A NOVEL DICTIONARY BASED SRC FOR FACE RECOGNITION

Ying Wen

Shanghai Key Laboratory of Multidimensional Information Processing,
Department of Computer Science and Technology,
East China Normal University, Shanghai, China, 200062
E-mail Address: ywen@cs.ecnu.edu.cn

ABSTRACT

The sparse representation based classification (SRC) performs not very well for small sample data. A discriminative common vector dictionary based SRC is introduced in this paper to address this issue. The contribution of this paper is that the dictionary of the proposed method is constructed by the discriminative common vector per class. The common vector represents the invariant property of each class, which is helpful to improve the performance of the proposed method for small sample database. Furthermore, the new dictionary has much less atoms than the original SRC based scheme, which reduces the computational cost. The experiments implemented on ORL, AR and LFW face databases demonstrate the effectiveness of the proposed method.

Index Terms— Sparse representation classification, discriminative common vector, face recognition

1. INTRODUCTION

Sparse representation-based classification (SRC) was introduced by Wright et al. [1] for face recognition, in which the training images are used as the dictionary to code an input testing image as a sparse linear combination of them via $l1$-norm minimization. Some extended SRC methods have been proposed to boost the research of sparsity based face recognition [2],[3],[4],[5]. Huang et al. introduced a transformation-invariant SRC for face recognition [6]. Yang et al. introduced a discriminative dictionary learning into the SRC to improve the accuracy of face recognition [7]. Deng et al. proposed a superposed sparse representation-based classifier (SSRC) for undersampled face recognition [8], [9].

In the SRC based methods, the dictionary is constructed by all training samples. When the number of samples is big, the dimension of the dictionary is high, which makes the SRC based methods be a very time-consuming task. In addition, the presence of noise, occlusion, varying viewpoints, background clutter, and illumination changes may make the dictionary contain too much redundant information. But if the number of samples is small, the performance of the SRC may be poor. It follows that the SR based classifier is time-consuming for big sample data and is hard to achieve a satisfactory result for small sample data.

To overcome the above shortcomings, we propose a discriminative common vector dictionary based SRC (DCV-SRC) for face recognition. The discriminative common vector method is based on a variation of Fisher's linear discriminant analysis for the small sample size case [10]. The common vector represents the invariant property of each class and are robust to the appearance of an individual's face, illumination, background, noise, etc. [11], [12]. Thus, using the discriminative common vector per class to construct the dictionary of the SRC can improve the performance of the classifier for small sample data. Since the number of discriminative common vectors equals to the number of classes, the size of the dictionary greatly decreases. As a result, our algorithm is faster than the SRC and its variants.

This work was supported by The National Natural Science Foundation of China (No.61273261) and the Science and Technology Commission of Shanghai Municipality under research grant no. 14DZ2260800.

2. METHODS

2.1. Sparse Representation-based Classification

Given sufficient training samples of the ith object class, any test sample from the same class will approximately lie in the linear span of the training samples associated with object i. Wright et al. [1] defined a matrix D for the entire training set as the concatenation of the N training samples of all C object classes:

$$D = [D_1, D_2, ..., D_C] = [x_1^1, x_2^1, ..., x_n^C] \tag{1}$$

where $D_i = [x_1^i, x_2^i, ..., x_n^i] \in R^{dim\times n}$ consists of the training samples of the ith class, x_i is the ith, $1 \leq i \leq N$, data object and n is the sample number of each class.

Then, the linear representation of y can be rewritten in terms of all training samples as:

$$y = Da_0 + z \tag{2}$$

where $a_0 \in R^n$ is a coefficient vector whose entries are zero except those associated with the ith class. $z \in R^{dim}$ is a noise term with bounded energy $\|z\|_2 < \varepsilon$. The sparse solution a_0 can be approximately recovered by solving the following stable ℓ^1-minimization problem:

$$(\ell_s^1) : \hat{a}_1 = \arg\min \|a\|_1 \text{ subject to} \|Da - y\|_2 \leq \varepsilon \quad (3)$$

Ideally, the nonzero entries in the estimate $\hat{a}_1$ will all be associated with the column of D from a single class. The classification result depends on the reconstruction of nonzero coefficients in each class. SRC classifies the signal y into the class which results in the minimal reconstruction error.

2.2. Discriminative Common Vector Dictionary based S-RC (DCV-SRC)

The dictionary of the SRC is constructed by all training samples. Generally, the dictionary size of the SRC is very large. The big dictionary of the SRC will increase the running time and complexity of algorithm. In addition, the dictionary constructed by the original samples contains redundant information, which may decline the algorithm performance. In some extended SRC methods, such as ESRC and SSRC, the dictionary consists of not only the differences between training samples and mean samples but also all training samples or mean samples, so that the dictionary size increases. How to remove the redundant information in the dictionary, decrease the dictionary size and improve the classifier's performance are still a challenge work.

In this paper, we use the discriminative common vector to construct a new dictionary of the SRC. Discriminative common vector [10] projects the face samples onto the null space Q of the within class scatter matrix S_w and then obtains the projection vectors by performing PCA. The method is described as follows:

Let x_m^i be a dim-dimensional column vector which denotes the mth sample from the ith class. The within class scatter matrix S_w is computed by:

$$S_w = \sum_{i=1}^{C} \sum_{m=1}^{n} (x_m^i - u_i)(x_m^i - u_i)^T \quad (4)$$

where u_i is the mean of samples in the ith class.

Choose any samples from each class and project it onto the null space Q of S_w to obtain the common vectors:

$$x_{com}^i = x_m^i - QQ^T x_m^i, i = 1, ..., C \quad (5)$$

where $Q = [\alpha_1 ... \alpha_r]$ is the set of orthonormal eigenvectors corresponding to the nonzero eigenvalues of S_w, and r is the rank of S_w .

In this way, we obtain C common vectors. The common vector of each person is obtained by removing the differences of face images belonging to the same person and presents common invariant property of ith face class [12]. Then we apply the principal component analysis to the common vectors.

Table 1. Dictionary comparison of different algorithms.

Method	Dictionary size	Storage size	Running time
NN	$dim \times n \times C$	large	fast
SRC	$dim \times n \times C$	large	slow
SSRC	$dim \times (n \times C + C)$	very large	very slow
DCV-SRC	$(C-1) \times C$	small	fast

$$S_{com} = \sum_{i=1}^{C} (x_{com}^i - u_{com})(x_{com}^i - u_{com})^T \quad (6)$$

where u_{com} is the mean of all common vectors.

According to Eq.(6), we form the projection matrix $W = [W_1 ... W_{C-1}]$, the eigenvectors corresponding to the nonzero eigenvalues of S_{com}. Then we obtain the projection vectors of common vectors:

$$\Omega_i = W^T x_{com}^i, i = 1, ..., C \quad (7)$$

where, Ω_i is the discriminative common vector (DCV). DCV only has $C-1$ dimension because the rank of S_{com} is $C-1$ when all common vectors are linearly independent. We use the discriminative common vectors to construct a new dictionary, which is defined as $\tilde{D}$:

$$\tilde{D} = [\Omega_1, \Omega_2, ..., \Omega_C] \quad (8)$$

The projection vector of a testing sample x_{test} is $\Omega_{test} = W^T x_{test}$, and the linear representation of Ω_{test} can be rewritten in terms of all discriminative common vectors as:

$$\Omega_{test} = \tilde{D} a_0 + z \quad (9)$$

Compare Eq.(2) with Eq.(9), a main difference between the SRC and the proposed method is the dictionary design, and the subsequent solution of our method is similar to that of the SRC.

In general, the number of training samples ($N = n \times C$) is much bigger than the number of classes, i.e., $N \gg C$. And the number of classes is much less than the image size or the holistic features' dimension, i.e., $C - 1 \ll dim$. Hence, the size of the proposed dictionary is much smaller than that of NN, SRC and SSRC. The information of different methods is listed in Table 1. It can be seen that the rank of the dictionary size is SSRC, NN, SRC and our method in descending order. In theory, NN spends less computational cost due to its simple computation, the second is ours and the third is SRC. However, with the increase of the training size, DCV-SRC may be faster than the NN because the low dimension of the sample. The highest computation cost is SSRC.

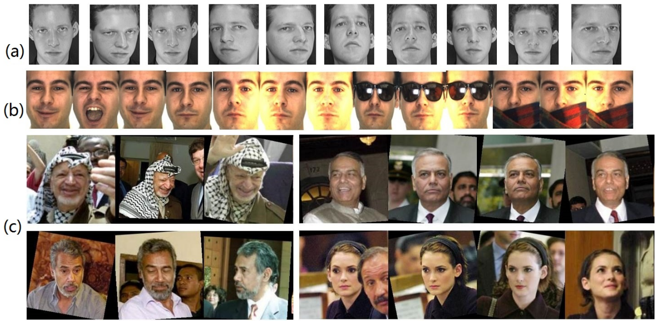

Fig. 1. The cropped images of one person from (a) ORL, (b) AR and (c) LFW databases.

3. EXPERIMENTAL RESULTS

In this section, we experimentally compare the proposed method with the SRC, SSRC, and NN on two simulated datasets and a wild dataset. The goal of selecting the simulated datasets is to validate the robustness to appearance, illumination and occlusion of the proposed method, while the goal of selecting a wild dataset is to validate the application performance of the proposed in real, various and complex conditions. In the following experiments, SRC, SSRC and DCV-SRC apply the Homotopy [13] method to solve the ℓ^1-minimization problem with the error tolerance $\varepsilon = 0.0001$.

3.1. Experimental on the simulated datasets

ORL and AR datasets are used to validate the performance of the proposed method. All images are cropped with the size of 32×32. We select $3 \sim 7$ images per individual in the datasets as the training set and the rest images as the testing set. For a given training size, we perform 20 times for all experiments and calculate the average recognition rates and the standard deviations. PCA is used to reduce the dimension of face image.

Robustness to Appearance: This experiment aims to evaluate the appearance robustness of the proposed method. The ORL (Olivetti-Oracle Research Lab) [14] database consists of 400 frontal-face images of 40 individuals. The facial expressions and details also vary. The images of one person are shown in Fig.1(a).

Table 2 shows dictionary size, accuracy and standard deviation of different methods implemented on the ORL dataset. For SRC, SSRC and NN, the dictionary size dramatically increases with the increase of the training size. However, the dictionary size of DCV-SRC is fixed and always is 40×39 ($C \times (C-1)$). It can be seen that the proposed method not only decreases the dictionary size but also achieves the highest recognition accuracies in various training size.

Robustness to Occlusion: This experiment aims to evaluate the robustness to facial occlusion. The AR database consists of over 4,000 frontal images for 126 individuals [15]. In the experiment, we choose a subset of the data set consisting of 702 frontal images for 54 individuals. These images include more facial variations and disguises. Thirteen images for an individual is shown in Fig.1(b).

Table 2. Experimental results of four methods from the ORL dataset (Dictionary size, Accuracy(%), Standard deviation)

Method	NN	SRC	SSRC	DCV-SRC
3	120×119 77.89 ± 2.65	120×119 80.93 ± 2.83	160×119 84.73 ± 2.24	**40×39** **89.04±1.88**
4	160×159 85.96 ± 2.45	160×159 88.29 ± 2.83	200×159 90.62 ± 2.00	**40×39** **93.48±2.06**
5	200×199 87.88 ± 2.83	200×199 91.07 ± 3.00	240×199 92.10 ± 2.83	**40×39** **94.98±2.46**
6	240×239 91.22 ± 2.45	240×239 94.69 ± 1.73	280×239 94.31 ± 2.00	**40×39** **96.12±1.82**
7	280×279 91.67 ± 2.65	280×279 95.58 ± 2.24	320×279 95.12 ± 2.24	**40×39** **96.54±1.69**

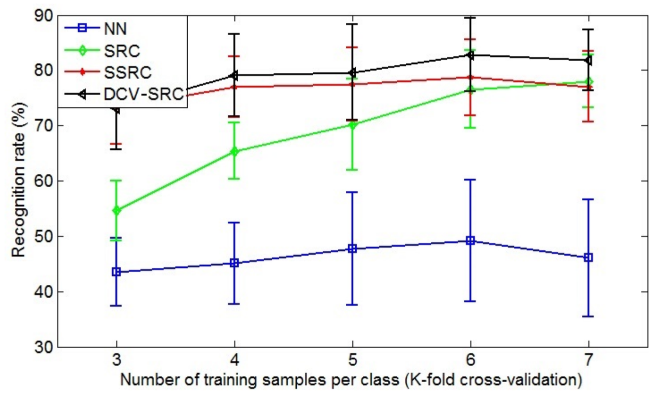

Fig. 2. Recognition rate on the occlusion images with sunglass from the AR dataset

First, we evaluate the performance of the proposed method by considering the occlusion image with sunglass. We randomly choose one image with sunglass and some neutral images as the training set and the remaining images as the testing. The average results and the standard deviations are shown in Fig.2. It can be seen that DCV-SRC performs better than the NN, SRC, SSRC for the sunglass occlusion images.

Next, we randomly choose one occlusion image with sunglass or scarf and some neutral images for the training set and the remaining images for testing. The average results and the standard deviations are shown in Fig.3. With the increase of training samples, the superiority of our algorithm is obvious and the proposed method presents a good performance for the sunglass or scarf occlusion images.

From the above experimental results, we can see that the

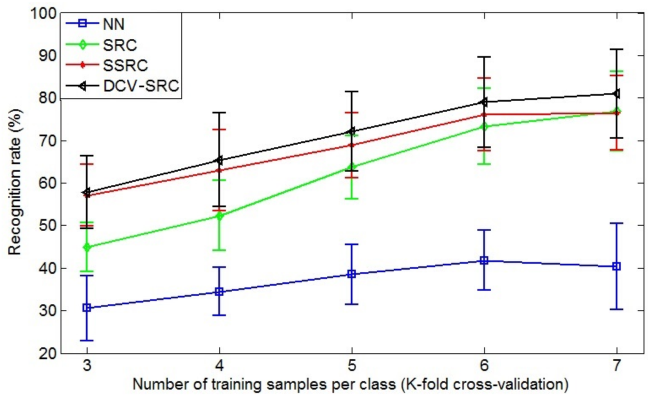

Fig. 3. Recognition rate on the occlusion images with sunglass or scarf from the AR dataset

proposed method is consistently higher than the others. Our proposed method is robust to illumination, appearance, and occlusion for face recognition, which contributes to the dictionary constructed by the discriminative common vector describing the unique property of its class.

3.2. Experiments in the wild dataset

Next, the proposed method is evaluated on the popular "Labeled Faces in the Wild dataset"(LFW) [16]. The dataset contains more than 13000 images and 1680 of the people pictured have two or more distinct photos in the data set. It is commonly regarded to be a challenging dataset for face recognition since the faces were detected from images taken from Yahoo! News and show large variations in pose, expression, lighting, and age etc. We use "deep funneled" dataset and some examples are shown in Fig.1(c). In the dataset, we select all persons with ≥ 3 photos to construct a dataset, in which there are 7613 images and 900 people. A person with the same number of photos is classified to the same subset. Finally, we obtain 13 subsets, in which there are $3 \sim 15$ photos per person separately.

We carry out 13 groups of experiment to evaluate the performance of DCV-SRC by "leave-one-out" method. Fig.4 shows average recognition rates and standard deviations of the NN, SRC, SSRC and DCV-SRC. It can be seen that the NN is the worst, the SRC comes second. The proposed DCV-SRC is the best in these four methods. The experimental results in the wild show that the performance of our method is stably superior to the NN, SRC and SSRC, and also demonstrate our method can be used for the classification in real, various and complex conditions.

3.3. Running Time

The running time comparison of different methods is implemented on the AR dataset. The number of the training samples per class is set from three to seven. And we repeat each experiment twenty times. All experiments are executed by MATLAB and performed on an Intel Core i3 M380 2.0GHz processor without any particular code optimization. The average running time per sample of each subset is reported in Table 3. From Table 3, we can see that NN is the fastest when the training size is small. Our method gets closer to the NN with the increase of the training size, and DCV-SRC is the fastest when the training size is 7. This is because our method has the small size of the dictionary although the training size increases. In these methods, the SSRC is the most time-consuming especially for the big training size.

All the experimental results show that the proposed method achieves a good performance with less time cost.

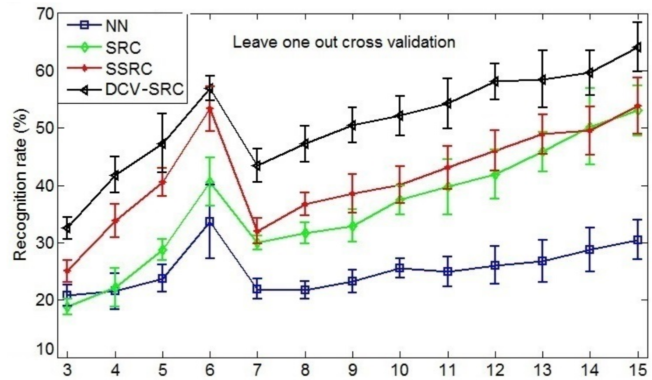

Fig. 4. Experimental results of four methods from the LFW dataset

Table 3. Average running time (ms) vs. training size using different methods.

Training size	3	4	5	6	7
NN	0.52	0.84	1.29	1.80	2.62
SRC	1.54	2.16	3.21	4.22	5.72
SSRC	28.52	49.26	98.78	211.59	372.89
DCV-SRC	0.89	1.16	1.48	1.89	2.44

4. CONCLUSION

In this paper, we propose a discriminative common vector dictionary based SRC for face recognition. In the method, we extract the discriminative common vector per class to construct a new dictionary for the SRC, which is robust to illumination, appearance and partial occlusion. Apart from the improvement of the recognition rate, one important contribution of our method is the dictionary, which has much less atoms than the original SRC based schemes. This greatly reduces the computational cost of sparse coding. The experimental results demonstrate the effectiveness of the proposed method.

5. REFERENCES

[1] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[2] A. Wagner, J. Wright, A. Ganesh, and Z. Zhou, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.

[3] H. Zhang, J. Yang, and Y. Zhang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," *IEEE Conference on Computer Vision*, pp. 770–777, 2011.

[4] J. Chen, J. Yang, L. Luo, J. Qian, and W. Xu, "Matrix variate distribution-induced sparse representation for robust image classification," *IEEE Trans. Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2291–2300, 2015.

[5] J. Wang, C. Lu, M. Wang, and P. Li, "Robust face recognition via adaptive sparse representation," *IEEE Trans. Cybernetics*, vol. 44, no. 12, pp. 2368–2378, 2014.

[6] J. Huang, X. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[7] M. Yang, D. Zhang, and X. Feng, "Fisher discrimination dictionary learning for sparse representation," *IEEE Conference on Computer Vision*, pp. 543–550, 2011.

[8] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 399–406, 2013.

[9] W. Deng, J. Hu, and J. Guo, "Extended src: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, 2012.

[10] H. Cevikalp, M. Neamtu, and M. Wilkes, "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 4–13, 2005.

[11] M. Gulmezoglu, V. Dzhafarov, and A. Barkana, "The common vector approach and its relation to principal component analysis," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 6, pp. 655–662, 2001.

[12] Y. He, L. Zhao, and C. Zou, "Face recognition using common faces method," *Pattern Recognition*, vol. 39, pp. 2218–2222, 2006.

[13] D. Donoho and Y. Tsaig, "Fast solution of l1-norm minimization problems when the solution may be sparse," *IEEE Trans. Information Theory*, vol. 34, no. 2, pp. 4789–4812, 2008.

[14] "http://www.cl.cam.ac.uk/research/dtg/ attarchive/facedatabase.html," .

[15] A. Martinez and R. Benavente, "The ar face database," *CVC Technical Report 24*, 1998.

[16] "http://vis-www.cs.umass.edu/lfw/," .