorithm(FLS-SVM). Our algorithm directly select the training sample points for training whose value are bigger. So our algorithm can find optimization classification hyperplane rapidly. The experiment results show our algorithm not only insure SVM generalization performance, but also improve training speed.

Let data \((x_k, y_k)\) are independent uniformly distributed training points, here \(k = 1, \cdots, N, x \in \mathbb{R}^d, y \in \{+1, -1\}\) are labels.

Least Squares SVM Classification Algorithm solve the problem\(^{[4]}\) below:

\[
\min_{w,b,c} \left( \frac{1}{2} w^T w + \frac{1}{2} \sum_{k=1}^{N} e_k^2 \right)
\]

Subject to the equality constraints:

\[
y_k [w^T \phi(x_k) + b] = 1 - e_k, k = 1, \cdots, N
\]

One defines the Lagrangian: \( L(w, b, e, \alpha) = \left( \frac{1}{2} w^T w + \frac{1}{2} \sum_{k=1}^{N} e_k^2 \right) - \sum_{k=1}^{N} \alpha_k \left( y_k [w^T \phi(x_k) + b] - 1 + e_k \right) \)

Where \( \alpha_k \) are Lagrange multipliers (which can be either positive or negative now due to the equality constraints as follows from the Kuhn-Tucker conditions).

The conditions for optimality:

\[
\begin{align*}
\frac{\partial L}{\partial w} = 0 & \rightarrow w = \sum_{k=1}^{N} \alpha_k y_k \phi(x_k) \\
\frac{\partial L}{\partial b} = 0 & \rightarrow \sum_{k=1}^{N} \alpha_k y_k = 0 \\
\frac{\partial L}{\partial e_k} = 0 & \rightarrow \alpha_k = \gamma e_k, k = 1, \cdots, N \\
\frac{\partial L}{\partial \alpha_k} = 0 & \rightarrow y_k [w^T \phi(x_k) + b] - 1 + e_k = 0, k = 1, \cdots, N
\end{align*}
\]

Can be written immediately as the solution to the following set of linear equations:
\[
\begin{bmatrix}
I & 0 & 0 & -Z^T \\
0 & 0 & 0 & -Y^T \\
0 & 0 & \gamma I & -I \\
Z & Y & I & 0
\end{bmatrix}
\begin{bmatrix}
w \\
b \\
e \\
\alpha
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
0 \\
1
\end{bmatrix}
\]

(7)

where \( Z = [\phi(x_1)^T \phi(x_2)^T \cdots \phi(x_N)^T]^T \), \( Y = [y_1 \cdots y_N]^T \), \( \gamma = [1, \cdots, 1]^T \), \( e = [e_1, \cdots, e_N]^T \), \( \alpha = [\alpha_1, \cdots, \alpha_N] \)

The solution is also given by

\[
\begin{bmatrix}
0 & -Y^T \\
Y & \Omega + \gamma^{-1} I
\end{bmatrix}
\begin{bmatrix}
b \\
\alpha
\end{bmatrix} =
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\]

(8)

Mercer condition can be applied again to the matrix \( \Omega = ZZ^T \)

\[
\Omega_{ij} = y_i y_j \phi(x_i)^T \phi(x_j)
\]

(9)

Hence the classifier (10) is found by solving the linear set of equation (8) and (9) instead of quadratic programming.

\[
y(x) = \text{sign} \left[ \sum_{k=1}^{N} \alpha_k y_k k(x, x_k) + b \right]
\]

(10)

The solutions of LS-SVM have no sparsity, because every \( \alpha \) are non-zero. Comparing with the classic SVM algorithm, the algorithm of LS-SVM is simpler, but computational cost is higher.

In the paper, we use our algorithm of FLS-SVM for training, our algorithm directly select the training sample points for training whose value are bigger. So our algorithm can find optimization classification hyperplane rapidly.

Based on researching the algorithm of LS-SVM, we conclude that all solutions \( \alpha \) are non-zero, each training point’s value \( \alpha \) has different effect to optimization classification hyperplane. \( \alpha \) are bigger, its effect are important[4]. We also conclude value \( \alpha \) are determined the distance of the point to optimization classification hyperplane.

\[
d_k = \frac{(w^T x_k + b)}{\|w\|} = \frac{(1-e_k)}{y_k \|w\|}
\]

(11)

where, \( \alpha_k = \gamma e_k \),

\[
d_k = \frac{(w^T x_k + b)}{\|w\|} = \frac{(1-e_k)}{y_k \|w\|} = \frac{(1-\alpha_k/\gamma)}{y_k \|w\|}
\]

(12)

\[
\Rightarrow \alpha_k = \gamma (1-d_k y_k \|w\|)
\]

(13)

where \( y_k \) can only equal to \( \pm 1 \). \( \|w\| \) is approximately no change. After selecting \( \gamma \), when \( d_k \approx 1 \), \( \|\alpha_k\| \) are bigger; when \( d_k \approx 1 \), \( \|\alpha_k\| \) are bigger. when \( d_k \) are middle value, \( \|\alpha_k\| \) are small[4]. The rule can instruct us to select training points.

Based on analysis and research, the algorithm of FLS-SVM is proposed. The algorithm has two steps.

Firstly, selecting the training points whose value \( \alpha \) are bigger.

Secondly, training LS-SVM by using selecting the training points.

Detailed algorithm can consult reference [4].

4 Experiment

In this section, we validate the aforementioned face recognition algorithm using ORL face database. ORL face database include face images of 40 persons, everyone has ten face images (different expression, having glasses or non-glasses). We select randomly five images as training set, other as testing set. All the images are cropped to the size 28 by 28 and rectified according to the manually located eye positions. The
value of cropped images are normalization to [0,1]. The features of all face images are extracting by ICA technique. We use extended Infomax algorithm\(^7\) to extract the features. We can attain a set of basis images \(A\). Training sets and testing sets pass through \(W\) filtering which corresponding \(A\), we can acquire a set of features\(^1\).

The features of training sets: \(u_{train} = WX_{train}\) \hspace{1cm} (14)

The features of testing sets: \(u_{test} = WX_{test}\) \hspace{1cm} (15)

SVM classifications are training by FLS-SVM technique. We adopt two the kernel functions below:

(1) Polynomial kernel (PK): \(k(x, y) = (x \cdot y)^d\) \hspace{1cm} (16)

(2) Gaussian kernel (GK): \(k(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right)\) \hspace{1cm} (17)

We apply one-to-the-rest method to implement classification of multi-class. We also compare our method with that of PCA method. The experiment results validate the performance of our method is superior to that of PCA method.

Fig.3 parts of cropped face images

Fig.4 first 30 eigenfaces of PCA

Fig.5 first 30 basis images of ICA

<table>
<thead>
<tr>
<th>Method of extracting features</th>
<th>classifier</th>
<th>The number of principal component</th>
<th>Correct recognition ratio(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>FLS-SVM</td>
<td>PK ((d = 2))</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PK ((d = 3))</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GK ((\sigma = 0.7))</td>
<td>80</td>
</tr>
<tr>
<td>ICA</td>
<td>FLS-SVM</td>
<td>PK ((d = 2))</td>
<td>60</td>
</tr>
</tbody>
</table>
5 Conclusion and Discussing

In this paper, we propose a new face recognition algorithm which based on ICA and FLS-SVM technique. The comparative performance of the algorithm against the Eigenfaces method on ORL database is excellent. Using the same principal components, the performance of our algorithm is superior to that of PCA method. Because, the features of PCA only include the quadratic relation of pixels, the features of ICA can include the high-level relation of pixels. Many useful information are include in the high-level relation of pixels. The information is very important for recognition.

But, in extracting features, the compute cost of ICA is higher than that of PCA. In experiment, extracting features of PCA is completed in several seconds, extracting features of ICA need several minutes. Each face recognition consumes several seconds.

Acknowledgment

This research is partially sponsored by STFC of Guangdong under contract No. 2008B090500185 and No. 2009A030200002.

References