Face Recognition Method Using Improved Discriminant Sparse Preserving Projection

Di Wu, Kezheng Lin, Youhu Rong, Liangwei Zhuang

Department of Computer Science and Technology, Harbin University of Science and Technology, Harbin, 150080, China
Email: link@hrbust.edu.cn, wudi1989wudi@126.com

Abstract. According to the question that sparse preserving projection can’t make full use of local structure information of the similar and heterogeneous samples, proposed a face recognition method using improved discriminant sparse preserving projection. It used kernel methods to increase discriminant information sparse preserving coefficient and local structure information between the samples, updating weight of sparse representation coefficient with information, and optimizing the method as a whole. Results show that the method has good resolution, and it has higher recognition rate and robustness.

Keywords: sparse preserving projections; kernel methods; discriminant information; weight of sparse representation coefficient

1 Introduction

Manifold learning method is popular in face recognition. It includes Local Preserving Projections (LPP) [1], neighborhood Preserving Embedding (NPE) [2] and other methods. However, these methods [3] are not strong robustness under the influence of factors such as illumination, expression, posture. Based on Sparse Representation Classification (SRC) [4], someone put forward the Sparse Preserving Projection (SPP) [5], it get sparse reconstruction relationship of data by sparse representation and has better discriminant ability.

According to the above problems, this paper proposes a face recognition method using improved discriminant sparse preserving projection based on SPP and other algorithms [6]. It used kernel methods [7][8] to increase the discriminant information sparse preserving coefficient, updating homogeneous and heterogeneous sample weights when reconstructing coefficients, whole orthogonal constraint transformation to avoid the sample error approximation, and improve the recognition rate.

2 Sparse Preserving Projections

The SPP construct weighting matrix based on sparse representation. The set of training samples \( \{x_i\} \) and \( x_i \in \mathbb{R}^n \), \( X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{n \times n} \). We get \( x_i \) and
sparse reconstruction weight vector $s_i$ by minimizing the modified $l_1$ norm, due to the presence of noise and sample observations error, we need to replace constraints $x_i = Xs_i$ with $\|Xs - x_i\| \leq \varepsilon$ to get the following constrained optimization problem.

$$\min ||x_i||, \quad s.t. \quad ||Xs - x_i|| \leq \varepsilon, \quad 1 = e^T s_i$$

Among, $s_i = [s_{i,1}, ..., s_{i,j}, ..., s_{i,n}]^T$, $s_{i,j}$ ($j \neq i$) is the contribution that reconstruct $x_i$ with each $x_j$. To solve coefficients reconstruction matrix $S = [s_1, s_2, ..., s_n]^T$. $S$ reflects the inherent sparse relations characteristics between the data, and it contains natural discriminant information. objective function of sparse preserving projection can be expressed as:

$$\min \sum_{i=1}^{n} ||w^T x_i - w^T Xs_i||^2 \quad s.t. \quad w^T X X^T w = 1$$

Finally, using simple linear algebra plan, the optimization problem can be solved by the following generalized eigenvalue problem solving:

$$X S X^T w = \lambda X X^T w$$

Among, $S = S^T - S^T S$, the optimal projection matrix is the $d$ largest eigenvalues of $(XX^T)^{-1}(XS X^T)$ corresponding to the eigenvector.

### 3 Orthogonal Sparse Preserving Projections of Kernel

Mapping sample data to high dimensional space to make sparse representation coefficient contains more discriminant information by kernel methods. Training sample is $A = [A_1, ..., A_m], A_i = [a_{i,1}, a_{i,2}, ..., a_{i,n}] \in \mathbb{R}^{n \times 1}$, $A$ nonlinear mapped to $X = \{\phi(x_1), \phi(x_2), ..., \phi(x_m)\} = \{X_1, X_2, ..., X_m\}$ by kernel transformation $\phi$.

We optimize sparse reconstruction coefficients, strengthening the local linear characteristics between similar data by solving the least squares.

$$\min \|x_i - X t\|_2 \quad s.t. \quad It = 1$$

$t = [t_1, t_2, ..., t_i, 0, t_{i+1}, ..., t_n]^T$, we can get the optimal solution reconstruction weight of the similar sample $i$ by formula(1), $i \in \mathbb{R}^d$. While guarantee the similar reconstruction error minimization, we need to reduce the reconstruction weight of the heterogeneous sample. $\hat{h} = [O, ..., O, i, O, ..., O] \in \mathbb{R}^d$, assuming same sample sparse reconstruction residual $er = x_i - \hat{i}, \hat{i} = x_i - X \hat{h}$, sparse reconstruction in heterogeneous data sets $\hat{X} = \{X_1, X_2, ..., X_i, O, X_{i+1}, ..., X_m\}$. The linear constraint condition $I s = 0, s = s^+ - s^-$, and satisfy the following conditions.

$$s^+ = \{s, \quad s > 0\}, \quad s^- = \{-s, \quad s < 0\}$$

$$\|s\| = \sum_{i} |s_i| = \sum_{i} |s_i| = \sum_{i} (s_i^+ + s_i^-), \quad \text{so we can get as follow.}$$
\[
\min_{\hat{s}, \hat{d}} \sum_{i} (s_i^r + x_i) \quad s.t. \quad \|t_i - \hat{X}^r - s_i^r\|_2 < \delta
\]  

(3)

According to formula(3), \( \hat{s} = [\hat{s}_1, \ldots, \hat{s}_n, \hat{d}, \hat{\delta}_1, \ldots, \hat{\delta}_n] \) is the reconstruction weight of the heterogeneous samples, \( \hat{d} = \hat{s} + \hat{h} = [\hat{s}_1, \ldots, \hat{s}_n, \hat{\delta}_1, \ldots, \hat{\delta}_n] \). \( \tilde{t} = 1, \tilde{t} = 0 \), so \( \lambda \hat{d} = 1 \). Minimum error objective function is:

\[
\min_{\hat{s}} \sum_{i} \|v^T x_i - v^T \hat{X} \|_1^2, \quad s.t. \quad v^T \hat{X}^r v = 1
\]  

(4)

Minimization problem can be equivalent to maximization problem, we will replace the \( v^T \hat{X}^r v = 1 \) with \( v^T v = 1 \) to get optimal solution space.

\[
\max_{v^T} v^T \hat{X} \beta \hat{X}^T v \quad s.t. \quad v^T v = 1
\]

Among, \( \beta = s + s^T - s^T s \). Using the lagrange multiplier method to solve, that is eigenvectors corresponding to \( d \) maximum eigenvalue of symmetric matrices \( xS, x^T \). This improves the ability of sparse reconstruction.

4 Experimental Analysis

In order to detect KODSPE algorithm is effective and feasible, we do an experiment analysis on the ORL and YALE_B face database, using matlab2010 test platform, and the nearest neighbor classifier as the classifier.

First we analyze kernel parameters, as shown in Figure 1, selecting different kernel function parameters effect on the recognition rate on the ORL and YALE_B face. on the ORL face database, the optimal gaussian kernel parameter \( t = 5 \), the optimal polynomial kernel parameter \( d = 2 \). on the YALE_B face database, the optimal gaussian kernel parameter \( t = 2 \), the optimal polynomial kernel parameter \( d = 2 \). According to the optimal kernel parameters, we compare the recognition rate of different algorithms, when we choose different training samples, as shown in Figure 2.

---

**Figure 1.** Recognition rates with kernel parameter changes

**Figure 2.** Different algorithms recognition contrast figure on the ORL face database
Experiments show that the algorithm has good recognition rate than other algorithms on the OLR and YALE_B face database, specific data as shown in Table 1, selecting different kernel functions corresponding to recognition rate on two databases respectively.

Table 1. Different training sample recognition rate comparison on different database

<table>
<thead>
<tr>
<th>parameters</th>
<th>SPP</th>
<th>KSPP</th>
<th>KODSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d=2</td>
<td>91.2</td>
<td>92.6</td>
<td>94.8</td>
</tr>
<tr>
<td>t=5</td>
<td>91.2</td>
<td>92.6</td>
<td>93.6</td>
</tr>
<tr>
<td>YALE_B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d=2</td>
<td>92.7</td>
<td>94.3</td>
<td>96.5</td>
</tr>
<tr>
<td>t=2</td>
<td>92.7</td>
<td>94.3</td>
<td>95.7</td>
</tr>
</tbody>
</table>

5 Conclusion

This paper proposed a face recognition method using improved discriminant sparse preserving projection. The algorithm get more discriminant information from high dimensional space, it can classify face more better, and update the coefficient of sparse representation to make similar face sample occupy more weight for reducing the error of feature extraction, finally improving overall information retention capacity by orthogonal transformation. Through experiments on multiple face library, the results show that the improved method has good stability and recognition rate.

References