Robust Coarse-to-Fine Sparse Representation for Face Recognition

Yunlian Sun and Massimo Tistarelli

Department of Sciences and Information Technology,
Univeristy of Sassari, 07100 Sassari, Italy
{elaine.sun717,mtista}@gmail.com

Abstract. Recently Sparse Representation-based classification (SRC) has been successfully applied to pattern classification. In this paper, we present a robust Coarse-to-Fine Sparse Representation (CFSR) for face recognition. In the coarse coding phase, the test sample is represented as a linear combination of all the training samples. In the last phase, a number of "nearest neighbors" is determined to represent the test sample to perform classification. CFSR produces the sparseness through the coarse phase, and exploits the local data structure to perform classification in the fine phase. Moreover, this method can make a better classification decision by determining an individual dictionary for each test sample. Extensive experiments on benchmark face databases show that our method has competitive performance in face recognition compared with other state-of-the-art methods.

Keywords: coarse-to-fine, sparse representation, face recognition.

1 Introduction

Over the past few decades, face recognition (FR), has always been a very challenging topic in pattern recognition [1]. This is due to a variety of reasons, including the fact that human faces are non-rigid 3D objects and many variables must be taken into account, including lighting conditions and temporal variations in the face appearance due to aging. As such, the recognition from 2D samples is an ill-conditioned problem which requires further efforts to become well-conditioned.

Towards this end, a number of approaches have been developed. Among them, global transform methods have been widely used in appearance-based FR, which includes principal component analysis (PCA) [2] and linear discriminant analysis (LDA) [3]. These methods usually use the global structure information of the entire training set to produce transform axes. Local transform methods, instead exploit local training samples to compute the transform axes in a more economical way [4–6].

The principles of locality and good approximation, underlying many learning algorithms applied in biometrics, are very important to ensure an accurate representation of the data samples to properly represent a subject. In many computer

A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 171–180, 2013.

vision applications, these principles have been pursued by means of a coarse-to-fine approach where an optimal representation is searched from the general to the particular through a proper sequence of frequency bands. Successful cases have been stereo disparity and optical flow computation [7]. A representative example is the scale-space theory [8] for edge and region detection. Remarkably, the Scale Invariant Feature Transform (SIFT), which has been largely applied in biometrics, is a successful application of the coarse-to-fine approach [9]. The same quest for locality and good approximation has been pursued in face-space analysis [4–6]. We postulate, that the same principles of locality and good approximation can be successfully enforced by means of a coarse-to-fine strategy to build an efficient face representation.

Recently, SR-based Classification (SRC) has demonstrated a good potential for FR [10]. Nonetheless, what makes up the underlying theoretical foundation for SR is still unclear. Zhang et al. [11]proposed the Collaborative Representation (CR) as the key element of SRC for face classification. More importantly, they proposed CR-based Classification with Regularized Least Squares (CRC_RLS), which has significantly less complexity than SRC but leads to very competitive results. In [12] Xu et al. also developed a Two-Phase Test Sample Sparse Representation (TPTSR) method to perform face classification.

In order to preserve both locality and approximation accuracy, in this paper we propose a coarse-to-fine strategy to build a sparse face representation, namely a Coarse-to-Fine Sparse Representation (CFSR). In the coarse phase the test sample is represented as a linear combination of the whole training set, and the classes and the nearest neighbors producing the least representation error are determined. In the fine phase the test sample is represented as a linear combination of only the selected nearest neighbors. The classification is performed by evaluating which class produces the least representation error. In this way, a smaller number of candidates are retained for classification reducing the required processing time.

Similarly to [11, 12] our method employs the l_2-norm to recover the best face representation. Unlike [12], in our approach not only a number of neighbors are selected, but the closest classes are also selected in the coarse phase.

The reminder of the paper is organized as follows. Section 2 briefly reviews SRC. In Section 3 the proposed CFSR is presented. The method is extensively discussed and compared with similar approaches in Section 4. Section 5 presents experimental results and Section 6 draws some conclusions.

2 How SRC Works for Face Recognition

Wright et al. [10] proposed SRC for face recognition. Denote by $A_i \in R^{m \times n}$ the set of training samples of the i^{th} subject class. Let $A = [A_1, A_2, ..., A_L]$ be the dictionary of training samples from L classes. Given a test sample y, The procedures of SRC are summarized as follows.

1) Normalize the columns of A to have unit l_2-norm.
2) Represent over A via l_1-minimization.

$$\hat{\alpha} = \arg \underset{\alpha}{min}\{\|y - A\alpha\|_2 + \lambda\|\alpha\|_1\} \tag{1}$$

where λ is a small positive constant balancing the reconstructed error of y and the sparsity of α.

3) Compute the residuals.

$$e_i(y) = \|y - A_i\hat{\alpha}_i\|_2 \tag{2}$$

where $\hat{\alpha}_i$ is the coefficients vector associated with only the i^{th} class.

4) Perform classfication via

$$identity(y) = \arg \underset{i}{min}\{e_i\} \tag{3}$$

3 Coarse-to-Fine Sparse Representation

To improve the performance of sparse representation for robust FR, we propose here a coarse-to-fine sparse representation (CFSR) scheme for FR, and present the details of CFSR in this section. Let $A = [A_1, A_2, ..., A_L]$ be the dictionary of training samples from L subjects, where A_i is the set of training samples of the i^{th} class. Given a test sample y The main steps of CFSR are as follows:

1) Build dictionary A using all normalized training samples.
2) Represent y over A via l_2-minimization and use the coarse phase to determine S classes best representing y with M nearest neighbors.
3) Build dictionary B using the selected $S \times M$ nearest neighbors.
4) Code y over B via l_2-minimization and use the fine phase to determine the identity of y.

3.1 The Coarse Phase of CFSR

The coarse phase of CFSR seeks to represent y as a linear combination of all the training samples and uses the representation result to identify $S(S < L)$ classes which produce the S greatest contributions in representing y. Differing from SRC in [10], we use the l_2-norm for classification:

$$\hat{\alpha} = \arg \underset{\alpha}{min}\{\|y - A\alpha\|_2 + \lambda\|\alpha\|_2\} \tag{4}$$

where λ is the regularization parameter. This can be solved as:

$$\hat{\alpha} = (A^T A + \lambda I)^{-1} A^T y \tag{5}$$

where I is the identity matrix.

Let $\{x_{1,1}, ..., x_{1,n}, ..., x_{L,1}, ..., x_{L,n}\}$ be the set of all training samples, where $x_{i,j}$ is the j^{th} sample of the i^{th} class. Now we rewrite $y = A\hat{\alpha}$ as follows

$$y \approx \hat{\alpha}_{1,1}x_{1,1} + ... + \hat{\alpha}_{L,n}x_{L,n} \tag{6}$$

where $\alpha = [\alpha_{1,1}, ..., \alpha_{L,n}]$, $\alpha_{i,j}$ is the coefficient of $x_{i,j}$. Note that each training sample makes its own contribution to representing y. We compute the sample specific representation residual $e_{i,j} = \|y - \hat{\alpha}_{i,j} x_{i,j}\|_2$ to evaluate the error of each training sample in representing y.

For each class, we exploit $e_{i,j}$ to identify M training samples that have the M smallest errors and refer to them as the M nearest neighbors of the test sample. Then for the i^{th} class, we take the cumulative residual of the M neighbors as its error, i.e. $e_i = \sum\limits_{j \in M neighbors} e_{i,j}$ in representing y. Next we exploit e_i to identify $S (S < L)$ classes which have the S smallest errors in representing the test sample. We consider that y is from one of the selected S classes. Consequently, CFSR classifies y into only one of the S classes. In other words, we perform a coarse classification of the test sample in this phase.

3.2 The Fine Phase of CFSR

Let $c = \{c_1, ..., c_S\}$ stand for the set of class labels of the determined S classes. Denote by $B = [B_1, B_2, ..., B_S]$ the dictionary of $S \times M$ nearest neighbors from the selected S classes, where B_i is the set of M neighbors of y from the c_i^{th} class. The fine phase of CFSR uses only the $S \times M$ neighbors to represent y and exploits the representation result to classify the test sample. That is, we code y by dictionary B via l_2-minimization. There is

$$\hat{\beta} = \arg \underset{\beta}{min} \{\|y - B\beta\|_2 + \mu \|\beta\|_2\} \tag{7}$$

where μ is the regularization parameter. The fine classification by $\hat{\beta}$ is the same as the classification by $\hat{\alpha}$ in SRC [10] as follows

$$identity(y) = \arg \underset{e_i}{min} \{y - B_i \hat{\beta}_i\} \tag{8}$$

4 Analysis of Coarse-to-Fine Strategy

The l_1-minimization adopted in [10] makes SRC very expensive. In [11] Zhang et al. revealed that it is the collaborative representation (CR) mechanism, but not the l_1-norm sparsity, that plays the essential role for face classification in SRC. Moreover, the l_2-norm was adopted to produce a certain amount of "sparsity" and consequently obtain a more efficient classification.

CFSR shares some differences and similarities with these two methods. They are different in that CFSR uses the linear combination of only a subset of the training samples to represent the test sample. In this regard, we can view CFSR as a local transform method. Local transform methods such as the local Fisher discriminant analysis proposed in [6] can achieve good classification performance by using the local data structure of patterns. Besides, CFSR is able to make a better classification decision by determining an individual dictionary for each test sample, while all the test samples share the same dictionary in SRC and

CRC_RLS. The adopt of l_2-norm in CFSR bears a lower computational complexity as compared to SRC.

CFSR is also similar to TPTSR [12]. This method first identifies Q nearest neighbors for the test sample and then uses these neighbors to represent and classify the test sample. We can consider the first stage of TPTSR as a special Nearest Neighbor (NN) classifier, while the coarse stage of CFSR can be viewed as an improvement to the Nearest Subspace (NS) method. More importantly, we do not use all the training samples in each class, but only a subset of them to represent the test sample.

The selective inclusion of representative samples, performed in the"coarse phase", allows to discard noisy samples (for example with occlusions or expression changes) which may compromise the representation and hence recognition accuracy. This makes the algorithm more robust to variations in facial appearance.

As for any parameterized learning scheme, the representation performance of CFSR is affected by the variability of the free parameters, in this case S and M. Moreover, as the optimal values of S and M depend on the structure of the training dataset, they cannot be determined theoretically. Therefore, as discussed in section 5.3, the optimal values of S and M are determined empirically from the training data.

5 Experimental Results

We tested the performance of CFSR on several representative face databases: Extended Yale B [14], AR [15] and ORL [16]. The competing methods included the original NS [17], SRC [10], CRC_RLS [11] and TPTSR [12]. Before performing the classification, we first applied PCA to reduce the dimensionality. In SRC we choose SPGL1 sparse reconstruction solver for l_1-minimization. The parameters λ and μ in both CFSR and TPTSR are set as 0.001 and 0.001. In CRC_RLS, we also set λ as 0.001. All the experiments were run using MATLAB on a 3.16GHz PC with 4.0GB RAM. In order to compare the performances of the algorithms under the same testing conditions, the reported results were all obtained by running the different algorithms on the same computing platform and on the same datasets, not from the results reported in the original papers.

5.1 Face Recognition

Extended Yale B Database: The Extended Yale B database consists of 2,414 frontal face images of 38 individuals. All the images are cropped and normalized to 54×48. For each subject, we randomly select 20 images for training and used the remaining for test. For feature dimension, we choose 30, 64, 150, 300 and 500. To make a reasonable comparison between CFSR and TPTSR, the numbers of selected nearest neighbors used in these two methods should be the same. That is the value of $S \times M$ should be equal to that of Q. In this experiment, we set M, S and Q as 4, 10 and 40, respectively. Fig. 1 (a) shows the recognition accuracy versus feature dimension by NS, SRC, CRC_RLS, TPTSR and CFSR. It can be seen that CFSR achieves better results than the other methods in all dimensions. With a feature dimension equal to 150, CFSR already achieves about 96.9% accuracy, compared to 94.9% for TPTSR, 95.4% for CRC_RLS, 93.5% for SRC and 95.0% for NS.

AR Database: The AR database consists of over 4,000 frontal images from 126 individuals [15]. For each individual, 26 pictures were taken in two separate sessions. As in [10], we choose a subset of the dataset consisting of 50 male subjects and 50 female subjects. For each subject, we select eight images with various illuminations, expressions and occlusion from Session 1 for training, and use all the thirteen images from Session 2 for test. The images were cropped to 60 × 43. We selected five feature dimensions: 30, 54, 120, 300 and 500. In this experiment, the parameters M, S and Q are set as 6, 50 and 300, respectively. Fig. 2 (a) shows the results. We can see that on this database when the dimension is very low, CFSR performs slightly worse than TPTSR and SRC. For example, when the dimension is 30, the accuracies of CFSR, TPTSR and SRC are 51.2%, 52.1% and 52.5%, respectively. However, increasing the feature dimension, CFSR performs better than all the other methods except for a feature dimension equal to 300. The best recognition accuracies of CFSR, TPTSR, CRC_RLS, SRC and NS are 85.2%, 84.3%, 85.1%, 73.1% and 62.4%, respectively.

ORL Database: The ORL database contains 400 images from 40 subjects, each providing 10 different images. In this experiment, we randomly select 5 images of each subject for training, with the remaining 5 images for test. We compute the accuracies with five dimensions 40, 80, 120, 160 and 200. In this experiment, the value of M, S and Q are respectively set as 3, 5 and 15. The results are illustrated in Fig. 3 (a) We can see that CFSR consistently outperforms all the other methods. With a feature dimension equal to 160, the accuracy for CFSR is 2.5%, 6.5%, 2% and 3% higher than that of TPTSR, CRC_RLS, SRC and NS, respectively. It can be seen also that CRC_RLS performs worse than all the other meth performs worse than all the other methods on the ORL database.

5.2 Computational Complexity

In this section, we compare the running time of CFSR and other competing methods. Eigenfaces of dimensionality of 150, 120 and 120 are respectively used as the input facial features on Extended Yale B, AR and ORL databases. Table. 1 shows the accuracies and the computation time required for different methods.

On Extended Yale B, with $M = 6$ and $S = 5$, CFSR achieves the best accuracy (96.9%), compared with the best accuracy of TPTSR (96.1%), 93.5% for SRC, 95.4% for CRC_RLS and 60% for NS. Although CFSR is much slower than NS and CRC_RLS, its computational speed is 2.51 and 1.37 times faster than TPTSR and SRC (measured on the same computing platform and processing conditions). On the AR database, CFSR achieves a significantly higer accuracy compared to all other methods, and it is faster than both TPTSR and SRC. For the experiments on ORL, TPTSR is slightly faster than CFSR but reports a

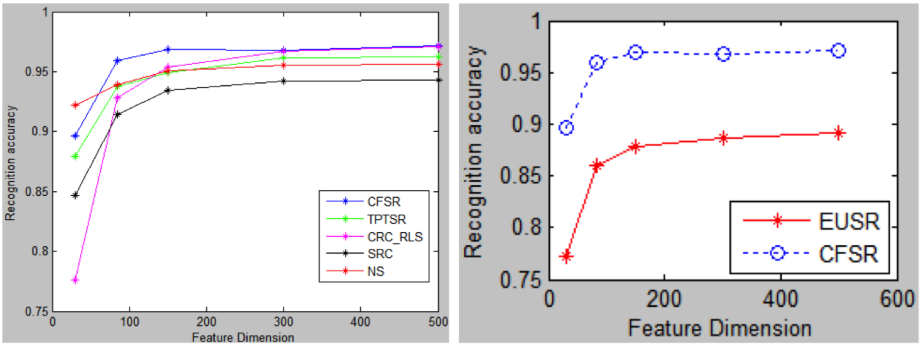

Fig. 1. Results on the Extended Yale B database (a) Comparison with state-of-the-art (b) Comparison with EUSR

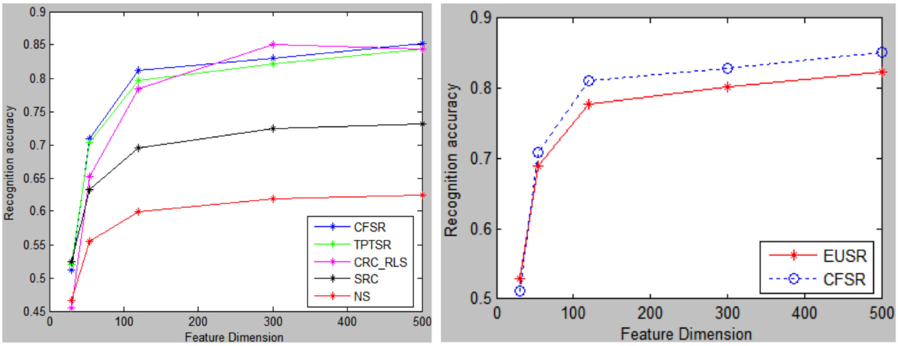

Fig. 2. Results on the AR database (a) Comparison with state-of-the-art (b) Comparison with EUSR

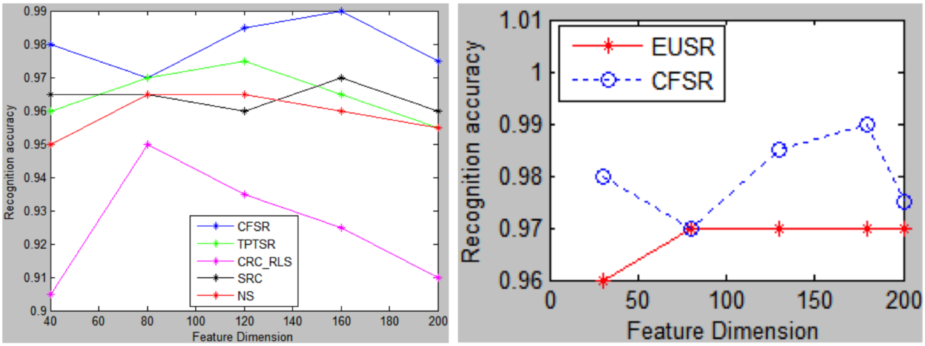

Fig. 3. Results on the ORL database (a) Comparison with state-of-the-art (b) Comparison with EUSR

Table 1. Computed best recognition accuracy and computation time required on the Extended Yale B, AR and ORL databases

Datasets	Extended Yale B		AR		ORL	
Algorithms	Rate(%)	Time(sec)	Rate(%)	Time(sec)	Rate(%)	Time(sec)
NS	95.0	0.0084	60.0	0.0058	96.5	0.0016
CRC_RLS	95.4	0.0018	78.7	0.0042	93.5	0.0008
SRC	93.5	1.203	69.5	1.0453	96.0	0.3044
TPTSR	96.1	0.6546	80.0	0.5615	98.0	0.0157
CFSR	96.9	0.479	81.2	0.5499	98.5	0.0158

lower accuracy. In comparison with SRC, CFSR is 19.27 times faster and reports 2.5% increase in recognition accuracy.

NS and CRC_RLS both code a test sample by a dictionary via l_2-minimization. The dictionary of NS is comprised of training samples from the same class, while the dictionary of CRC_RLS consists of all the training samples. In all the competing methods, NS and CRC_RLS are the two most efficient methods. However, CFSR performs better than both of them in accuracy, especially when the feature dimension is low, CFSR achieves much higher accuracy than CRC_RLS.

5.3 Influence of S and M

The total number of training samples and classes, the distribution of the samples and the local structure of the test samples can all affect the optimal values of S and M. For example, if a test sample is very close to the boundaries of several classes, S and M should be large to capture the complexity of the local structure of the face manifold. On the contrary, if the test sample is close to a class center, S and M can be set to be arbitrarily small. Fig. 4 illustrates the relationship between the recognition accuracy and the values of S and M, as computed on the AR database. As it can be noticed, the accuracy may not be directly related to increasing values of S and M. For example, if $S = 20$ and $M = 4$, the accuracy is equal to 77.9%, while it drops to 75.6% when $S = 20$ and $M = 8$. Nonetheless, despite the complexity of the AR dataset, which includes both changes in facial expression and occlusions, the figure demonstrates a graceful degradation of accuracy when varying the S and M parameters.

5.4 The Rationale for the Coarse Phase

One of the main objectives for the coarse phase is to select the most representative training samples for a test face. An obvious argument is whether it is more practical to directly compute the distance with all other face samples in the training set instead of deducing the data from the sparse representation. In order to test this option we run the same classification tests described in section 5.1, with a modified version of the CFSR algorithm (dubbed EUSR) where, in the coarse phase, the Euclidean distance between the test sample and all samples in the training set is computed to select the nearest neighbors. The fine

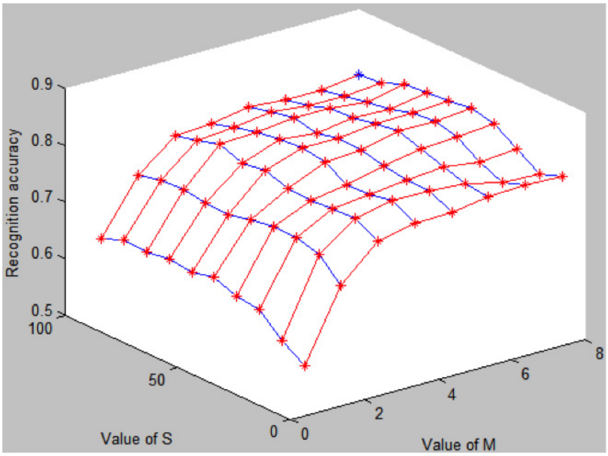

Fig. 4. Relationship between S, M and recognition accuracy on the AR database

phase of EUSR is the same of CFSR. The accuracy versus the values of feature dimension, computed on the three datasets, are reported in Fig. 1 (b), Fig. 2 (b), and Fig. 3 (b).

As it can be noticed, the coarse-to-fine strategy entirely based on the sparse representation consistently achieves the best recognition accuracy, as compared to the algorithm based on the Euclidean distance. This demonstrates the effectiveness of the proposed approach.

6 Conclusion

The large variability of face appearance resulted in the design of a large number of classification algorithms together with several different representations for face classes. The ill-conditioning nature of the face recognition problem makes it very hard to devise a unique methodology which both allows to well represent a large number of classes of individuals and, at the same time, achieving robustness to noise, induced by the data capturing.

This paper proposed a novel method to achieve both accurate representation and robustness to noise through a Coarse-to-Fine Sparse Representation (CFSR)technique. The coarse phase determines an individual dictionary for each test sample by selecting S classes which can produce the least residual when representing the test sample with their corresponding M nearest neighbors. The fine phase seeks to represent each test sample with their corresponding dictionaries and perform the final classification by exploiting the resulting representation. Our approach has been tested on several representative face databases. The experimental results show the competitive performance in comparison with other state-of-the-art techniques.

Acknowledgments. This work has been partially supported by the European Commission COST Action IC1106 and by the Australian ARC Discovery project DP110100827.

References

1. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Computing Surveys 35, 399–458 (2003)
2. Yang, J., Zhang, D., Song, F.X., Yang, J.Y.: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Transactions on PAMI 26, 131–137 (2004)
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces versus fisherfaces: Recognition using class specific linear projection. IEEE Transactions on PAMI 19, 711–720 (1997)
4. Liu, Z.Y., Chiu, K.C., Xu, L.: Improved system for object detection and star/galaxy classification via local subspace analysis. Neural Networks 16, 437–451 (2003)
5. Vural, V., Fung, G., Krishnapuram, B., Dy, J.G., Rao, B.: Using local dependencies within batches to improve large margin classifiers. J. Mach. Learn. Res. 10, 183–206 (2009)
6. Fan, Z.Z., Xu, Y., Zhang, D.: Local linear discriminant analysis framework using sample neighbors. IEEE Transactions on NN 22, 1119–1132 (2011)
7. Grosso, E., Tistarelli, M.: Active/dynamic stereo vision. IEEE Transactions on PAMI 17, 868–879 (1995)
8. Lindeberg, T.: Scale-space theory in computer vision. Kluwer Academic Publishers (1994)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–100 (2004)
10. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on PAMI 31, 210–227 (2009)
11. Zhang, L., Yang, M., Feng, X.C.: Sparse representation or collaborative representation: which helps face recognition? In: Proceedings ICCV 2011, pp. 471–478 (2011)
12. Xu, Y., Zhang, D., Yang, J., Yang, J.Y.: A two-phase test sample sparse representation method for use with face recognition. IEEE Transactions on T-CSVT 21, 1255–1262 (2011)
13. Sugiyama, M.: Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. J. Mach. Learn. Res. 8, 1027–1061 (2007)
14. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on PAMI 27, 684–698 (2005)
15. Martinez, A., Benavente, R.: The AR face database. CVC Tech. Report 24 (1998)
16. `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabasee.html`
17. Ho, J., Yang, M., Lim, J., Lee, K., Kriegman, D.: Clustering appearances of objects under varying illumination conditions. In: Proceedings 2003 IEEE Computer Society Conference on CVPR, pp. 11–18 (2003)