CrossMark

ORIGINAL ARTICLE

Undersampled face recognition based on virtual samples and representation classification

Jun Yang[1,3] ⓘ · **Yanli Liu**[2]

Abstract Classifiers such as sparse representation or collaborative representation can get good performance in face recognition. But these methods require a number of train samples in each class to construct the dictionary. On the condition of undersampled train samples, their performance decreases dramatically. A novel method is proposed in this paper to address undersampled face recognition problem. Firstly, virtual face images are generated by principal component analysis and mirror transform. Secondly, the test sample is collaboratively represented by the augmented train samples and is recognized by classifier based on representation. A number of face recognition experiments on three benchmark face database show that the recognition accuracy of our method is greater than that of a similar method, while time efficiency of our method is competitive to similar method.

Keywords Face recognition · Virtual face images · Representation classification · Principal component analysis

✉ Jun Yang
  jkxy_yjun@sicnu.edu.cn

1 College of Computer Science, Sichuan Normal University, Chengdu, China

2 College of Mathematics and Software, Sichuan Normal University, Chengdu, China

3 Visual Computing and Virtual Reality Key Laboratory of Sichuan Province, Sichuan Normal University, Chengdu, China

1 Introduction

Face recognition technology can be widely used in business and security fields, such as attendance system, criminals monitoring, and Id verification. So face recognition [1–7] has been a hot research topic in the field of pattern recognition over the past decades. Researchers have proposed many algorithms for face recognition. Among them, the subspace analysis methods such as the principal component analysis [8–14], the linear discriminant analysis [15–20], and the independent component analysis [21, 22] were used most widely. These methods are usually composed of two stages: feature extraction and classification. In feature extraction stage, a projection transformation matrix is got by learning on the training set, and then samples are projected onto the feature subspace by the matrix. In classification stage, some classifiers such as the nearest neighbor or support vector machine classifier are used to classify testing samples.

Recently, a novel method called sparse representation classifier (SRC) [23] was proposed and achieved excellent experiment results for face recognition. The core of the algorithm is to use all the training samples to represent the test sample by linear combination [23–28]. The combination coefficient is constrained by the sparse, that is to say, most of the component of the coefficient are zeros. After combination coefficient was got, for a test sample, the representation residual to each class can be calculated by the test sample and linear combination result of the training samples in that class. Finally, the test sample is classified to class with minimum residual.

Previous researches have already proved that the sparse representation method can obtain a very good performance for face recognition [29]. However, the computational cost of the method is very high because it depends on an

iterative algorithm to calculate the sparse representation coefficient. Zhang [30] proposed a method called collaborative representation classifier based on regularized least squares (CRC_RLS). This method utilizes collaborative representation (linear combination of all training samples) to represent the test samples and the linear combination coefficient is deduced by regularized least squares. The recognition rate of CRC_RLS is comparable with that of SRC, but computational cost of the CRC_RLS is far less than that of SRC. Xu et al. [26] proposed a simple and fast representation-based (SFR) face recognition method. The method selects only the nearest training sample to the test sample from every class and then expresses the testing sample as a linear combination of all the selected training samples. The method can classify more accurately than the nearest neighbor classification. Tang et al. [27] proposed a novel sparse representation method based on virtual samples. The virtual samples were created by adding small noise to the original training samples. Wang et al. [35] improved the method proposed in [27]. They divided an original image into several parts and generated a corresponding virtual training sample via adding different random noise to pixels of different parts. So, various illumination changes could be simulated.

In this paper, a novel collaborative representation method based on virtual samples (CRVS) is proposed. Firstly, the virtual samples of training samples are produced by principal component analysis (PCA) reconstruction and image mirror transform. Then, by the original sample and new virtual samples as the dictionary, the CRC_RLS is employed to express the test sample by the augmented training samples and to classify the test sample. A number of FR experiments show that our proposed method outperforms the method proposed in [26] and other methods based on virtual samples.

The rest of this paper is organized as follows: Sect. 2 introduces the proposed algorithm. Section 3 shows the experiments results for face recognition. Finally, Sect. 4 comes to the conclusion.

2 The representation classification based on augmented sample set

2.1 Generating virtual images

The proposed method in the paper includes two steps. Firstly, the virtual samples are created by PCA reconstruction and image mirror transform. Secondly, the new training sample set that consists of the original and virtual samples is used as the dictionary to represent the test sample. CRC_RLS is utilized to calculate the representation coefficient and classify the test sample.

To keep virtual sample with similar to the original image, PCA reconstruction is employed to produce virtual sample. PCA is a common statistical method, often used for data dimension reduction. Here, we take advantage of the nature of the minimum reconstruction error for PCA and select the most important base vectors of PCA to reconstruct the original sample.

Let $x_{k,i}(k = 1, 2, \ldots, C \ i = 1, 2, \ldots, m)$ denotes a train sample. Here, the k denotes that the sample is from the k th class and i denotes that it is the i th sample in that class. C and m, respectively, are the number of class and sample number of each class.

In the process of PCA, first of all, the samples are subtracted by the mean. Let $x'_{k,i} = x_{k,i} - \overline{x}$, where $\overline{x}$ denotes the mean of all training samples. Then, the covariance matrix can be created by Eq. (1).

$$M = \sum_{k=1}^{C} \sum_{i=1}^{m} x'_{k,i} x'_{k,i}{}^{T} \tag{1}$$

By calculating the eigenvalue and eigenvector of matrix M, a series of eigenvector can be got. Let $v_j(j = 1, 2, \ldots N - 1)$ denotes an eigenvector, where N is the number of training sample. The eigenvalue is corresponding to the important degree of the eigenvector for PCA reconstruction and those eigenvectors with large eigenvalues can be able to reconstruct the original sample better. So we sort eigenvectors from large to little according to corresponding eigenvalues. A sample is projected into PCA subspace and the projected coefficients are got by Eq. (2).

$$a_j = x_{k,i}{}^{T} v_j \tag{2}$$

If the first L eigenvectors are selected as the basis vectors, the reconstruction sample of the original sample is got by following equation.

$$x^1_{k,i} = \sum_{j=1}^{L} a_j v_j \tag{3}$$

As shown in Fig. 1, we can see that sample of PCA reconstruction is similar to original sample.
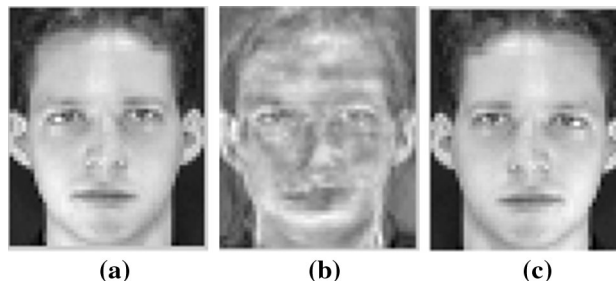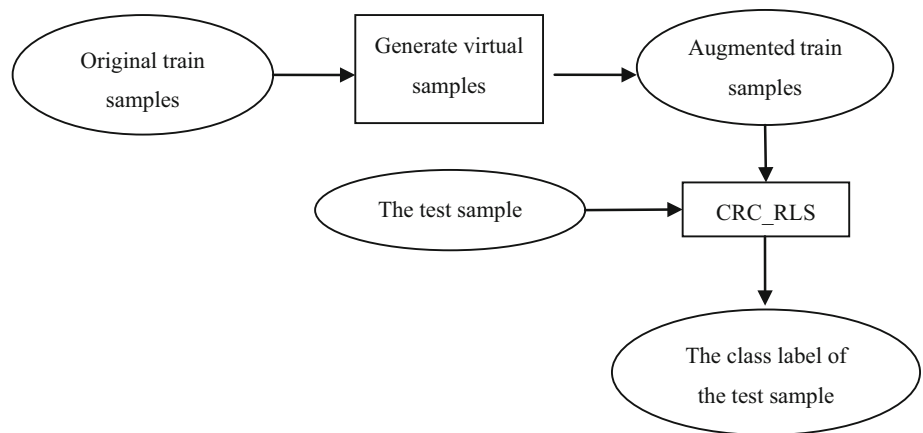

Fig. 1 An original face and virtual faces of it produced by PCA reconstruction and mirror transform. **a** The original face. **b** The PCA reconstruction face (first 90% eigenvectors are selected). **c** The mirror face

Fig. 2 The flow diagram of the proposed method

In order to improve the recognition ability under change of test samples, the proposed method employs mirror transform to generate new sample. Front face image is usually symmetric, and the mirror transform does not destroy the overall structure of the face image. A new face with mild pose change can be got and with the new face, the matching probability of training samples and test samples will be improved. We get the mirror image by Eq. (4).

$$x_{k,i}^2(r,c) = x_{k,i}(r, nc - c + 1) \tag{4}$$

In above equation, (r, c) denotes the pixel position of an image and nc is the width of the image. Figure 1 shows an original training sample and virtual samples produced by it.

2.2 Collaborative representation classification by extensive samples

By the previous process, we get some virtual samples and these virtual samples will be used in the following step. Then, it comes to the second phase of the proposed method. We create a matrix X including the all original train samples and their virtual samples. Let y denotes the test sample. Regarding X as the dictionary, we can represent y by the linear combination of the column vectors of X. In other words, it assumes that the following equation is approximately satisfied:

$$y = \sum_{i=1}^{N} a_i x_i \tag{5}$$

where N is the size of column vectors of X, and x_i is the i th column of X and a_i is the coefficient of x_i for linear combination. The equation can be rewritten as follows:

$$y = Xa \tag{6}$$

All face images have similar appearance, so it is reasonable that expressing collaboratively new face by the linear combination of all the training images. The

regularized least squares method can be used to solve the equation and the liner combination coefficient can be got by Eq. (7).

$$a = (X^T X + \lambda I)^{-1} y \tag{7}$$

where, I is the unitary matrix and λ is positive constant as the regularization parameter. After the combination coefficient a is obtained, we can get the representation result from train samples of each class by following equation.

$$y_k = \sum_{i=1}^{nk} a_i x_{k,i} \quad k = 1, 2, \ldots L \tag{8}$$

where, nk is the samples number of kth class in X. The deviation of test sample and representation sample can be calculated by following equation.

$$d_k = \|y - y_k\| \tag{9}$$

The deviation reflects the ability to represent test sample of that class. The smaller d_k is, the greater contribution the samples from the k th class makes to represent the test sample. So, y will be classified into the class that produces the smallest deviation. That is to say, if $d_t = \min(d_k)$, the test sample is classified into the tth class.

The whole flow of recognizing a face in our proposed method is shown in Fig. 2.

3 Experimental results

To evaluate the performance of the proposed method, a number of experiments are done on three benchmark face image databases: ORL, Yale, and AR. For showing the effectiveness of our proposed method, other similar methods proposed in [26, 27, 35] were implemented and experimented based on same face data. Moreover, as a preprocessing method before training, the getting virtual samples method proposed in this paper can be combined with many classification algorithms. Recently, some

complex but efficient algorithms-based SRC were proposed, such as the Regularized Robust Coding (RRC) in [36]. In experiments, the RRC method was used for face recognition, and the combination of virtual samples with RRC was also done.

The ORL face database [31] consists of 40 subjects. Every subject is composed of 10 images with various facial expressions, varying illumination and facial details. We get cropped images from ref [26] and the size of the image is 56×46. Figure 3 shows some cropped images of one person in ORL database. For the dictionary in representation classification is need to be over completed, we resize each image to 14×11 in the experiment.

The Yale face database [32] consists of 165 face images of 15 individuals, each providing 11 different images. The images are in upright, frontal position under various facial expressions and lighting conditions. In our experiments, each image is manually cropped and resized to 32×32 [33]. Figure 4 shows some cropped images of one person in Yale database. We resize each image to 16×16 in the experiment.

For experiments on ORL and Yale face database, if m samples of all the n samples per subject are chosen as train samples and the remaining samples are used as test samples, then the number of possible training sets and test sets is: C_n^m. The experiment results are the mean of these test sets.

The AR face database [34] contains over 3200 frontal face images of 126 different individuals. Each individual has 26 different images taken in two different sessions separated by 2 weeks intervals, and each session consists of 13 faces with different facial expressions, illumination conditions and occlusions. In our experiment, we choose a subset of AR consisting of 50 men and 50 women as [30] did, and all images are cropped to 60×43 pixels and converted to grayscale. Figure 5 shows the images of one

person used in our experiments. We resize each image to 15×11 in experiments.

We segment the AR database into three subsets, and experiments were done on the subsets. Subset one includes 1st–4th images of each person in session 1. The images in the subset are variant facial expressions. Subset two includes 1st, 5th–7th images of each person in session 1. The images in the subset are variant illumination. In experiments on subset one and subset two, first one image is selected as train sample and the remaining samples are used as test samples. Subset three includes all images showed in Fig. 5 and the images in session 1 are included in train set and others are in test set.

In all experiments, we firstly convert each sample vector into a unit vector with the length of 1 in advance. We use the regularized least squares to obtain the coefficients for linear combination and set the regularized parameter a to be 0.01. When getting the virtual sample by PCA reconstruction, we select first 95% eigenvectors as the reconstruction basis.

For Yale database, the number of train samples of each class is from 1 to 5. The experiment result is shown in Table 1.

From the table, we can see that the recognition rate of our proposed method is always better than that of other methods on this database, no matter how many images are selected as train samples. In addition, by combing the virtual samples and RRC, the recognition accuracy of RRC can be increased on Yale face database. The average recognition accuracy is higher 13.22% than that of method proposed in [26]. When five samples of each person are selected as train samples, we can get 462 train sets by combination. Figure 6 shows the recognition rate (%) on each train set. From Fig. 6, we can see that the recognition rate of our method is greater than that of [26] on most train sets.

Fig. 4 Images of a person from Yale database

Fig. 5 the images of a person from AR database used in our experiment. (The images in first row are from session 1 and the images in second row are from session 2)

Table 1 Means of the recognition rates (%) of our proposed method and other methods on the Yale face database

Number of training samples per class	1	2	3	4	5
The method proposed in [26]	46.73	58.03	63.33	66.78	69.34
The method proposed in [27]	46.42	65.70	74.43	78.90	81.76
The method proposed in [35]	49.09	68.26	76.63	81.02	83.59
RRC [36]	44.73	62.81	69.86	73.70	76.20
Virtual samples + RRC	53.76	66.34	72.61	76.04	78.73
Our proposed method	54.91	70.44	78.00	82.24	84.72
Number of training sets	11	55	165	330	462

For ORL database, the number of train samples of each class is from 1 to 4. The experiment result is shown in Table 2.

From the table, we can see that the accuracy of our proposed method is higher than that of others when 1 or 2 images are selected as train samples. When 3 or 4 images are selected as train samples, the accuracies of all methods are similar. It shows that our method is more efficient at scenario with undersampled train set, especially when the number of training samples is very little, such as one training sample. When 4 images of each person are selected as train samples, there are 210 training sets. The variances of accuracy of our method and the method in [27] are 1.61% and 1.84%. That shows our method is more stable to the method in [27].

The experiment results on AR database are shown in Table 3.

Fig. 6 The comparison of the recognition rate of each train set on Yale database between the proposed method and the method mentioned in [26]
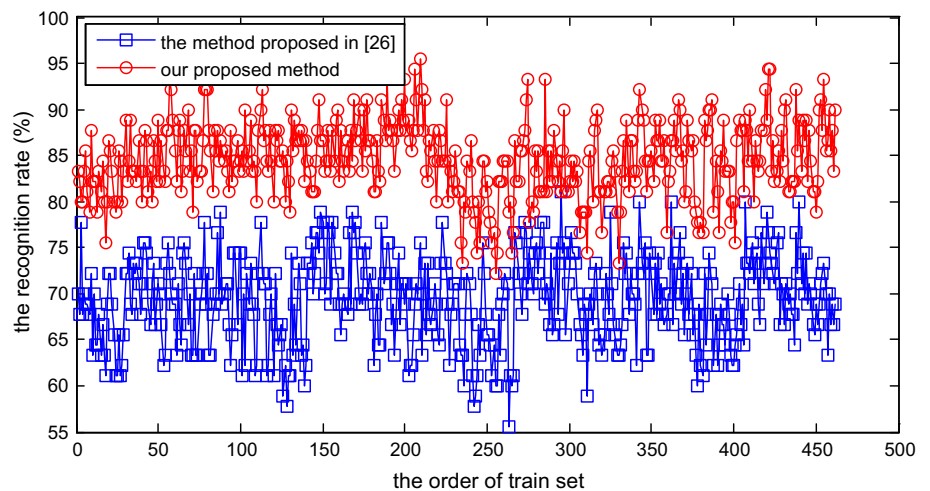
Table 2 Means of the recognition rates (%) of our proposed method and other methods on the ORL face database

Number of training samples per class	1	2	3	4
The method proposed in [26]	71.28	83.80	89.54	92.75
The method proposed in [27]	70.11	84.22	90.20	93.16
The method proposed in [35]	70.22	84.10	89.76	92.87
RRC [36]	62.22	77.89	85.07	88.96
Our proposed method	73.48	85.45	90.06	92.58
Number of training sets	10	45	120	210

Table 3 The recognition rates (%) of our proposed method and other methods on the AR face database

Subset	One(expressions)	Two(illumination)	Three(session)
The method proposed in [26]	85.66	45	84.55
The method proposed in [27]	85.33	49	86.70
The method proposed in [35]	83.67	80	86.12
RRC [36]	46.67	86.33	96.72
Our proposed method	79.67	78.67	88.98

From the table, we can see that although the performance of our proposed method is inferior to that of other methods on expressions subset, the recognition rate of our proposed method is comparative to that of others on illumination subset. On this database, the RRC method got most high recognition rate. But on expressions subset, the performance of RRC is poor. The recognition accuracy is higher 4.43% than that of method proposed in [26] when images in session 1 are selected as train samples (subset three).

The algorithm mentioned in [26] is simple and fast. Similar to that algorithm, our method also calculates the representation coefficients by regularized least squares. So the calculation efficiency of our method is comparable with algorithm mentioned in [26]. Different from method of [26], our method uses more samples including original and virtual samples to represent a test sample. So an additional step is needed for generating virtual samples in our method, and this will increase the computational burden of our proposed method. But the additional work can be done at training step, and the time efficiency of recognition step can be still kept. We record the times on experiment at 462 train sets, when five samples of each person are selected as train samples at Yale database. The time of our method is 25.23 s and the time of [26] is 45.53 s. The running time of our method is far less than that of [26]. So in time efficiency of test sample recognition, our proposed method is far better than method proposed in [26].

4 Conclusions

For representing the test sample better, we exploit the original and virtual images to construct new train sets. The method has two advantages. Firstly, the train samples are undersampled in face recognition problem and the dictionary composed by these samples in collaborative representation problem is usually not over completed. The number of train samples increases by generating virtual samples and the new dictionary can be over completed. Secondly, new generated sample represents some change of a face. So when matching a test face with unknown facial change, it seems that a more similar train sample can be found on the augmented samples. A lot of experiments show that the proposed method outperforms to some similar methods proposed in [26, 27, 35].

Compliance with ethical standards

Conflict of interest Authors declare that they have no conflict of interest.

References

1. Turk M, Pentland A (1991) Eigenfaces for recognition. J Cogn Neurosci 3(1):71–86
2. Wang X, Tang X (2005) Subspace analysis using random mixture models. In: IEEE computer society conference on computer vision and pattern recognition, 2005, CVPR 2005. IEEE
3. Lu J, Tan Y-P (2011) Improved discriminant locality preserving projections for face and palmprint recognition. Neurocomputing 74(18):3760–3767
4. Fukunaga K (2013) Introduction to statistical pattern recognition. Academic Press, London

5. Kafai M, An L, Bhanu B (2014) Reference face graph for face recognition. IEEE Trans Inf Forensics Secur 9(12):2132–2143
6. Ding C, Xu C, Tao D (2015) Multi-task pose-invariant face recognition. IEEE Trans Image Process 24(3):980–993
7. Moeini A, Moeini H (2015) Real-world and rapid face recognition towards pose and expression variations via feature library matrix. IEEE Trans Inf Forensics Secur 10(5):969–984
8. Kirby M, Sirovich L (1990) Application of the KL procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12(1):103–108
9. Yang J, Zhang D, Frangi AF, Yang JY (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Mach Intell 26(1):131–137
10. Wang H, Chen S, Hu Z, Luo B (2008) Probabilistic two dimensional principal component analysis and its mixture model for face recognition. Neural Comput Appl 17(5–6):541–547. doi:10.1007/s00521-007-0155-1
11. Sun N, Wang H, Ji Z, Zou C, Zhao L (2008) An efficient algorithm for Kernel two-dimensional principal component analysis. Neural Comput Appl 17(1):59–64
12. Xu Y, Zhang D, Yang J (2010) A feature extraction method for use with bimodal biometrics. Pattern Recogn 43(3):1106–1115
13. Yang W, Sun C, Ricanek K, Yang W, Sun C, Ricanek K (2012) Sequential row–column 2DPCA for face recognition. Neural Comput Appl 21(7):1729–1735. doi:10.1007/s00521-011-0676-5
14. Zhu Q, Xu Y (2012) Multi-directional two-dimensional PCA with matching score level fusion for face recognition. Neural Comput Appl. doi:10.1007/s00521-012-0851-3
15. Mika S, Ratsch G, Weston J, Bernhard et al (1999) Fisher discriminant analysis with kernels. In: Proceedings of the IEEE neural networks for signal processing workshop, pp 41–48
16. Xu Y, Yang J, Jin Z (2004) A novel method for Fisher discriminant analysis. Pattern Recogn 37(2):381–384
17. Wang H, Li P, Zhang T (2008) Histogram feature-based Fisher linear discriminant for face detection. Neural Comput Appl 17(1):49–58
18. Li J, Pan J, Lu Z (2009) Kernel optimization-based discriminant analysis for face recognition. Neural Comput Appl 18(6):603–612
19. Li JB, Pan JS, Lu ZM (2009) Face recognition using Gabor-based complete Kernel Fisher discriminant analysis with fractional power polynomial models. Neural Comput Appl 18(6):613–621
20. Zhang B, Qiao Y (2010) Face recognition based on gradient Gabor feature and efficient Kernel Fisher analysis. Neural Comput Appl 19(4):617–623
21. Liu C, Yang J (2009) ICA color space for pattern recognition. IEEE Trans Neural Netw 20(2):248–257
22. Zhang L, Gao Q, Zhang D (2008) Directional independent component analysis with tensor representation. In: CVPR 2008, Anchorage, Alaska, U.S., 23–28 June, pp 1–7
23. Wright J, Yang AY, Ganesh A et al (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
24. Wright J, Ma Y, Mairal J et al (2009) Sparse. Representation for computer vision and pattern recognition. In: Proceedings of IEEE, pp 1–8
25. Xu Y, Zhang D et al (2011) A two-phase test sample sparse representation method for use with face recognition. IEEE Trans Circuits Syst Video Technol 21(9):1255–1262
26. Xu Y, Zhu Q (2013) A simple and fast representation-based face recognition method. Neural Comput Appl 22:1543–1549
27. Tang D, Zhu N, Yu F et al (2014) A novel sparse representation method based on virtual samples for face recognition. Neural Comput Appl 24(3–4):513–519
28. Zhu N, Li S (2014) A kernel-based sparse representation method for face recognition. Neural Comput Appl 24(3–4):845–852
29. Kroeker KL (2009) Face recognition breakthrough. Commun ACM 52(8):18–19
30. Zhang L, Yanga M et al (2011) Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of IEEE international conference on computer vision, pp 471–478
31. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
32. http://cvc.yale.edu/projects/yalefaces/yalefaces.html
33. www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html
34. http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html
35. Wang Y, Wang M, Chen Y et al (2014) A novel virtual samples-based sparse representation method for face recognition. Optik 125(15):3908–3912
36. Yang M, Zhang L, Yang J, Zhang D (2013) Regularized robust coding for face recognition. IEEE Trans Image Process 22(5):1753–1766