

Analysis of WD Face Dictionary for Sparse Coding Based Face Recognition

Shejin Thavalengal and Anil Kumar Sao

School of Computing and Electrical Engineering
Indian Institute of Technology Mandi, India
s.t@students.iitmandi.ac.in, anil@iitmandi.ac.in

Abstract. This paper deals with the analysis of WD Face dictionary for sparse coding based face recognition. WD (weighted decomposition) Face dictionary emphasizes subject specific unique information of a person. This dictionary has an advantage to adapt to the nature of training images. In the resultant dictionary rows are uncorrelated, which is an essential criterion for dictionary to ensure sparse representation of coefficient vector. The range of sparsity determined by calculating the lower and upper bounds of sparse recovery of coefficient vector for WD Face dictionary exhibits its capability to sparsely represent a test image as a linear combination of training images, even when available training images are small in number. Experimental results solidify our proposal that sparse coding based face recognition with WD Face dictionary is preferable to the existing face recognition techniques.

Keywords: Sparse coding, dictionary, face recognition.

1 Introduction

Face recognition stands as one of the most interesting and promising area in computer vision and pattern recognition, inspired by the ability of human being to recognize and discriminate the faces of his fellow people. Automatic face recognition has been studied extensively for the last two decades owing to its significance in biometrics and other applications [20]. Recently sparse coding based approaches have been introduced and studied in face recognition [5, 18]. These methods allow to represent a test face image of a person (\mathbf{y}_i) as a linear combination of the training images of the same person, as

$$\mathbf{y}_i = \sum_{j=1}^{n_i} a_{i,j} \mathbf{v}_{i,j}, \quad (1)$$

where n_i is the number of training images available for the i^{th} person, $\mathbf{v}_{i,j} \in \mathbb{R}^m$ is the j^{th} training face image of the same person and corresponding coefficient is denoted by $a_{i,j}$. This equation can be rewritten as

$$\mathbf{y} = \mathbf{D}\mathbf{a}, \quad (2)$$

where $\mathbf{D} \in \mathbb{R}^{m \times n}$ denotes dictionary (sensing matrix), which contains all the available n training face images and $\mathbf{a} \in \mathbb{R}^n$ is denoted as coefficient vector, which contains zero

corresponding to those face images in dictionary which do not belong to the same class of the test face image.

The choice of dictionary \mathbf{D} plays a crucial role in any application in sparse coding framework. The available dictionary in literature can be classified in to two classes namely, (i) pre-constructed dictionaries, such as wavelets, random matrix and down sampling matrix and (ii) empirically-learned dictionary such as KSVD, MOD and WD Face dictionaries [7, 16]. The pre-constructed dictionaries are generic and do not exploit the nature of the signal, while empirically learned dictionaries can adapt to the family of signals.

Extensive details and information regarding sparse coding based face recognition and recovery of coefficient vector \mathbf{a} can be found in [2, 5, 16, 18]. Reference [18] has shown that dictionary is not significant if very large number of training images are available. Assuming there are sufficient images available, this method uses severely downsampled face images (DS Face) or face images projected to a random matrix (Randomface) as dictionary. This work is extended to incorporate mis-alignment in reference [17]. Reference [5] attempted to overcome the issue of requirement of large number of training samples by generating an intra-class variant dictionary. This requires prior knowledge of all variations which may be present during face recognition and generation of all possible intra-class variant images. In reference [16] it is shown that when a very large number of training images is not available, dictionaries generated using DS Face and Randomface fail to deliver a good face recognition performance. WD Face dictionary was proposed in reference [16] to overcome this problem by emphasizing subject specific unique information of a person, which helps in discriminating face images of several people. The work presented in this paper focuses on demonstrating the characteristics of WD Face dictionary and how well this dictionary ensures recovery of sparse coefficient vector \mathbf{a} via l_1 minimization.

The experiments carried out in this paper follow the algorithm described in [16]. Results are demonstrated using Extended Yale Face Database B [9], which contains 2432 images of 38 human subjects under 64 illumination conditions [13]. All images are resized to 100×100 prior to any operation. Experiments were conducted using four sets of training face images consisting of 1216 (32 face images per person), 570 (15 images per person), 380 (10 images per person), and 190 (5 images per person) face images. Half of the database is randomly chosen to generate first set of training images (1216 images, 32 images per person). The successive sets are randomly chosen subset of their predecessor. This approach is designed to remove the dependency of results on types of face images to maintain uniform nature in experiments.

Rest of the paper is organized as follows: Section 2 explains the necessary conditions to be satisfied by the dictionary (\mathbf{D}) for good estimation of sparse coefficient vector (\mathbf{a}). The deterministic construction of dictionary for face recognition which ensures sparse signal recovery is explained in Section 3. Evaluation of performance of WD Face dictionary is explained in Section 4, and Section 5 summarizes the paper.

2 Conditions for Sparse Signal Recovery

The fundamental concern in sparse coding based face recognition is how well the sensing matrix (dictionary) \mathbf{D} is created. In reference [16] it was shown that it is necessary

to create a dictionary which emphasizes the subject specific unique information of a person. Moreover, the dictionary \mathbf{D} should ensure the reconstruction of sparse coefficient vector \mathbf{a} . Restricted Isometry Property (RIP) is a sufficient condition for unique and sparse reconstruction of \mathbf{a} via l_1 minimization [3]. The dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$ follows RIP with parameters δ and $2s$, for all $\mathbf{a} \in \mathbb{R}^n$, if

$$\sqrt{1-\delta}\|\mathbf{a}\|_2 \leq \|\mathbf{D}\mathbf{a}\|_2 \leq \sqrt{1+\delta}\|\mathbf{a}\|_2, \quad (3)$$

where $\delta \in (0, 1)$ and $s \in \mathbb{N}$. If $\delta < \sqrt{2} - 1$, then one can define upper bound on sparse recovery as explained in [3].

$$\|\bar{\mathbf{a}} - \mathbf{a}\|_1 \leq 2(1-\rho)^{-1} [\alpha\varepsilon\sqrt{s} + (1+\rho)\|\mathbf{a} - \mathbf{a}^s\|_1], \quad (4)$$

where $\bar{\mathbf{a}} \in \mathbb{R}^n$ is the coefficient vector obtained via l_1 minimization, $\rho = \frac{\sqrt{2\delta}}{1-\delta}$, $\alpha = \frac{2\sqrt{1-\delta}}{1-\delta}$, ε is reconstruction error and $\mathbf{a}^s \in \mathbb{R}^n$ is a vector with largest s absolute values of \mathbf{a} and remaining elements as zero. But it is an NP hard problem to check whether a given matrix satisfies RIP condition. In this context, reference [11] derived a verifiable condition for sparse recovery of a given sensing matrix. A parameter $\hat{\gamma}_s(\mathbf{D}, \beta)$ is defined as

$$\hat{\gamma}_s(\mathbf{D}, \beta) = \max_a \left[\|\mathbf{a}\|_{s,1} - \beta \|\mathbf{D}\mathbf{a}\|_2 : \|\mathbf{a}\|_1 \leq 1 \right], \quad (5)$$

where $0 \leq \beta \leq \infty$ and $\|\mathbf{a}\|_{s,1}$ is the sum of largest s absolute values of \mathbf{a} . If $\hat{\gamma}_s(\mathbf{D}, \beta) \leq \frac{1}{2}$ and $\beta \leq \infty$, then the upper bound on sparse recovery is similar to the one obtained using equation (4), and written as in [11].

$$\|\bar{\mathbf{a}} - \mathbf{a}\|_1 \leq (1 - 2\hat{\gamma}_s(\mathbf{D}, \beta))^{-1} [2\beta\varepsilon + 2\|\mathbf{a} - \mathbf{a}^s\|_1 + \nu], \quad (6)$$

where ε denotes reconstruction error and ν is the inaccuracy in solving equation (6).

This method allows to obtain upper bound and lower bound of sparsity of the given dictionary. The lower bound is computed using linear programming of $(2n^2 + n) \times (n(m + n + 1))$ constrained matrix and upper bound is computed from lower bound by sequential convex approximation [11].

3 Analysis of WD Face Dictionary

The transformed face images such as Down-Sampling face (DS Face) where the face image is downsampled to 15×12 or Randomfaces where a face image is projected in to a random matrix of size $350 \times m$ whose values are drawn from independent and identically distributed Gaussian (where the image of size $\sqrt{m} \times \sqrt{m}$ is converted to a vector of $m \times 1$ before projection) can be used as dictionary for sparse coding based face recognition [18]. Even though the above mentioned random projection guarantees sparse recovery with high probability [4, 6], its performance in face recognition is not good since Randomfaces does not emphasize subject specific unique information of face image [16].

Dictionary derived using Weighted Decomposition face (WD Face) representation is given as

$$\hat{\mathbf{D}} = \mathbf{W}\Psi^T\mathbf{D}, \quad (7)$$

where $\hat{\mathbf{D}}$ is the derived dictionary, \mathbf{W} is a weight matrix and Ψ is the matrix consisting of eigenvectors obtained from the eigen analysis of the set of training images. WD Face is generated by assigning proper weights to those underlying components of a face image which emphasize subject specific unique information of the person. A physical inter-

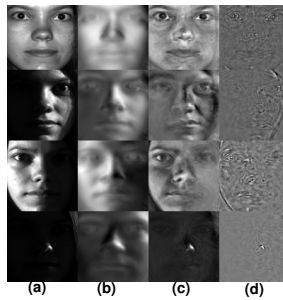


Fig. 1. Illustration of face decomposition in to different components. (a) Gray level image. The reconstructed face image using (b) first 10 eigenvectors, (c) 11-350 eigenvectors, (d) 351-remaining eigenvectors.

pretation of WD Face is given based on the assumption that a face image can be decomposed into three components [12] as shown in Fig. 1. The first component (Fig. 1(b)) corresponds to a component which is common to all the face images. The third component (Fig. 1(d)) can be considered as a noise component which contains negligible information about the identity of the person. The second component (Fig. 1(c)) carries all the person specific unique information which accounts for the discrimination of a person from the other. These components are derived from the eigenvectors of the covariance matrix of the dictionary \mathbf{D} . These eigenvectors are arranged according to the descending order of corresponding eigenvalues. Fig. 1(b) depicts the reconstruction of Fig. 1(a) using 10 eigenvectors which corresponds to the 10 largest eigenvalues. In a similar manner Fig. 1(c) and Fig. 1(d) are reconstructed using 11-350 eigenvectors and 351-10000 eigenvectors, respectively. WD Faces are generated by giving higher weightage to the component shown in Fig. 1(c). This can be done with the weight matrix $\mathbf{W} = \Lambda^{-1/2}$ where Λ is a diagonal matrix containing the eigenvalues corresponding to the eigenvector matrix Ψ . One has to take care of the situation where eigenvalues are very small (close to zero), which results in infinitely large weights for the less significant eigen components. This can be done by removing the least significant eigen components with the help of a thresholding operator. The eigen analysis incorporated in the computation of WD Face dictionary requires all the training images to be well-aligned [15].

Apart from just giving higher weights to more discriminative features of face image, this transformation projects the face images to an orthogonal subspace. This transformation take advantage of the nature of signal. Moreover, we can see that

$$E\{\hat{\mathbf{D}}\hat{\mathbf{D}}^T\} = \mathbf{W}\Psi^T E\{\mathbf{D}\mathbf{D}^T\}(\mathbf{W}\Psi^T)^T, \quad (8)$$

where $E\{\cdot\}$ denotes the expectation operator. We can decompose $E\{\mathbf{D}\mathbf{D}^T\}$, which is the covariance matrix of all the training face images, in terms of eigenvectors and eigenvalues as $\Psi\Lambda\Psi^T$. Then equation (8) can be rewritten as

$$\begin{aligned} E\{\hat{\mathbf{D}}\hat{\mathbf{D}}^T\} &= \mathbf{W}\Psi^T\Psi\Lambda\Psi^T\Psi\mathbf{W}^T \\ &= \Lambda^{-\frac{1}{2}}\Psi^T\Psi\Lambda\Psi^T\Psi\Lambda^{-\frac{1}{2}} \\ &= \mathbf{I}, \end{aligned} \quad (9)$$

where \mathbf{I} is an identity matrix. That is, WD transform makes the rows of derived dictionary uncorrelated [10], which is one of the criteria for dictionary to ensure the sparse recovery of coefficient vector. One can observe the advantage of WD Face representation from Fig. 2.

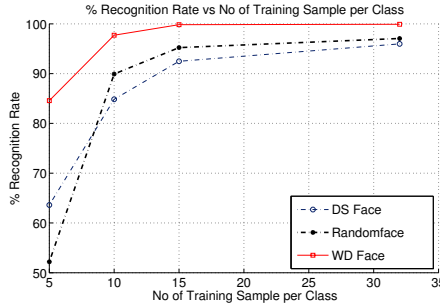


Fig. 2. Face recognition performance for different dictionaries on Extended Yale Face Database B as a function of number of training images per class

In addition, lower and upper bounds on sparse recovery (equations (5) and (6)) can be used to analyze the statistical RIP to check the efficiency of dictionary. The lower and upper bounds computed in this fashion are shown in Fig. 3. It has to be noted that the bounds are not calculated for the case where half of the data is used for training and remaining for testing (32 images per class, total 1216 Images for training), since it will result in solving linear programming of $(2n^2 + n) \times (n(m + n + 1)) = 5.6373 \times 10^{12}$ constrained matrix which is computationally intensive. Fig. 3(a) compares the lower bound for different dictionaries generated for sparse coding based face recognition. In the worst case scenario, where there are just five training images per class, the dictionaries generated using DS Face and Randomfaces return a lower bound of 10 and 17

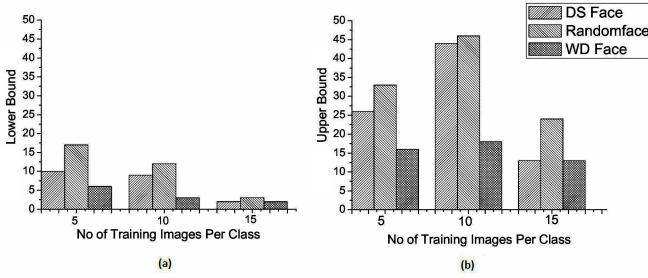


Fig. 3. Illustration of (a) lower bound and (b) upper bound on sparse recovery for different dictionaries as function of number of training images per class

respectively. It is higher than the number of actual available training images per class. Hence, these representations do not ensure a sparse recovery of the coefficient vector \mathbf{a} with five coefficients, which explains their poor performance (Fig. 2). On the other hand, dictionary generated using WD Face provides a lower bound of 6 which is very close to the number of training images available for each class. Hence the probability to recover sparse coefficient vector \mathbf{a} is higher as compared to vectors generated using DS Face and Randomface, which justifies the better performance of face recognition using WD Face based dictionaries. We can notice that as the number of training samples increases, the lower bound of sparsity is decreasing for all the dictionaries. This is because of the availability of new training faces which can represent the corresponding class efficiently. When number of training samples reaches 15 or more images per class, the lower bound of sparsity remains more or less the same for all the dictionaries, which explains the claim of reference [18] that when 'sufficient' training images are available, dictionary is no more significant.

Fig. 3(b) compares the upper bound of sparsity derived for the different dictionaries. The lower bound and upper bound together define the range of sparsity for the given dictionaries. It is proposed that the upper bound should not be very large as compared to the lower bound, and hence the range of sparsity. As the range of sparsity is small, it will help the recovery algorithm to ensure recovery of sparse coefficient vector \mathbf{a} . Here, in the case of WD Face dictionary, we can note that the upper bound is 15 and lower bound is 6 for 5 training images per person. While, upper bound and lower bounds for DS Face dictionary are 26 and 10; and for Randomface dictionary are 32 and 16 respectively. Hence, the range of sparsity for WD Face dictionary is smaller as compared to the that for DS Face dictionary and Randomface dictionary. Also, as the number of training samples increases, the range of sparsity becomes narrower for all the three dictionaries because of the availability of new training faces which can represent the corresponding class efficiently. From Fig. 3, we can conclude that the dictionary generated using WD Faces is superior over its counterparts in sparse coding based face recognition, since it ensures the sparse recovery of the coefficient vector even when 'sufficient' training samples are not available and the difference between upper bound and lower bound is small as compared to its counterparts.



Fig. 4. Face recognition performance for different dictionaries on CUBiC FacePix Database as a function of number of training images per class

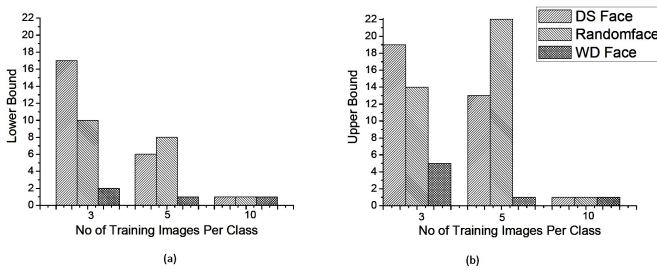


Fig. 5. Illustration of (a) lower bound and (b) upper bound on sparse recovery for different dictionaries as function of number of training images per class on CUBiC FacePix Database

4 Experiments and Results

In addition, analysis of WD Face dictionary is carried out on CUBiC FacePix database [1, 14]. A subset of CUBiC FacePix database corresponding to frontal face images captured with a point light source at various illumination angles without an ambient light source is used for our experiments. This subset consists of a total number of 5430 images of 30 human subjects. Three sets of training images are chosen in the same manner as explained in the previous section. Performance of various dictionaries is compared in Fig. 4. The lower bound and upper bound for these dictionaries are shown in Fig. 5. The results validate our claim that dictionary created using WD Face is superior to Randomfaces and DS Faces especially when the number of training images is very small.

In order to compare the results obtained by WD Face dictionary in sparse coding based face recognition with other studies on these databases, we extended our experiments as in references [8] and [1]. Table 1 compares the performance of WD Face dictionary with different classifiers as in [18] and [8]. These experiments were carried out in Extended Yale Face Database B, where half of the database was used for training and remaining half for testing. Face recognition performance of different methods with feature space dimensions 30, 56, 120, and 504, respectively is shown. We have to note that maximum number of valid Fisher faces is $k - 1$, where k refers total number of

Table 1. Face Recognition Results for Different Methods in Extended Yale Face Database B

Dimension	30	56	120	504
Eigen+NS	89.9%	91.1%	92.5%	93.2%
Eigen+SVM	70.6%	84.3%	93.1%	96.8%
Eigen+CRC [19]	67.97%	85.62%	94.41%	98.55%
Eigen+KSR-Polynomial [8]	91.74%	95.75%	97.42%	98.28%
Eigen+KSR-Gaussian [8]	89.01%	94.42%	97.49%	99.16%
Laplacian+NS	89.0%	90.4%	91.9%	93.4%
Laplacian+SVM	72.0%	85.0%	94.0%	97.7%
Laplacian+CRC [19]	72.25%	86.89%	94.54%	97.91%
Laplacian+KSR-Polynomial [8]	92.25%	95.14%	97.07%	99.11%
Laplacian+KSR-Gaussian [8]	88.86%	94.24%	97.11%	98.12%
DS Face+NS	80.8%	88.2%	91.1%	93.4%
DS Face+SVM	48.9%	69.5%	79.0%	91.6%
DS Face+CRC [19]	67.97%	82.84%	92.93%	97.23%
DS Face+KSR-Polynomial [8]	87.20%	93.09%	95.81%	97.47%
DS Face+KSR-Gaussian [8]	83.57%	91.65%	95.31%	97.80%
Fisher+NS	81.9%	NA	NA	NA
Fisher+SVM	86.7%	NA	NA	NA
Fisher+CRC [19]	68.10%	NA	NA	NA
Fisher+KSR-Polynomial [8]	91.74%	NA	NA	NA
Fisher+KSR-Gaussian [8]	88.93%	NA	NA	NA
WD Face+Sparse Coding	88.90%	96.05%	98.44%	99.91%

classes in the database, which is 38 in Extended Yale Face Database B. Hence recognition performance for the methods using Fisher face is limited to 30 dimensions. It can be noted that, sparse coding based face recognition, which incorporates the proposed WD Face dictionary in sparse coding based face recognition consistently outperforms all the other approaches (even the most recent kernel sparse representation method [8]) for all the dimensions except for the case of 30 dimensions. The drop-off in performance of WD Face dictionary may be because 30 dimensions won't be sufficient to represent person specific unique information.

Table 2. Face Recognition Results for Different Methods in FacePix Database

Number of Training Images	3	5
PCA	71.71%	90.33%
LDA	79.52%	94.92%
HMM	37.38%	59.37%
BIC	79.10%	93.54%
WD Face+Sparse Coding	93.18%	98.24%

Similarly, Table 2 shows the comparison of our method against existing approaches in CUbiC FacePix database as in [1]. Two different sizes of training image set are used; the first one contains 3 images per class for training (total 90 images) and the second set contains 5 images per class for training (total 150 images). We can observe from the table that, in both cases sparse coding based face recognition with WD Face dictionary significantly outperforms the other methods.

5 Summary

In this paper we analyzed the properties of WD Face dictionary for sparse coding based face recognition. WD Face dictionary, which is generated using WD face representation reflects person specific unique information, which is critical in face recognition. The WD Face transform makes the rows of the derived dictionary uncorrelated. Moreover, WD Face dictionary has the ability to adapt to the training face data. These properties of the dictionary make it a potential candidate for use in sparse coding based face recognition. Upper and lower bounds for sparsity computed experimentally serve to strengthen our proposal that a dictionary generated in this fashion ensures sparse representation of test image as a linear combination of training images even when training images are very small in number. Further experiments strengthen the proposal that sparse coding based face recognition using WD Face dictionary is favourable as compared to the existing approaches.

References

1. Black Jr., J.A., Gargesha, M., Kahol, K., Kuchi, P., Panchanathan, S.: A framework for performance evaluation of face recognition algorithms. *ITCOM, Internet Multimedia Systems II*, 163–174 (July 2002)
2. Candes, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Information Theory* 52(2), 489–509 (2006)
3. Candes, E.J.: The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique* 346(910), 589–592 (2008)
4. Candes, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* 59(8), 1207–1223 (2006)
5. Deng, W., Hu, J., Guo, J.: Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 34(9), 1864–1870 (2012)
6. Donoho, D.L.: Compressed sensing. *IEEE Trans. on Information Theory* 52(4), 1289–1306 (2006)
7. Elad, M.: *Sparse and Redundant Representations - From Theory to Applications in Signal and Image Processing*. Springer, New York (2010)
8. Gao, S., Tsang, I.W.-H., Chia, L.-T.: Sparse representation with kernels. *IEEE Trans. on Image Processing* 22(2), 423–434 (2013)
9. Georgiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 23(6), 643–660 (2001)

10. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley (2001)
11. Juditsky, A., Nemirovski, A.: On verifiable sufficient conditions for sparse signal recovery via l_1 minimization. *Mathematical Programming* 127, 57–88 (2011)
12. Sao, A.K., Yegnanarayana, B.: Analytic phase-based representation for face recognition. In: Seventh International Conference on Advances in Pattern Recognition, Kolkata, India, pp. 453–456 (February 2009)
13. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27(5), 684–698 (2005)
14. Little, D., Krishna, S., Black, J., Panchanathan, S.: A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, USA, vol. 2, pp. 89–92 (March 2005)
15. Shejin, T., Sao, A.K.: Dictionary for sparse coding based pose invariant face recognition. unpublished
16. Shejin, T., Sao, A.K.: Significance of dictionary for sparse coding based face recognition. In: 11th International Conference of the Biometrics Special Interest Group, pp. 1–6 (2012)
17. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., Ma, Y.: Towards a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 34(2), 372–386 (2012)
18. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 31(2), 210–227 (2009)
19. Zhang, L., Yang, M., Feng, X.: Sparse representation or collaborative representation: Which helps face recognition? In: 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, pp. 471–478 (November 2011)
20. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Computer Survey* 35(4), 399–458 (2003)