Fast Sparse Representation Classification Using Transfer Learning

Qi Zhu[1,2(✉)], Baisheng Dai[3], Zizhu Fan[4], and Zheng Zhang[5]

[1] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
zhuqi@nuaa.edu.cn
[2] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing, China
[3] School of Computing Science and Technology, Harbin Institute of Technology, Harbin, China
csbsdai@gmail.com
[4] School of Basic Science, East China Jiaotong University, Nanchang, China
zzfan3@163.com
[5] Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China
zhengzhang0812@163.com

Abstract. Under certain conditions, the sparsest solution to the combination coefficients can be achieved by L1-norm minimization. Many algorithms of L1-norm minimization have been studied in recent years, but they suffer from the expensive computational problem, which constrains the applications of SRC in large-scale problems. This paper aims to improve the computation efficiency of SRC by speeding up the learning of the combination coefficients. We show that the coupled representations in the original space and PCA space have the similar sparse representation model (coefficients). By using this trick, we successfully avoid the curse of dimensionality of SRC in computation and develop the Fast SRC (FSRC). Experimental results on several face datasets illustrate that FSRC has comparable classification accuracy to SRC. Compared to PCA +SRC, FSRC achieves higher classification accuracy.

Keywords: Biometrics · Sparse representation · Fast SRC · PCA

1 Introduction

Recently, by using compressed sensing, the sparse representation based classification (SRC) is proposed for face recognition [1], which has attracted much attention due to its good performance. The works [2–4] show SRC has better performance than the previous methods in robust face recognition. SRC needs to solve the l_1 norm optimization problem, which costs much computation time. Many researchers suggested solving the representation model with l_2 norm optimizer [5]. For example, Yanguses the whole training set to perform representation [4]. TPTSR is a local representation method [5]. The l_2 norm optimizer based methods are more efficient than SRC, but they cannot provide the sparse representation model, which plays important role in

Z. Huang et al. (Eds.): ICCCS 2015, LNCS 9483, pp. 501–509, 2015.
DOI: 10.1007/978-3-319-27051-7_43

classification. The computation efficiency of SRC constrains its applications. The main computation time of SRC is consumed in solving the sparse representation coefficients. This part of time increases greatly as the dimensionality of the sample increases. Consequently, the SRC is very time consuming or even unfeasible in many face recognition problems. Therefore, it is necessary to improve the classification efficiency of SRC. Learning the model from the other domain may be much easier and appropriate for classifying the data in the original domain, which is the main idea of transfer learning [6]. Transfer learning allows the domains used in training and test to be different, such as transfer via automatic mapping and revision (TAMAR) algorithm [7].

This paper aims to speed up the classification procedure of SRC by using transfer learning. Suppose that there exist coupled representations, i.e. high-dimensional representation (HR) and low-dimensional representation (LR), of the same image, the training samples in these two representations compose a pair of dictionaries. We assume the image has the similar sparse representation model on this pair of dictionaries. This assumption allows us to get the approximate solution of the coefficients in a low-dimensional space with a relatively low computation cost. In our method, we first convert the original (HR) test and training samples to the low high-dimensional space by K-L transform, and get the LR of the samples. The coefficients are learned by sparsely coding the LR test sample on the low-dimensional dictionary. Then, we reconstruct the HR test image by each class-specific HR face images with the obtained sparse coefficients. Finally, we classify the sample according to reconstruction error. It should be noted that our method is distinctly different from PCA+SRC that is the SRC performed in PCA space. Because the representation model of our method is learned in PCA space, the representation error of our method is calculated in original space. Otherwise, if representation error is directly calculated in the low-dimensional space, some discriminative features may be lost. This may explain why SRC after PCA, random-sampling or sub-sampling does not perform well shown in [6].

2 Preliminaries

2.1 SRC

Suppose there are n training samples from t classes. Class i has n_i training samples, and $n = \sum n_i$. Each image sample is stretched to column vector. Image sample is represented by $\mathbf{x}_i \in \mathbf{R}^{m \times 1}, m$ is the dimensionality of the sample. The test sample, e.g. $\mathbf{y} \in \mathbf{R}^{m \times 1}$, is represented as:

$$\mathbf{y} = \mathbf{XA} \tag{1}$$

We can get the solution by:

$$\mathbf{a} = \arg\min(||\mathbf{y} - \mathbf{Xa}||_2^2 + \mu||\mathbf{a}||_1) \tag{2}$$

where, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$,$\mu$ is the regular parameter, $||\ ||_1$ denotes the l_1 norm, and $||\ ||_2$ denotes the l_2 norm. For example, for the vector $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$, its l_1 norm is

$\sum_{i=1}^{k}|\mathbf{v}_i|$, and the l_2 norm is $\sqrt{\sum_{i=1}^{k}\mathbf{v}_i^2}$. After getting the coefficients, the representation error by each class can be derived by:

$$e_i = ||\mathbf{y} - \mathbf{X}_i\mathbf{a}_i||_2^2 \qquad (i = 1,\ 2,\ \ldots,\ t) \tag{3}$$

where $\mathbf{X}_i \in \mathbf{R}^{m\times n_i}$, and $\mathbf{a}_i$ is the combination coefficient vector corresponding to $\mathbf{X}_i$. Finally, $\mathbf{y}$ is classified according to the representation error.

2.2 PCA

PCA or Karhunen-Loeve (KL) transform is a useful dimensionality reduction used in signal processing. PCA finds d directions in which the data has the largest variances, and projects the data along these directions. The covariance matrixis defined as:

$$\mathbf{C} = \sum_{j=1}^{n}(\mathbf{x}_j - \overline{\mathbf{x}})^T(\mathbf{x}_j - \overline{\mathbf{x}}), \quad (j = 1,\ 2, \ldots,\ n) \tag{4}$$

where $\overline{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}x_j$. The transform projection vectors are d eigen-vectors, $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$, of the d largest eigen-values of the covariance matrix. PCA transform space is defined as $\Phi = span\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d\}$.

In many face recognition methods, PCA is usually used for dimensionality reduction before the classification. E.g., people often use the framework of PCA plus SRC. Indeed, performing PCA before SRC has the lower computational cost than SRC, and makes the solution of the combination sparser. But classifying the data using SRC in PCA space, i.e. PCA+SRC, cannot achieve promising accuracy, even much worse than that in original space. In next Section, we propose a novel framework than can speed up the classification of SRC without decrease of accuracy.

3 Fast SRC

SRC needs to solve l_1 norm minimization problem. Taking into account the very high dimensionality of the face image, SRC is very time consuming in face recognition. We aim to develop an efficient way to learn the representation model of SRC.

Proposition 1. The test image sample y can be coded by $\mathbf{Xa}$ in the original space, where $\mathbf{a} = \arg\min(||\mathbf{y} - \mathbf{Xa}||_2^2 + \mu||\mathbf{a}||_1)$. We use the function f to denote the transform from the original space to the new space, and then the test image in the new space, i.e., $f(\mathbf{y})$, can be coded by $f(\mathbf{X})\mathbf{a}'$, where $\mathbf{a}' = \arg\min(||f(\mathbf{y}) - f(\mathbf{X})\mathbf{a}'||_2^2 + \mu||\mathbf{a}'||_1)$. There exist some transforms from the original space to the new space, by which A is very close to A'.

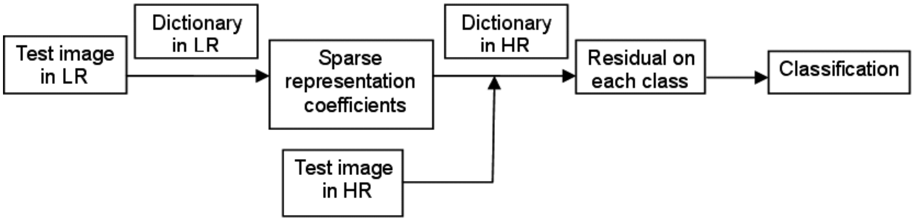

Fig. 1. Framework of the fast SRC

For example, with subsampling, the models in original space and sub-sampled space should be similar. By the l_1 norm optimizer, image y can be sparsely represented by $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + 0\mathbf{x}_3 + 0\mathbf{x}_4 + 0\mathbf{x}_5$, ($c_1 \neq 0$ and $c_2 \neq 0$), and y' can be sparsely represented by $c'_1\mathbf{x}'_1 + c'_2\mathbf{x}'_2 + 0\mathbf{x}'_3 + 0\mathbf{x}'_4 + 0\mathbf{x}'_5$, ($c'_1 \neq 0$ and $c'_2 \neq 0$). In the first model, if the representation coefficient vector $(c_1\ , c_2,\ 0,\ 0,\ 0)$ is replaced by $(c_1{}',\ c_2{}',\ 0,\ 0\ ,\ 0)$, the representation result of y becomes $c'_1\mathbf{x}_1 + c'_2\mathbf{x}_2 + 0\mathbf{x}_3 + 0\mathbf{x}_4 + 0\mathbf{x}_5$. In this example, the coefficient vector in the sub-sampled space should be an approximate or suboptimal solution of the coefficient vector in the original space.

In many image classification problems, the original test and training samples are always in high-dimensional representation (HR) space. In framework shown in Fig. 1, we project them onto the subspace, and low-dimensional representation (LR) of the images. The test face image is reconstructed class by class using HR face images with the corresponding sparse representation coefficients learned on the LR dictionary. After devising the framework of Fast SRC, we need to find the transform (from dictionary in HR to dictionary in LR) having the following two properties: (1) computationally efficient to calculate the transform (low cost for getting the dictionary in LR) (2) the similar sparse representation model to that in the original space. In the next two subsections, we will demonstrate K-L transform (PCA) meets the above requirements.

We will show the relationship between the representation modelsin original space and K-L transformspace by the following intuitive explanation. It is reasonable to assume the prior probability distribution of the training samples coincides with the real probability distribution of the samples. Let $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d$ be all the d orthonormal eigenvectors having non-zero eigenvalues obtained by PCA, and we denote $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_d)$. If y is coded by $\mathbf{Xa}$, where $\mathbf{a} = \arg\min(||\mathbf{y} - \mathbf{Xa}||_2^2 + \mu||\mathbf{a}||_1)$, we have $\mathbf{a} = \arg\min(||\mathbf{P}^T\mathbf{y} - \mathbf{P}^T\mathbf{Xa}||_2^2 + \mu_1||\mathbf{a}||_1)$, where μ_1 is the scalar constant.

Because, we know that the l_2 norm of the vector transformed into the K-L transform space is equal to that in original space. Then, if $||\mathbf{y} - \mathbf{Xa}||_2^2 > ||\mathbf{y} - \mathbf{Xa}'||_2^2$, we have $||\mathbf{P}^T\mathbf{y} - \mathbf{P}^T\mathbf{Xa}||_2^2 > ||\mathbf{P}^T\mathbf{y} - \mathbf{P}^T\mathbf{Xa}'||_2^2$. Hence, we can determine $\mathbf{a}$ is also a vector of the same test sample y in the K-L transform space.

The algorithm of FSRC:

Input: Training face image set X, test sample y, and projection number c.

Output:The prediction class of test sample

Step 1. Construct the covariance matrix of the training samples, and get the its c orthonormal eigenvectors $\tilde{\mathbf{P}} = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_c)$ corresponding to the c largest eigen-values. Project the training and test samples onto the low dimensional space by $\tilde{\mathbf{P}}$.

Step2.Get the coefficient vector of y in PCA space: $\mathbf{b} = \arg\min(\| \tilde{\mathbf{P}}^T \mathbf{y} - \tilde{\mathbf{P}}^T \mathbf{X}\mathbf{b} \|_2^2 + \mu \| \mathbf{b} \|_1)$

Step 3. In original space, we calculate the representation error by each class $e_i = \| \mathbf{y} - \mathbf{X}_i \mathbf{b}_i \|_2^2$, where i =1,2,..., t and $\mathbf{b}_i$ is the combination coefficient vector corresponding to $\mathbf{X}_i$

Step 4. The FSRC classifies the sample according to reconstruction error

$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_c$ corresponding to the c largest non-zero eigen-values, where c is the integer smaller than d, play more important role than $\mathbf{p}_{c+1}$, $\mathbf{p}_{c+2}, \ldots,$ $\mathbf{p}_d$ in the K-L transform. We use s_i to denote the eigen-value corresponding to the eigen-vector $\mathbf{p}_i$. In most face recognition problems, $\sum_{i=1}^{c} s_i / \sum_{i=1}^{d} s_i$ is close to 1, and we denote $V_c = \sum_{i=1}^{c} s_i / \sum_{i=1}^{d} s_i$. In most face recognition problems, $V_c = \sum_{i=1}^{c} s_i / \sum_{i=1}^{d} s_i$ is very close to 1, even though c is a very small integer. Denoted by the $m \times c$ matrix $\tilde{\mathbf{P}} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_c)$, we have $||\tilde{\mathbf{P}}^T \mathbf{y} - \tilde{\mathbf{P}}^T \mathbf{X}\mathbf{a}||_2^2$ is very close to $||\mathbf{P}^T \mathbf{y} - \mathbf{P}^T \mathbf{X}\mathbf{a}||_2^2$. Since the l_2 norm of the vector in the K-L transform space is equal to that in original space, i.e. $||\mathbf{y} - \mathbf{X}\mathbf{a}||_2^2 = ||\mathbf{P}^T \mathbf{y} - \mathbf{P}^T \mathbf{X}\mathbf{a}||_2^2$, we have $||\tilde{\mathbf{P}}^T \mathbf{y} - \tilde{\mathbf{P}}^T \mathbf{X}\mathbf{a}||_2^2$ is very close to $||\mathbf{y} - \mathbf{X}\mathbf{a}||_2^2$. Therefore, we can get that the approximate solution of the sparse representation coefficient vector in c-dimensional PCA space, ($c < < d \leq m$).Based on the framework of fast SRC and the analysis of PCA, we proposed the FSRC algorithm.

As we know, the SRC can reconstruct the test sample $\mathbf{y}$ by $\mathbf{Xa}$. Although the sparse representation coefficient vector of FSRC is calculated in the low-dimensional space, FSRC also can reconstruct the test sample with the same dimensionality as that of original space by $\mathbf{Xb}$. Figure 2 shows some examples reconstructed by SRC and FSRC, respectively. In Fig. 2, the first row shows some images from ORL dataset, the second row shows the images reconstructed by SRC, and the third row shows the images reconstructed by FSRC.

From Algorithm 1, we find that the complexity of FSRC consists of three parts, feature extraction using PCA, sparse coding on the low-dimensional space, and calculating the representation error in original space. No matter how many test samples the test set has, the PCA transformation procedure in FSRC is performed only once. Supposing FSRC transforms the samples into c dimensional PCA for getting the sparse

Fig. 2. Some face images and the reconstructed images by SRC and FSRC, respectively.

model, the time complexity of PCA procedure for each test sample can be considered as the O(cm^2/t), where m and t denote dimensionality of the sample and test sample size, respectively. FSRC calculates representation error in original space, and the time complexity of this part is O(mn). The first two columns of Table 1 give the computational complexity comparison between SRC and FSRC. Clearly, FSRC can reduce the computational complexity from O(m^3) to O($c^3 + cm^2/t$), where c is much less than m.

4 Experiments

ORL, FERET, Extended Yale B and AR datasets were used in the experiments [8–11]. On ORL dataset, we randomly choose 5 images from each class for training, and the others are used as test samples. To fairly compare the performance, the experiments are repeated 30 times with 30 randomly possible selections of the training images. On FERET dataset, we randomly select 4 images from each subject for training. The experiments are repeated with 10 randomly possible selections. On the Extended Yale B dataset [10], 5 images of each subject were selected for training. The experiments are also repeated 10 times. The experiments are carried on the above three datasets, respectively. Two state-of-the-art methods CRC and SRC are employed as comparisons. Table 1 shows the classification accuracies of the methods SRC,CRC, PCAcFSRC and FSRC. From the results, we find our method achieves the comparable classification accuracy to SRC and CRC. Table 2 shows that the classification efficiency of FSRC is much higher than SRC on the first three datasets. In PCA+SRC, the representation error is directly calculated in the low-dimensional space. In the proposed method of FSRC, the representation error is calculated in the original space. For revealing the correlation between where the representation error is calculated and the accuracy, we also carried PCA+SRC on the datasets. Compared to the proposed method of FSRC, PCA+SRC always obtains a lower accuracy. In computation, FSRC achieves the higher classification efficiency than SRC. Each image of AR dataset is occupied by the Gaussian noise. The mean is 0 and variance is 0.01. Figure 3 gives the 26 noised images from the first subject.

Table 1. The classification accuracies (%) of the methods on four face datasets

	SRC	CRC	PCA+SRC	FSRC
ORL	92.08	91.72	88.53	92.62
FERET	65.00	58.83	54.53	64.50
Yale B	78.37	78.19	67.40	77.25
AR with noise	93.33	91.27	87.11	92.06

Table 2. The classification time (s) of the methods on four face datasets

	SRC	CRC	PCA+SRC	Random+SRC	Subsample+SRC	FSRC
ORL	1563.4	353.2	683.5	572.9	558.4	734.3
FERET	23752.1	894.3	2135.6	1443.2	1526.7	2556.3
Yale B	45958.0	2375.2	4772.4	3795.2	3419.5	5218.4
AR	21330.5	1372.7	4583.1	2836.8	2583.1	4036.7

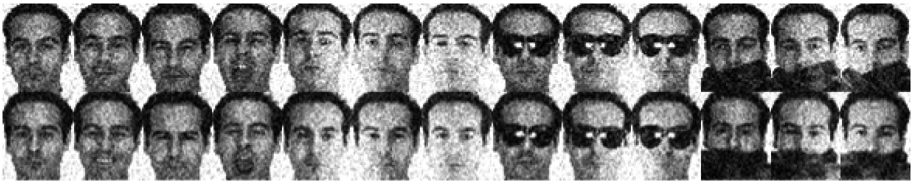

Fig. 3. The noised images from the first subject

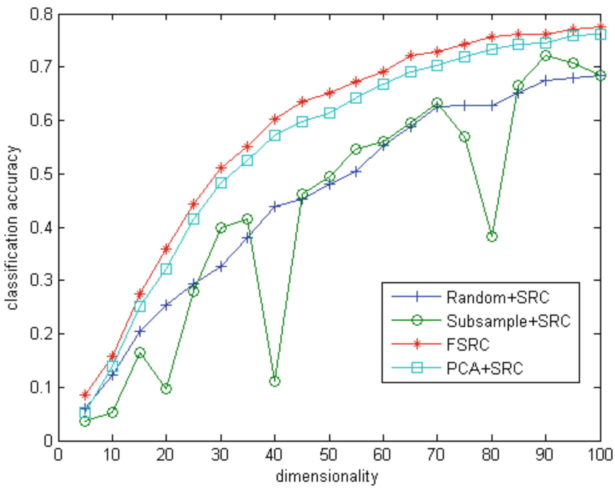

Fig. 4. The classification accuracies of the four fast SRC methods on AR dataset.

The classification accuracies and time of the methods SRC, CRC, PCA+SRC and FSRC on this dataset are shown in the last row of Tables 1 and 2, respectively. As we know, some feature selection or extraction plus SRC based methods, such as PCA +SRC, Random+SRC and Subsample+SRC [6], also have higher efficiency than SRC. Our method is distinctly different from the above methods. Seeing from step 4 of our algorithm, it is clear that all the features are participated in our classification method. PCA+SRC, Random+SRC and Subsample+SRC are also employed in the experiments for comparisons. The classification accuracies of our method, PCA+SRC, Random +SRC and Subsample+SRC, as their dimensionalities range from 5 to 100 with the interval of 5, are shown in Fig. 4. The Subsample+SRC case shows two significant drops in classification accuracy when the dimensionality is 40 and 80. We belive this phenomenon is led by the drawback of subsample method. When extracting the features, subsample method considers the position of the pixel rather than the discriminant information. Some pixels having discriminant information may be lost in sub-sample method. Then the features generated by sub-sample method may be not appropriate to classify using SRC. The results show that FSRC outperforms the other three methods with the same number of dimensionality. Table 2 also shows the running time of these four fast SRC methods on the noised AR dataset. Seeing form Fig. 4 and Table 2, we find that the four fast SRC methods cost the roughly equal running time, and our method achieves the best classification performance.

5 Conclusion

This paper proposed a fast sparse representation classification (FSRC) algorithm for face recognition. FSRC learns the approximate solution of the sparse representation coefficients in a low-dimensional space with a relatively low computation cost. Based on the idea of transfer learning, FSRCreconstructs the test image in original space rather than low-dimensional space. Therefore, FSRC can achieve the comparable accuracy to SRC, and much higher computational efficiency. It is necessary to point that the framework ofFSRC is independent on optimization algorithms. We evaluated the proposed method on four face datasets. Compared with SRC, FSRC is with significantly lower complexity and has very competitive accuracy. Compared with PCA +SRC and the other two SRC based fast classification frameworks, FSRC achieves the best classification results.

Acknowledgements. This article is partly supported by Nature Science Foundation of China under grants(Nos. 61375021, 61472138, 61362031and 61263032), China Postdoctoral Science Foundation funded project (No. 2015M570446), Jiangsu Planned Projects for Postdoctoral Research Funds (No.1402047B), Open Project Program of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education (No. JYB201503), and Natural Science Foundation ofJiangsu Province (No. BK20131365).

References

1. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE TPAMI **31**(2), 1–17 (2009)
2. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Ma, Y.: Towards a practical face recognition system: robust registration and illumination via sparse representation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2009
3. Zhu, Q., Sun, H., Feng, Q.X., Wang, J.H.: CCEDA: building bridge between subspace projection learning and sparse representation-based classification. Electron. Lett. **50**(25), 1919–1921 (2014)
4. Yang, M., Zhang, L., Zhang, D., Wang, S.: Relaxed collaborative representation for pattern classification. In: CVPR (2012)
5. Xu, Y., Zhang, D., Yang, J., Yang, J.-Y.: A two-phase test sample sparse representation method for use with face recognition. IEEE Trans. Circ. Syst. Video Technol. **21**(9), 1255–1262 (2011)
6. Pan, SinnoJialin, Yang, Qiang: A survey on transfer learning. TKDE **22**(10), 1345–1359 (2010)
7. Yang, Qiang, Pan, SinnoJialin, Zheng, Vincent W.: Estimating location using Wi-Fi. IEEE Intell. Syst. **23**(1), 8–13 (2008)
8. http://www.uk.research.att.com/facedataset.html
9. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.: The FERET evaluation methodology for face recognition algorithms. IEEE TPAMI **22**, 1090–1104 (2000)
10. Lee, K., Ho, J., Kriegman, D.: Acquring linear subspaces for face recognition under variable lighting. IEEE TPAMI **27**, 684–698 (2005)
11. Martinez, A., Benavente, R.: The AR face database, The Computer Vision Center, Technical report (1998)