A Novel Approach to Improve the Collaborative Representation for Face Recognition

YuJie Jia¹, JiaLu Deng¹, Xu Zhang², Qiong Wang² and Shi Yan¹,*

¹ School of Information Science and Engineering, Lanzhou University, Lanzhou, 730000, China
² State Grid Gansu Information & Telecommunication Company, Lanzhou, 730000, China

E-mail: jiayj17@lzu.edu.cn; dengjl17@lzu.edu.cn; zhangxu@gs.sgcc.con.cn; wangqiong-xt@gs.sgcc.con.cn; m51@163.com*

Abstract. Collaborative representation based classification (CRC) codes a test sample as a linear combination of all the training samples. However, the recognition rate of CRC is not ideal when available training samples are insufficient, as the test sample cannot be accurately represented by limited training samples. In order to address this issue, we propose a novel idea of training samples fragmentation. First, all training samples are divided into two new training sample sets according to the similarity among them. Next, the test sample simultaneously uses the two new training sample sets to perform CRC, which ultimately uses $2M$ "nearest neighbors" from the two training sample sets to represent and classify the test sample. In addition, this method also takes advantages of a new fusion classification mechanism based on histogram similarity and Euclidean distance, which has been proven to perform better than Euclidean distance classification. The experimental results reveal that the proposed method performs better in face recognition compared with the most representation based classification methods.

1. Introduction

Face Recognition (FR) is a challenging but interesting question that attracts researchers from different fields and various techniques have been proposed to deal with different issues, such as illumination, pose, occlusion and small sample size [1]. A typical face recognition method involves two steps: feature extraction and classification. Since Wright et al. [2] proposed a face recognition method based on sparse representation based classification (SRC), SRC and its variants [3,4] have been widely used in face recognition. In the SRC, a test sample is linearly reconstructed from all training samples and classified based on the reconstruction errors associated with each class. Unlike traditional classification methods, SRC does not require feature extraction.

Recent investigations have shown that CRC [5] has achieved excellent results in face recognition by using l2 sparse constraint in the optimization problem. In the practical face recognition system, the number of available training samples is generally limited such that face recognition becomes a small sample size (SSS) problem. In order to solve this, Xu et al. [6] used symmetry of the image to generate two symmetrical virtual facial images to increase the number of training samples and then the recognition rate was greatly increased. Similarly, Tang et al. [7] used an optic flow and expression ratio image to generate ‘virtual’ facial expression. In addition to generating virtual samples, many
researchers are committed to improving the CRC. For example, Dong et al. [8] proposed a two-stage learning algorithm, which mainly used unlabeled samples to reconstruct marker samples to obtain collaborative representation coefficients. Tian et al. [9] proposed a new fusion mechanism, which combines CRC and SRC and has the advantages of both l2 sparse constraint and l1 sparse constraint in the optimization problem.

In this paper, we propose a novel approach to improve the collaborative representation for face recognition. All training samples are firstly divided into two new training sample sets according to the similarity among them, and then the test sample simultaneously uses the two new training sample sets to perform CRC, which ultimately uses $2M$ "nearest neighbors" from the two training sample sets to represent and classify the test sample. The aim of our method is to reduce the interference of these training samples that are not same class as the test sample. Beside, since the Euclidean distance treats different attributes of samples equally, which leads to the same contribution of the residuals in different positions of two images. This sometimes fails to meet the actual requirements. Therefore, a new fusion classification mechanism based on histogram similarity and Euclidean distance is used to judge the residual. Experimental results show that the proposed method performs well in face recognition compared with the most representation-based methods.

2. The proposed method

Assume that there are $c$ classes in all training samples, all training samples are $X=[X_1,X_2,...,X_i]$, where $X_i \in R^{n \times m}$, $n$ is the number of $i$ class training samples.

The purpose of using the inter-class scattering is to extract the most discriminative low-dimensional features from the high-dimensional feature space, which can help different classes of samples to be separated as much as possible. The inter-class dispersion matrix of the training samples is calculated by

$$S_i = \sum_{j=1}^{c_i}(X_j - \bar{X})(X_j - \bar{X})^T,$$

where $\bar{X}$ is the average face of all training samples and $X_j$ is the average face of the $i$-class training sample. Then the classes of the training samples are sorted according to the size of the discrete values. Finally, all training samples are cross-divided into two new training sample sets.

The training samples $X$ are divided into two new training sample sets, i.e., $X = X_1 \cup X_2$ ($c = c_i + c_s$, $X_1 = [x_{i1}, x_{i2}, ..., x_{iocc_1}]$, and $X_2 = [x_{s1}, x_{s2}, ..., x_{soc_2}]$). Given the test sample $y$ and two new training sets $X_1$ and $X_2$, the model of the CRC can be expressed as

$$\alpha_i = \arg \min_{\alpha_i} \left( \|y - X_i \alpha_i \|^2_2 + \lambda \|\alpha_i\|^2_1 \right)$$

and

$$\alpha_s = \arg \min_{\alpha_s} \left( \|y - X_s \alpha_s \|^2_2 + \lambda \|\alpha_s\|^2_1 \right),$$

where $\lambda$ is a regularization parameter, and $\alpha_i$ and $\alpha_s$ are collaborative representation coefficients that linearly represent test sample $y$ by $X_i$ and $X_s$.

The role of the regularization term is twofold: it stabilizes the solution of the least squares method and introduces some sparsity. The collaborative representation based on the regularized least squares solution can be derived as

$$\alpha_i = P_i y$$

and

$$\alpha_s = P_s y,$$

where $P_i = (X_i^T X_i + \lambda I)^{-1} X_i^T$, $P_s = (X_s^T X_s + \lambda I)^{-1} X_s^T$, $\alpha_i = [a_{i1}, a_{i2}, ..., a_{iocc_1}]$ and $\alpha_s = [a_{s1}, a_{s2}, ..., a_{soc_2}]$.

For the residuals of the test sample and the reconstructed test samples from the two new training sample sets, a new fusion mechanism is proposed by
is the coefficient. The test sample is finally assigned to the training sample and \( \text{class} \), and \( \text{class} \), which is linearly represented by the new training set 

\[
d_s = \|y - \sum_{j=1}^{k} a_j x_j\| + \sum_{j=1}^{k} \sqrt{pp_j} \tag{6}
\]

and

\[
d_s = \|y - \sum_{j=1}^{m} a_j x_j\| + \sum_{j=1}^{m} \sqrt{pp_j} \tag{7}
\]

where \( p \), \( p_j \), and \( p_i \) represent histogram data of the test sample \( y \), \( q \)-th training sample and \( k \)-th training sample, respectively, and the range is between \([0, 1]\). The closer value of the histogram similarity is to 0, the more similar the two images are.

The two residual sets are arranged in ascending order, and each of the first \( M \) residuals is taken. The class of \( 2M \) residuals is denoted as \( r_i \)-th, \( r_2 \)-th, \ldots, \( r_{2M} \)-th, and are recomposed into a new training sample set \( X' \). The test sample \( y \) is linearly represented by the new training set \( X' \)

\[
y = f_1 x_1 + \ldots + f_{2M} x_{2M} \tag{8}
\]

where \( f_i \) is the coefficient. The equation (8) is rewritten as

\[
y = X' F \tag{9}
\]

where \( F \) is calculated by \( F = (X' X + \gamma I)^{-1} X' \) with \( \gamma \) being a small positive constant and \( I \) also denotes the identity matrix. The ultimate effect on representing the test sample of the \( h \)-th class can be evaluated by

\[
u_h = \|y - \sum_{j=1}^{M} f_j x_j\| \tag{10}
\]

If \( m = \arg \min_{h} (u_h) \), then the test sample \( y \) is finally assigned to the \( m \) class.

3. Insight into the proposed algorithm

In this section, we give a detailed analysis to the proposed algorithm. CRC employs all training samples to linearly represent the test sample and then determines which class of training samples leads to the largest contribution of the test sample. However, a limited number of available training samples have become one bottleneck of CRC. These training samples cannot accurately reconstruct the test sample, which leads to the interference from the training samples that are not same class as the test sample. In order to avoid this kind of interference, we calculate the similarity between the training samples and put them into two new training sample sets, which lets the test sample to find its class more easily and quickly.

Take the ORL database as an example, which is a classic face database commonly used in face recognition. The second image in the ORL is used as the test sample, and the first one in each class of image is used as the training sample. Figure 1 shows the five training samples with the closest to the test sample and test sample’s coding coefficients by the two-step classification without virtual samples. Figure 2 shows the five training samples with the closest to the test sample and test sample’s coding coefficients by our method. Because the test sample is assigned to the class of minimum residual, the second image in the ORL is recognized wrongly in the two-step classification without virtual samples [8]. However, this image can be easily identified correctly by our method.

We add the idea of training samples fragmentation to the two-step classification. Compared with the two-step classification, our method reduces the interference of the training samples that are not same class as the test sample, so the recognition rate has been improved. Figure 3 shows our method compared with the CRC [5] and the two-step classification without virtual samples in the FERET database [8].
Figure 1. The closest 5 classes of face images and residuals in the second image of the ORL database by the two-step classification.

In addition to the inter-class scattering, histogram similarity and the new fusion method can also measure the similarity between each class of training samples. Histogram similarity obtains similarity values only by comparing the histogram information of each class of training samples. Therefore, the new fusion mechanism (6) based on histogram similarity and Euclidean distance not only contains histogram information, but also makes Euclidean distance supplement the similarity value. The recognition rates of them are better than the inter-class scattering, but their time and computational complexity are higher. Detailed experimental results are shown in Section 4.

4. Simulation Results
We conduct face recognition on three different databases including ORL, GT, and FERET database to test our method. In these training databases, our method not only effectively compensates for the defect of Euclidean distance, but also reduces the interference of these training samples that are not same class as the test sample. Besides, to show the performance of our proposed algorithm (PA), single-scale pixel processing method (MSA) [10], linear Regression classification (LRC) [11], RBTM [12], CIRLRC [13], NFRFR [14], KRBM [15], FSSP [9], DSSR [16] are used as comparisons.

4.1. Experiments on the ORL database
The ORL database contains images from 40 people, and each person provides 10 images. Figure 4 shows one person from the ORL face database. The first 3 samples of each class are used as training samples, and the rest are used as test samples. The experimental results are shown in Table 1. The training samples fragmentation reduces the interference of these training samples that are not same class as the test sample, and the new fusion mechanism compensates for the defect of Euclidean distance, which makes PA have advantage over the most representation-based methods.

Table 1. The rate of classification errors of different methods on ORL database (%)

| Training sample per class | PA (inter-class scattering) | PA (histogram) | PA (new fusion method) | MSA | LRC | RBT | CIRLRC | NFRFR | KRBM | DSRR | FSSP |
|--------------------------|---------------------------|---------------|-----------------------|-----|-----|-----|--------|-------|------|------|------|
| 1                        | 20.00                     | 20.28         | 20.00                 | 23.06| 32.25| 31.94| 26.39  | 29.44 | 27.22| 28.06| 28.62|
| 2                        | 8.44                      | 7.50          | 6.88                  | 12.15| 20.62| 20.62| 13.12  | 12.81 | 11.87| 13.13| 14.69|
| 3                        | 6.43                      | 4.64          | 5.71                  | 11.43| 18.57| 21.79| 10.36  | 10    | 10.71| 10.36| 10.36|
4.2. Experiments on the FERET database
We use a subset of the FERET database to test our method. The FERET database contains 700 images from 100 people, and each person provides 7 images of different poses and illuminations. Let the first 1, 2, and 3 face images of each person be used as training samples, and the remaining face images as test samples. The images of FERET database and the experimental results are shown in Figure 5 and Table 2 respectively. The proposed method shows strong robustness to angle variations, so the recognition accuracy rates have been improved greatly compared with the most representation-based methods.

4.3. Experiments on the GT database
The GT database includes images of 50 people. Everyone has 15 color images and the background is complex. These images show the front side with different expressions, lighting conditions and angles. Figure 6 shows an example of the image from the GT database. The first 3 images of each person are used as training samples, and the other images are test samples. The experimental results show in Table 3. Because the histogram information of the images is greatly helpful to improve the recognition rate of GT database, the proposed method fuses histogram similarity and Euclidean distance to respectively compare with the most representation-based methods, so that it obtains better performance.

5. Conclusions
In this paper, we have proposed the ideas of training samples fragmentation and new classification fusion mechanism. They are helpful for overcoming the drawback of limited training samples in the practical face recognition system. The training samples fragmentation has reduced the interference
from those training samples that are not same class as the test sample, and three methods of training samples fragmentation have measured the similarity between the training samples well. Furthermore, the new classification fusion mechanism has compensated for the weak point of Euclidean distance. These have increased the recognition rate of the CRC. The experimental results show that the proposed method is superior to the most representation-based methods.

Acknowledgments
This article is partly supported by the Natural Science Foundation of Gansu Province of China (18JR3RA286), the Fundamental Research Funds for the Central Universities (lzujbky-2018-128), CERNET IPv6 Innovation Project (NGII20170602), Key Laboratory for Research of Design and Electromagnetic Compatibility of High Speed Electronic Systems, Ministry of Education No.2017001.

References
[1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[2] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Machine Intell. 31 (2) (2009) 210–227.
[3] M. Yang, L. Zhang, J. Yang, D. Zhang, Robust sparse coding for face recognition, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2011, pp. 625–632.
[4] K. Hui, C. Li, L. Zhang, Sparse neighbor representation for classification, Pattern Recogn. Lett. 33 (5) (2012) 661–669.
[5] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition?, in: Proc Int. Conf. Computer Vision, 2011, pp. 471–478.
[6] Xu Y, Zhu X, Li Z, et al. Using the original and ‘symmetrical face’ training samples to perform representation based two-step face recognition[J]. Pattern Recognition, 2013, 46(4).
[7] B. Tang, S. Luo, H. Huang, High performance face recognition system by creating virtual sample, in: Proceedings of the International Conference on Neural Networks and Signal Processing, 2003, pp. 972–975.
[8] Dong X, Zhang H, Sun J, et al. A two-stage learning approach to face recognition[J]. Journal of Visual Communication and Image Representation, 2017, 43:21-29.
[9] Tian C, Sun G, Zhang Q, et al. Integrating Sparse and Collaborative Representation Classifications for Image Classification[J]. International Journal of Image and Graphics, 2017, 17(02):1750007.
[10] Y. Xu, B. Zhang, Z. Zhong. “Multiple representations and sparse representation for image classification”, Pattern Recognition Letters, 2015, 68: 9-14.
[11] I. Naseem, R. Togneri, M. Bennamoun. “Linear regression for face recognition”, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010, 32(11): 2106-2112.
[12] Y. Xu, Q. Zhu, Z. Fan, Y. Wang, J.-S. Pan, From the idea of sparse representation to a representation-based transformation method for feature extraction, Neurocomputing, 113,168-176,2013.
[13] Y. Xu, X. Li, J. Yang, Z.Lai, D.Zhang, Integrating conventional and inverse representation for face recognition, IEEE Transactions on Cybernetic s., 44(10), 1738-1746, 2014.
[14] Y. Xu, X. Fang, J. You, Y. Chen, H.Liu, Noise-free representation based classification and face recognition experiments. Neuro-computing, 147,307-314,2015.
[15] Y.Xu,Z.Fan,Q.Zhu,Feature space-based human face image representation and recognition, Optical Engineering. 51 (1) (2012) 017205-017205-7
[16] Xu Y, Zhong Z, Yang J, et al. A New Discriminative Sparse Representation Method for Robust Face Recognition via l2 Regularization[J]. IEEE Trans Neural Netw Learn Syst, 2017, 28(10):2233-2242.