Extended CRC: Face Recognition with a Single Training Image per Person via Intraclass Variant Dictionary

Guojun LIN†a, Mei XIE†, Nonmembers, and Ling MAO†, Member

SUMMARY For face recognition with a single training image per person, Collaborative Representation based Classification (CRC) has significantly less complexity than Extended Sparse Representation based Classification (ESRC). However, CRC gets lower recognition rates than ESRC. In order to combine the advantages of CRC and ESRC, we propose Extended Collaborative Representation based Classification (ECRC) for face recognition with a single training image per person. ECRC constructs an auxiliary intraclass variant dictionary to represent the possible variation between the testing and training images. Experimental results show that ECRC outperforms the compared methods in terms of both high recognition rates and low computation complexity.

key words: face recognition, sparse representation, collaborative representation, single training image

1. Introduction

In the past two decades, face recognition has become one of the most potential recognition techniques of the biometric characteristic for the merits of being natural, directly perceived, safe and convenient. Many face recognition methods, such as Eigenfaces, Fisherfaces and Laplacianfaces, have been proposed. Recently, Wright et al. proposed Sparse Representation based Classification (SRC) for face recognition with a single training image per person. SRC can achieve good recognition performance, but requires sufficient training images per person. However, in many specific scenarios, such as law enforcement, passport identification and homeland security, there is only a single training image per person that can be used. Face recognition with a single training image per person has become one of the challenges in the real applications. Deng et al. proposed Extended Sparse Representation based Classification (ESRC) [3] for face recognition in the case of insufficient or even a single training images per person. ESRC can achieves higher recognition rates than SRC but with the high computation complexity. Zhang et al. observed that it was collaborative representation rather than $l_1$-norm sparsity that made SRC powerful for face recognition, and proposed Collaborative Representation based Classification (CRC) [4, 5] for face recognition with sufficient training images per person. CRC has significantly less complexity than SRC and achieves the similar recognition rate. For face recognition with a single training image per person, CRC has significantly less complexity than ESRC. However, CRC gets lower recognition rates than ESRC.

In order to combine the advantages of CRC and ESRC, we propose Extended Collaborative Representation based Classification (ECRC) for face recognition with a single training image per person. By using the observation that the intraclass variability, caused by variable expressions, illuminations and disguises, can be shared across different subjects, ECRC constructs an intraclass variant dictionary to represent the variation between the testing and training images. The testing image can be approximated as a linear combination of all the training images and the intraclass variant bases, and the large coefficients are concentrated on the training image with the same class as the testing image and on the related intraclass variant bases. Experimental results on the AR and Extended Yale B databases show that ECRC outperforms the compared methods and achieves both high recognition rates and low computation complexity.

The rest of the paper is organized as follows. Section 2 reviews CRC algorithm. Section 3 presents in detail the ECRC algorithm. Section 4 presents the experimental results. Section 5 concludes the paper.

2. Review of CRC

We suppose that a face database has $k$ classes of training samples, which is denoted by a matrix $A = [A_1, A_2, \ldots, A_k] \in R^{mxm}$. Each column of $A_i \in R^{mxn}$ is a training sample of class $i$. Then, the linear representation of a testing sample $y$ can be written in terms of all training samples

$$y = Ax_0 + z \ (1)$$

where $x_0$ is a coefficient vector whose entries are nearly zero except those associated with the $i$th class, and $z \in R^m$ is a noise term with bounded energy $\|z\|_2 < \varepsilon$. In order to collaboratively represent the testing sample $y \in R^m$ over $A$ with low computational burden, the regularized least square method is used and can be represented as:

$$\hat{x} = \arg \min_x \{\|y - A \cdot x\|_2^2 + \lambda \|x\|_2^2\} \ (2)$$

where $\lambda$ is the regularization parameter, $\hat{x}$ is a coefficient vector and the solution. The solution of Eq. (2) can be analytically derived as:

$$\hat{x} = (A^T A + \lambda \cdot I)^{-1} A^T y \ (3)$$
3. Extended Collaborative Representation Based Classification

We suppose that there is a gallery face set with a single nature image per subject. The testing images in the real world environment may contain complex variations, such as expressions, illuminations and disguises. Because the testing image deviates largely from the same class gallery image, the dominant coefficients of \( \hat{x} \) in Eq. (2) cannot concentrate on the correct gallery image. CRC gets low recognition rates. In [3], Deng et al. assumed that the intraclass variation of any gallery face can be approximated by a sparse linear combination of the intraclass differences from sufficient number of generic faces. Based on the assumption, the large deviation between the testing image and the same class gallery image can be linearly approximated by the intraclass differences of generic subjects. This intuition inspires us to propose ECRC for face recognition with a single training image per subject.

Suppose a basis matrix \( D \) represent the universal intraclass variant bases. If the gallery face data matrix \( A \) is the matrix with a single natural image per subject, Eq. (1) can be modified to account for large variation between the testing and training images and rewritten as follows:

\[
y = Ax_0 + Db_0 + z
\]  

where the intraclass variant matrix \( D \) usually represents unbalanced lighting changes, exaggerated expressions and occlusions. \( b_0 \) is a coefficient vector for \( D \). Eq. (2) can be modified as follows:

\[
\begin{bmatrix}
\hat{x} \\
\hat{\beta}
\end{bmatrix}
= \arg \min_{x, \beta} \left\{ \| y - [A, D] \begin{bmatrix} x \\ \beta \end{bmatrix} \|_2^2 + \lambda \left( \| x \|_2^2 + \| \beta \|_2^2 \right) \right\}
\]

The solution of Eq. (8) can be derived as:

\[
\begin{bmatrix}
\hat{x} \\
\hat{\beta}
\end{bmatrix}
= \arg \min_{x, \beta} \left\{ \| y - [A, D] \begin{bmatrix} x \\ \beta \end{bmatrix} \|_2^2 + \lambda \left( \| x \|_2^2 + \| \beta \|_2^2 \right) \right\}
\]

Let \( P = (A^T A + \lambda I)^{-1} A^T \). We can see that \( P \) is independent of \( y \) and can be precalculated as a projection matrix. Given a testing sample \( y \), we can just simply project \( y \) onto \( P \) via \( Py \) to get \( \hat{x} \). We can see that the matrix \( P \) is very important for CRC, because \( P \) makes CRC fast. The CRC algorithm is shown in Table 1.

### Table 1 The CRC algorithm.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Normalize the columns of ( A ) to have unit ( l_2)-norm.</td>
</tr>
<tr>
<td>2.</td>
<td>Code ( y ) over ( A ) by</td>
</tr>
<tr>
<td></td>
<td>( \hat{x} = Py )</td>
</tr>
<tr>
<td>3.</td>
<td>Compute the regularized residuals</td>
</tr>
<tr>
<td></td>
<td>( r_i = \frac{| y - A_i \cdot \hat{x}_i |_2}{| \hat{x}_i |_2} )</td>
</tr>
<tr>
<td>4.</td>
<td>Output the identity of ( y ) as</td>
</tr>
<tr>
<td></td>
<td>( \text{identity}(y) = \arg \min_i (r_i) )</td>
</tr>
</tbody>
</table>

### Table 2 The ECRC algorithm.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Input: a matrix of training samples ( A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times k} ) for ( k ) classes, a matrix of intraclass variant bases ( D \in \mathbb{R}^{m \times p} ), a testing sample ( y \in \mathbb{R}^m ).</td>
</tr>
<tr>
<td>2.</td>
<td>Normalize the columns of ( A ) and ( D ) to have unit ( l_2)-norm.</td>
</tr>
<tr>
<td>3.</td>
<td>Code ( y ) over ( [A, D] ) by</td>
</tr>
<tr>
<td></td>
<td>( \begin{bmatrix} \hat{x} \ \hat{\beta} \end{bmatrix} = Qy )</td>
</tr>
<tr>
<td>4.</td>
<td>Compute the regularized residuals</td>
</tr>
<tr>
<td></td>
<td>( r_i = \frac{| y - A_i \cdot \hat{x}_i |_2}{| \hat{x}_i |_2} )</td>
</tr>
<tr>
<td>5.</td>
<td>Output the identity of ( y ) as</td>
</tr>
<tr>
<td></td>
<td>( \text{identity}(y) = \arg \min_i (r_i) )</td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
\hat{x} \\
\hat{\beta}
\end{bmatrix}
= (\begin{bmatrix} A, D \end{bmatrix}^T [A, D] + \lambda \cdot I)^{-1} [A, D]^T y
\]

where \( \hat{x} \) is a coefficient vector for \( A \), \( \hat{\beta} \) is a coefficient vector for \( D \). Let \( Q = ([A, D]^T [A, D] + \lambda \cdot I)^{-1} [A, D]^T \). Given a generic face set with multiple images per subject, there is a natural image for each subject. The face set is denoted by a matrix \([D_1, D_2, \ldots, D_l] \in \mathbb{R}^{m \times l}\), where the submatrix \( D_i \in \mathbb{R}^{m \times i} \) is the samples of class \( i \), \( i = 1, 2, \ldots, l \). The bases of dictionary \( D \) can be got by subtracting the natural image from the other images of the same class. The intraclass variant dictionary \( D \) can be calculated as follows:

\[
D = [D_1 - a_1e_1, D_2 - a_2e_2, \ldots, D_l - a_le_l] \in \mathbb{R}^{m \times (l-1)}
\]

where \( e_i = [1, 1, \ldots, 1] \in \mathbb{R}^{1 \times (l-1)} \), \( a_i \) is the natural sample of class \( i \), and \( D_i \) is the reduced matrix of class \( i \) which removes the natural sample. The proposed ECRC algorithm is summarized in Table 2.

Based on Eq. (7), the testing sample can be represented by a linear combination of the training samples and the intraclass variant bases. The large coefficients are concentrated on the training sample with the same class as the testing sample and on the related intraclass variant bases.

In order to illustrate the principle of ECRC work, Figure 1 shows the coefficients of the testing image with sunglasses on the training images and the intraclass variant bases using ECRC. The coefficients of \( \hat{x} \) are the former 80 coefficients which correspond to 80 gallery training images with a single image per subject, and the coefficients of \( \hat{\beta} \) are the latter 240 coefficients which correspond to the intraclass variant bases. Several intraclass variant bases marked by red lines represent the intraclass variant bases with sunglasses. From Fig. 1, we can see that the testing image with sunglasses is approximated by the linear combination of the gallery training image of the same class and several intraclass variant bases related to sunglasses. The large coefficients of the testing image are concentrated on the gallery training image of the same class and several intraclass variant bases with sunglasses. The smallest residual corre-
sponds to the correct subject. Besides the disguise case, this schemes have been worked on the testing images with variable illuminations and expressions.

4. Experimental Results

In this section, ECRC, CRC and SRC with fast \( l_1 \) minimization methods, including L1LS [6], FISTA [7] and Homotopy [8] are compared on the AR and Extended Yale B databases. The parameter \( \lambda \) is set as 0.05. We also compare the running time of ECRC and ESRC with fast \( l_1 \) minimization methods, including L1LS, FISTA and Homotopy on the AR and Extended Yale B databases.

4.1 AR Database

In the experiment, we randomly chose a subset consisting of 80 subjects from session 1 of the AR database [9]. For each subject, the single natural image is used for training, and the other 12 images with illumination change, expressions and facial disguises are used for testing. The images of size \( 165 \times 120 \) are converted to gray scale and stacked to be 19800 dimensional vectors. Figure 2 shows the 13 images of one person of the AR database. In order to construct the intraclass variant dictionary for ECRC, another 20 subjects different from the 80 testing subjects are selected, each subject has 13 images. The intraclass variant dictionary consists of 240 bases, which are computed by Eq. (10). Table 3 shows the recognition rates of the proposed ECRC and the compared methods on the AR database. We can see that the recognition rate of ECRC is higher than those of the compared methods. ECRC achieves a recognition rate of 91.67%, which is about 37% higher than those of CRC and three SRC methods.

4.2 Extended Yale B Database

The Extended Yale B database consists of 2414 frontal face images of 38 individuals [10]. The cropped and normalized \( 192 \times 168 \) face images were taken under varying illumination conditions [11]. In the experiment, the images are resized to \( 96 \times 84 \) pixels and stacked to be 8064 dimensional vectors. We randomly choose 32 subjects from the database. Figure 3 shows the 13 images of one person of the Extended Yale B database. For each subject, the single natural image is selected from subset 1 for training, 12 images with illumination change of subset 3 is used for testing. In order to construct the intraclass variant dictionary, the remainder 6 subjects different from the 32 testing subjects are selected, each subject has 13 images. The intraclass variant dictionary consists of 72 bases, which are computed by Eq. (10). Table 4 shows the recognition rates of the proposed ECRC and the compared methods on the Extended Yale B database. We can see that ECRC achieves the best performance. The recognition rate of ECRC is 36.3% higher than that of CRC and 42.6% higher than that of SRC(L1LS). The recognition rate of ECRC is about 45% higher than those of
Table 5 Recognition rate and speed of different methods on the AR database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition rate (%)</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESRC(L1LS)</td>
<td>84.79</td>
<td>8.4272</td>
</tr>
<tr>
<td>ESRC(FISTA)</td>
<td>78.96</td>
<td>8.0376</td>
</tr>
<tr>
<td>ESRC(Homotopy)</td>
<td>84.79</td>
<td>0.4749</td>
</tr>
<tr>
<td>ECRC</td>
<td>91.67</td>
<td>0.0387</td>
</tr>
</tbody>
</table>

Table 6 Recognition rate and speed of different methods on the Extended Yale B database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition rate (%)</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESRC(L1LS)</td>
<td>95.30</td>
<td>8.5206</td>
</tr>
<tr>
<td>ESRC(FISTA)</td>
<td>90.34</td>
<td>8.0074</td>
</tr>
<tr>
<td>ESRC(Homotopy)</td>
<td>96.61</td>
<td>0.1036</td>
</tr>
<tr>
<td>ECRC</td>
<td>97.39</td>
<td>0.0059</td>
</tr>
</tbody>
</table>

SRC(FISTA) and SRC(Homotopy).

4.3 Running Time

We compare the running time of ECRC and ESRC with various fast $\ell_1$ minimization methods, including L1LS, FISTA and Homotopy. All the experiments are implemented using Matlab on a PC with Dual Core 3.0 GHz Pentium CPU and 2 GB RAM. Table 5 shows the recognition rate and speed of different methods on the AR database, and Table 6 shows the recognition rate and speed of different methods on the Extended Yale B database.

In the ECRC algorithm, $Q$ can be precalculated as a projection matrix. A testing sample $y$ is directly projected onto $Q$ via $Qy$ to get $\hat{x}$ fast for face recognition. ECRC is operated fast and has low computation complexity. On the contrary, ESRC hasn’t the precalculated projection matrix and requires operating $l_1$ minimization to get $\hat{x}$ for face recognition, while $l_1$ minimization is operated slowly and has high computation complexity. So ESRC is operated slowly and has high computation complexity.

On the AR database, the proposed ECRC achieves the best recognition rate and the fastest speed. The recognition rate of ECRC is about 7% higher than those of ESRC(L1LS) and ESRC(Homotopy), the speed of ECRC is 217.8 times faster than that of ESRC(L1LS) and 12.3 times faster than that of ESRC(Homotopy). The recognition rate of ECRC is 12.7% higher than that of ESRC(FISTA), the speed of ECRC is 207.7 times faster than that of it. On the Extended Yale B database, ECRC has the best recognition rate of 97.39%, the speed of ECRC is 17.6 times faster than that of ESRC(Homotopy), especially, 1444.2 times faster than that of ESRC(L1LS) and 1357.2 times faster than that of ESRC(FISTA).

5. Conclusions

In this paper, we proposed Extended Collaborative Representation based Classification (ECRC) for face recognition with a single training image per person. ECRC combines well the advantages of CRC and ESRC. Experimental results on the AR and Extended Yale B databases show that ECRC outperforms the compared methods and achieves both high recognition rates and low computation complexity. As a result, ECRC is a better method for face recognition.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No.61172117 and Science and Technology Planning Project of Guangdong Province under Grant No.2011B031600004.

References