Significance of Dictionary for Sparse Coding Based Face Recognition

Shejin T, Anil Kumar Sao

School of Computing and Electrical Engineering
Indian Institute of Technology Mandi
Mandi, Himachal Pradesh - 175001
India
s_t@students.iitmandi.ac.in
anil@iitmandi.ac.in

Abstract: Sparse representation based classification (SRC) successfully addresses the problem of face recognition under various illumination and occlusion conditions, if sufficient training images are given. This paper discusses the significance of dictionary in sparse coding based face recognition. We primarily address the problem of sufficiency of training data in various illumination conditions. The dictionary is generated using a lower dimensional representation of image, which emphasizes the subject specific unique information of the face image. This representation is called weighted decomposition (WD) face image, because it attempts to give more weightage to unique information of face image. The effect of illumination in computation of WD face image is reduced using edginess based representation of image, which is derived using one-dimensional (1-D) processing of image. 1-D processing provides multiple partial evidences, which are combined to enhance the face recognition performance. The experimental results suggest that the proposed approach addresses the issue of sufficiency of training data efficiently.

1 Introduction

Automatic face recognition remains as one of the challenging problems in pattern recognition due to large variations in illumination, pose and expressions during image acquisition [GBMK11, BCF06]. In recent years sparse coding has been widely studied in different applications of pattern recognition including face recognition [WMM+10, DHG12]. The experimental results show that the sparse coding based face recognition performs better than most of the existing approaches of face recognition such as Nearest Neighbor (NN), Nearest Subspace (NS) and Linear Support Vector Machine (SVM) classifiers [WYG+09]. In sparse coding based face recognition a face image $\mathbf{y} \in \mathbb{R}^m$ of ith person is represented as the linear combination of the training face images associated with the same person. That is

$$\mathbf{y} = \sum_{j=1}^{n_i} a_{i,j}\mathbf{v}_{i,j}, \tag{1}$$

where $\mathbf{v}_{i,j} \in \mathbb{R}^m$ is the jth training face image of ith person and corresponding coefficient is denoted by $a_{i,j}$. Equation 1 can be rewritten as

$$\mathbf{y} = \mathbf{Da}, \tag{2}$$

where $\mathbf{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_k]$, also known as the dictionary, denotes the training face images of all the k persons (available in the database). $\mathbf{D}_i = [\mathbf{v}_{i,n_1}, \mathbf{v}_{i,n_2}, \ldots, \mathbf{v}_{i,n_i}]$ denotes all the training face images of ith person and total number of training face images of the same person is denoted by n_i. The coefficient vector is denoted using $\mathbf{a} = [0, \ldots, a_{i,1}, a_{i,2}, \ldots, a_{i,n_i}, 0, \ldots, 0] \in \mathbb{R}^n$. Where n refers to total number of training face images, i.e., $n = \sum_{i=1}^{k} n_i$. For a given $\mathbf{y}$ and $\mathbf{D}$, several approaches have been proposed in the literature to compute coefficient vector $\mathbf{a}$ using sparse coding approximation [CRT06, TG07, KKL$^+$07]. In practice, the coefficients that correspond to the false class in Equation 2 will be relatively small (close to zero) as compared to the coefficients that correspond to true class. Hence, the identity of a given test face image can be computed using the derived coefficient vector $\mathbf{a}$.

One of the necessary conditions for working on sparse coding based face recognition is under-determined condition of matrix ($\mathbf{D}$), i.e., the number of columns should be more than number of rows [Don06]. Thus, if we use gray level representation of 30×30 face image, then we need at least 900 training face images, which is a difficult criterion to satisfy in practical face recognition. This issue is addressed by using transformed face image as

$$\hat{\mathbf{y}} = \mathbf{Ty}, \tag{3}$$

where $\hat{\mathbf{y}}$ is lower dimensional representation of original face image $\mathbf{y}$ and $\mathbf{T}$ is a transform operator. Several operators are proposed in the literature such as down-sampling, K-L transform operator, discrete cosine transform (DCT) etc. [TP91, AGRA08]. Reference [WYG$^+$09] shows that the performance of the face recognition using random matrix as transform operator is comparable to any other operator. As discussed in the literature, the K-L transform [TP91] (where transform operator consists of eigenvectors corresponding to few largest eigenvalues) does not emphasize the class specific unique information of face image, which is very important in the case of face recognition.

In this work we propose the use of a representation which emphasizes the subject specific unique information from face image to form dictionary for sparse coding based face recognition. In addition, edginess based representation is used to overcome the issue of illumination sensitivity [SYK07]. The edginess based representation is computed using one dimensional (1-D) processing of images, which gives multiple partial evidences for a given face image [SYK07]. Each evidence is used separately in sparse coding based recognition, and combined suitably to improve the performance of the face recognition. The experimental results demonstrated on Extended Yale Face Database B [GBK01] quantify our proposal. The cropped version of Extended Yale Face Database B contains 2432 images of 38 human subjects under 64 illumination conditions [LHK05]. All images are resized to 100×100 prior to any operation.

The organization of the paper is as follows: Section 2 explains the use of subject specific unique representation for sparse coding based face recognition. The dictionary is further

modified by incorporating edginess based representation, which is explained in Section 3. The comparison of different representation schemes is discussed in Section 4. Section 5 gives the summary of the proposed approach.

2 Dictionary Using Subject Specific Unique Information

In any task that involves sparse coding approach, dictionary plays a significant role in performance of the task. As mentioned earlier in sparse coding based face recognition, transformed face images derived from Equation 3 can be used to form dictionary. Figure 1 shows the transformed face images using operators as down-sampling matrix and random matrix. These images are called Down-Sampling face (DS Face) and Randomface respectively. Here, we used the image down-sampled to 15×12 and random matrix of size $350 \times m$ (Note that the image of size $\sqrt{m} \times \sqrt{m}$ is converted to a vector of $m \times 1$ prior to projecting in to the random matrix) as transform operators. The results show that the transformed images do not contain much person specific unique information, which could play crucial role in discriminating face images of two persons. The subject specific unique information can be emphasized using a representation as explained in [KSY09]. In this approach, a given face image is divided into three components- a common component, noise component and a component which gives subject specific unique feature of the face image. These three components are shown in Figure 2 . These components are computed using different eigenvectors which are derived from covariance matrix of several face images. The eigenvectors are arranged according to the descending order of corresponding eigenvalues. Figure 2(b) shows the reconstruction of Figure 2(a) using first 10 eigen vectors. Similarly Figure 2(c) and 2(d) are reconstructed using 11-350 eigenvectors and 351-10000 eigenvectors respectively. The eigenvectors corresponding to the largest eigenvalues give information common to all the training samples. On the other hand, the noise or unwanted information is present in the eigenvectors corresponding to small eigenvalues. The unique information of a face image can be highlighted by removing the common and noisy components of the given face image. It can be observed from Figure 1 and 2 that the transformed face image shown in Figure 2(c) has perceptually more discriminating information as compared to Figures 1(b) and 1(c). Thus, the following transform operator is used to emphasize the class specific unique information of the given image $\mathbf{v}_i$. That is

$$\mathbf{v}_i^d = \mathbf{W}\mathbf{\Psi}^T\mathbf{v}_i, \tag{4}$$

where $\mathbf{W} = diag\left\{\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_N}}\right\}$ is the weight matrix, $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ are the eigenvalues corresponding to the eigenvectors, $\mathbf{\Psi}$ is the matrix consisting of eigenvectors. This transformed face image obtained from Equation 4 is denoted as WD Face (Weighted Decomposition of Face, $\mathbf{v}_i^d$) and is used to form the dictionary.

The results of face recognition obtained using different choice of dictionaries for different sets of training face images of Extended Yale Face Database B are shown in Figure 3.

Figure 1: (a) Gray level image, (b) DS Face of Figure 1(a) - Face images down-sampled to 15×12, (c) Randomface of Figure 1(a) - Face images projected to a normalized random matrix of size 350×10000.
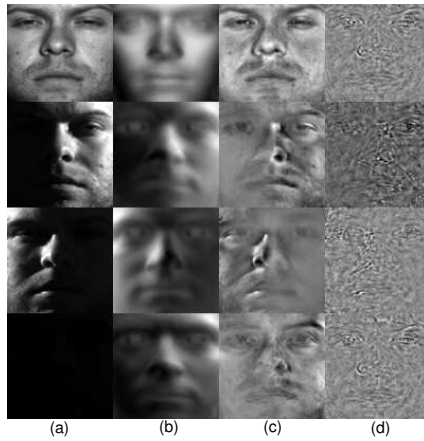

Figure 2: Illustration of subject specific unique information derived using eigen decomposition. (a) Gray level image. The reconstructed face image using (b) first 10 eigenvectors, (c) 11-350 eigenvectors, (d) 351-10000 eigenvectors.

Table 1: Sparse Coding Based Face Recognition

Algorithm for sparse coding based face recognition
Step1: Initialize training images as column vectors of dictionary $\mathbf{D} \in \mathbb{R}^{m \times n}$ for k subjects, test image $\mathbf{y} \in \mathbb{R}^m$ as a column vector, transform operator $\mathbf{T} \in \mathbb{R}^{r \times m}$, threshold ϵ
Step2: Normalize $\mathbf{y}$ and columns of $\mathbf{D}$ to unit length
Step3: Transform $\hat{\mathbf{y}} = \mathbf{Ty}$ and $\hat{\mathbf{D}} = \mathbf{TD}$
Step4: Solve min $\|\mathbf{a}\|_1$ subject to $\|\hat{\mathbf{y}} - \hat{\mathbf{D}}\mathbf{a}\|_2 \leq \epsilon$
Step5: Compute residual $r_i(\mathbf{y}) = \|\hat{\mathbf{y}} - \hat{\mathbf{D}}\delta_{\mathbf{i}}(\mathbf{a})\|_2$ for $i = 1, 2, ...k$, where $\delta_i(\mathbf{a}) \in \mathbb{R}^n$ is a vector whose non zero entries are the entries in $\mathbf{a}$ that associated with class i
$\mathrm{I}(\mathbf{y}) = \arg\min\limits_{i} r_i(\mathbf{y})$, where $\mathrm{I}(\mathbf{y})$ is the identity of the given test image $\mathbf{y}$

Experiments were conducted using five 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. The algorithm for sparse coding based face recognition is given in Table 1. The results are compared using four choices of transform operators: DS (Down-sampled face image), randomface, eigenface and WD face (Equation 4). The size of transform operator is chosen in such a way that the dimension of transformed face image is 350, thus the condition of under determined matrix can be realized. This parameter is changed to 185 in the case of 190 training face images. The discrimination ability of WD face representation can be seen from Figure 3 for a different set of training face images.

In addition, as the number of training samples per class is decreased, the reduction in performance of WD based face representation is much less as compared to the other three representations.

3   Dictionary Using Edginess Based Representation

The transformed face image, which emphasizes the subject specific unique information, is computed using gray level values of the image. It is conjectured that gray level values of an image are sensitive to illumination variations. This issue can be addressed by using a representation that is less sensitive to illumination variations such as, Gabor feature [VS06], Haar transform based feature [PLY+09] and edginess based representations [SYK07]. In this work, we have explored edginess based representation to compute the subject specific unique information of a given face image. The edginess based representation is computed using one dimensional (1-D) processing of image, which gives the edge evidences along different directions [SYK07]. Figure 4 shows the edge evidences for a given face image, for four different directions $\theta = 0°, 45°, 90°$ and $135°$. The WD transformed image (Equation 4) for each partial evidence is computed separately and is shown in Figure 5.

Figure 3: Face recognition performance for different methods as a function of number of training samples per class
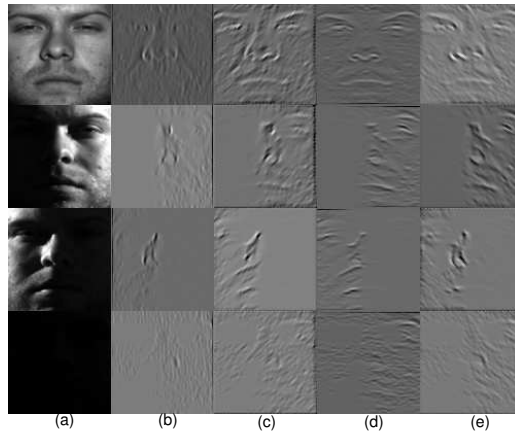

Figure 4: (a) Gray level image. Edginess based representation of face images shown in Figure 4(a) using (b) $\theta = 0°$,(c) $\theta = 45°$, (d) $\theta = 90°$ and (e) $\theta = 135°$.
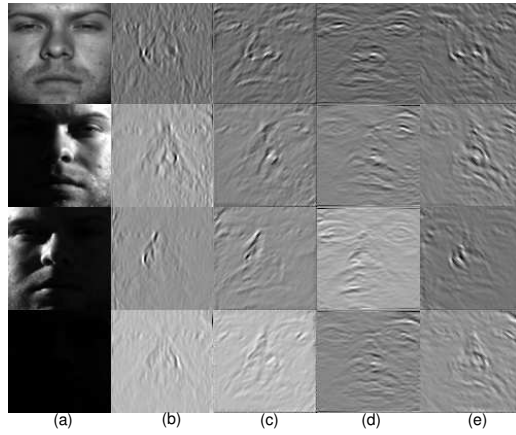
Figure 5: (a) Gray level image. Weighted decomposition of edginess based face representation using 11-350 eigen components and (b) $\theta = 0°$ (c) $\theta = 45°$ (d) $\theta = 90°$ (e) $\theta = 135°$.
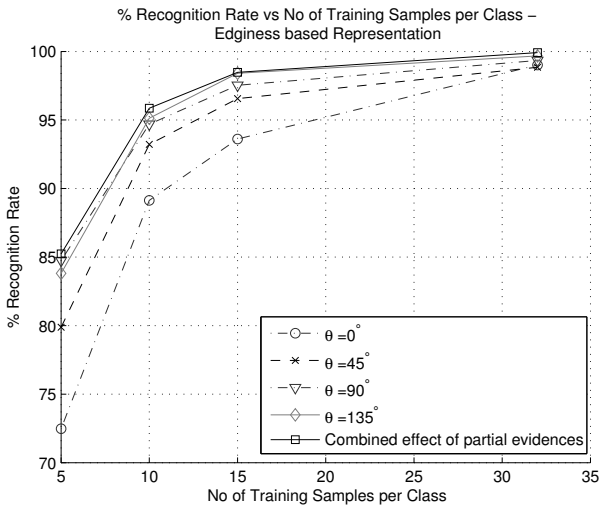

Figure 6: Face recognition performance of WD edginess based representation as a function of number of training samples per class
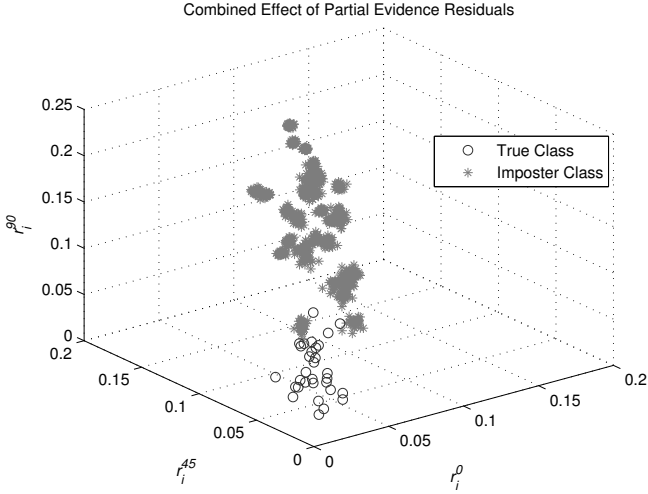
Figure 7: Combined effect of partial evidence residuals for discrimination

Each evidence gives different information of face image; and hence, is used separately for sparse coding based face recognition. The experiments were repeated as explained in the previous section for different sets of training face images using transformed face images of each partial evidence. Figure 6 compares the performance of face recognition for all the partial evidences. The recognition rates obtained using evidences for the directions $\theta = 90°, 135°$ are better than $\theta = 0°, 45°$ since the processing along these directions results in representations which emphasize more facial features. One can also observe that each evidence gives the complimentary information of given face image. Thus the performance can be improved further by combining the residual error obtained from each partial evidence. Let r_i^θ denote the residual error obtained using edge evidence along θ direction of face image $\mathbf{y}$ in algorithm mentioned in Table I. The residual errors for different values of θ are combined as follows

$$P_i(\mathbf{y}) = \left\{ \sum_\theta \left(\mathbf{r_i^\theta}(\mathbf{y}) \right)^{\mathbf{q}} \right\}^{\frac{1}{q}} \tag{5}$$

For $i = 1, 2..., k$ where k is the number of classes, $\theta \in [0, 45, 90, 135]$ and q is a constant parameter. The identity of the person can be established as

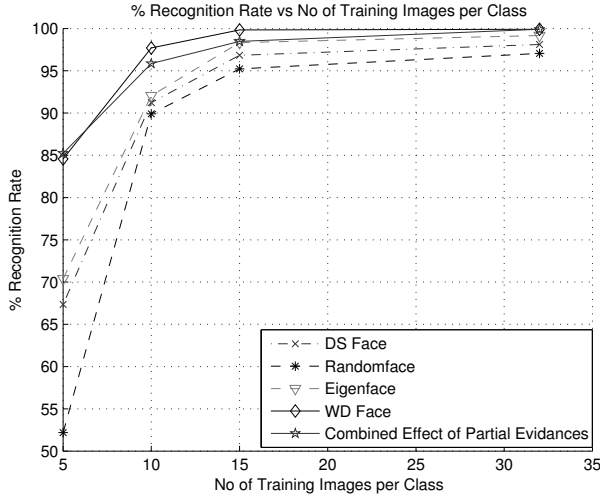$$\mathrm{I}(\mathbf{y}) = \arg \min_i P_i \tag{6}$$

252

Figure 8: Comparison of face recognition performance between different face representation schemes and combined effect of partial evidences.

If we choose a very high value of q ($q \gg 1$), then the weightage to larger residuals (among the residuals for different θ) will be high in the combined evidences, which is not desirable in the proposed approach, because the combined evidences should be small for true class and high for imposter classes. On the other hand, for very small value of q, the weightage to smaller residuals will be high and combined residual value will become low for true class and high for imposter classes. We experimentally found that $q = 0.2$ gives the best performance for all sizes of dictionary. From Figure 6 we can observe that combining information obtained from all the partial evidences and classifying according to Equation 6 outperforms the recognition rates obtained by the edginess representation along individual evidence. This is because, by combining the residuals obtained for each partial evidence, we are projecting the face images to a higher dimensional space and ascertaining higher degree of freedom, where true class can be discriminated from imposter classes easily. This is depicted in Figure 7 for the dictionary of 1216 training images. In this figure, we have shown the results using only three partial evidences ($\theta = 0°, 45°$ and $, 90°$) for visualization purpose. Here, a face image can be represented as a point in the three dimensional space spanned by the residuals along three different direction of partial evidence. We can observe that true class lies around origin and can be easily discriminated from imposter classes.
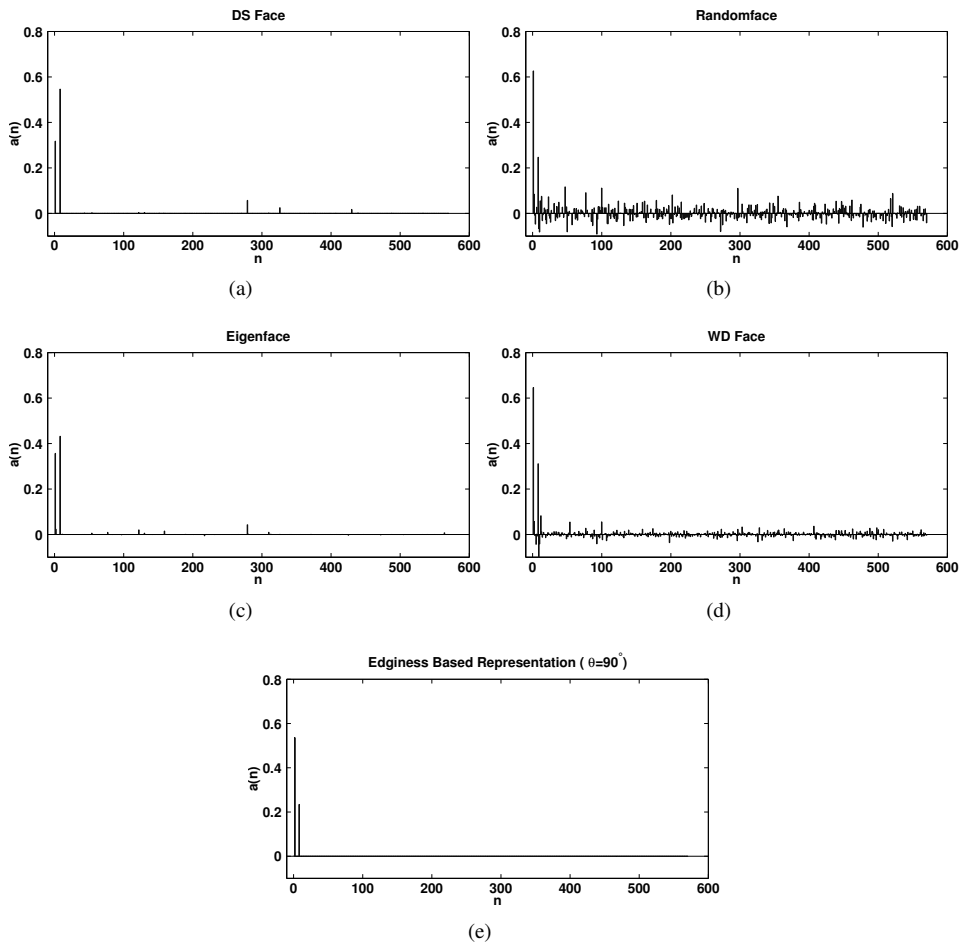
Figure 9: Sparse coefficient vector **a** for a given face image obtained using (a) DS Face, (b) Randomface, (c) Eigenface, (d) WD Fce and (e) Partial evidence (for $\theta = 90°$).

4   Comparison of Results

Figure 8 compares the recognition rates obtained by different face representation schemes as discussed previously and the combined effect of partial evidences. Recognition using Equation 5 yields better results when the training set is very small since edginess representation provides relatively less illumination variant representation and combined effect of partial evidences compliments this with better discrimination ability. Moreover, we can compare the sparse coefficient vector **a** obtained for different face representation schemes as discussed in this paper. The sparse coefficients for a particular test image from class 1 for different choice of dictionary are shown in Figure 9. The x-axis stands for each person in the dictionary and y-axis indicates the value of the sparse coefficient vector **a** for the corresponding person. It can be observed from the Figure 8 that edginess representation of face images gives highly sparse coefficients as compared to the other representation schemes.

5   Summary

In this paper, we have discussed the significance of dictionary in sparse coding based face recognition. The issue of sufficiency in training set was addressed using weighted decomposition representation of face images (WD Face). The weights were given in a manner which reflects the person specific unique information in the resulting representation. In order to avoid the illumination sensitivity of the gray scale images, edginess based representations are used. These representations were found to be highly subject specific and illumination invariant. When these representations are implemented, sparse coding based face recognition performs well even when number of training samples is reduced drastically. By combining the residuals obtained from these partial evidences, a better discrimination measure between true and impostor class was obtained, which further improves the performance. Experimental results indicate that with suitable representation of training images (dictionary), sparse coding based face representation can address the issue of illumination and sufficiency of training images.

References

[AGRA08]   A. Amine, S. Ghouzali, M. Rziza, and D. Aboutajdine. Investigation of feature dimension reduction based DCT/SVM for face recognition. In *IEEE Symposium on Computers and Communications, 2008.*, pages 188 –193, july 2008.

[BCF06]   Kevin W. Bowyer, Kyong Chang, and Patrick Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding*, 101(1):1 – 15, 2006.

[CRT06]   E.J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489 – 509, feb. 2006.

[DHG12]   W. Deng, J. Hu, and J. Guo. Extended SRC: Undersampled Face Recognition via Intra-Class Variant Dictionary. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1, 2012.

[Don06]   D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289 –1306, april 2006.

[GBK01]   A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[GBMK11]   Ralph Gross, Simon Baker, Iain Matthews, and Takeo Kanade. Face Recognition Across Pose and Illumination. In Stan Z. Li and Anil K. Jain, editors, *Handbook of Face Recognition, 2nd Edition*, pages 197–221. Springer, 2011.

[KKL⁺07]   Seung-Jean Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An Interior-Point Method for Large-Scale l1-Regularized Least Squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606 –617, dec. 2007.

[KSY09]   A. Kumar Sao and B. Yegnanarayana. Analytic Phase-based Representation for Face Recognition. In *Seventh International Conference on Advances in Pattern Recognition, 2009*, pages 453 –456, feb. 2009.

[LHK05]   K.C. Lee, J. Ho, and D. Kriegman. Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[PLY⁺09]   Yanwei Pang, Xuelong Li, Yuan Yuan, Dacheng Tao, and Jing Pan. Fast Haar Transform Based Feature Extraction for Face Representation and Recognition. *IEEE Transactions on Information Forensics and Security*, 4(3):441 –450, sept. 2009.

[SYK07]   Anil Kumar Sao, B. Yegnanarayana, and B. V. K. Vijaya Kumar. Significance of image representation for face verification. *Signal, Image and Video Processing*, 1(3):225–237, 2007.

[TG07]   J.A. Tropp and A.C. Gilbert. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*, 53(12):4655 –4666, dec. 2007.

[TP91]   M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991*, pages 586 –591, jun 1991.

[VS06]   K.B. Vinay and B.S. Shreyas. Face Recognition Using Gabor Wavelets. In *Fortieth Asilomar Conference on Signals, Systems and Computers, 2006.*, pages 593 –597, 29 2006-nov. 1 2006.

[WMM⁺10]   J. Wright, Yi Ma, J. Mairal, G. Sapiro, T.S. Huang, and Shuicheng Yan. Sparse Representation for Computer Vision and Pattern Recognition. *Proceedings of the IEEE*, 98(6):1031 –1044, june 2010.

[WYG⁺09]   J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Yi Ma. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210 –227, feb. 2009.