ROBUST FACE RECOGNITION USING LOCALLY ADAPTIVE SPARSE REPRESENTATION

Yi Chen, Thong T. Do, and Trac D. Tran

Department of Electrical and Computer Engineering
The Johns Hopkins University

ABSTRACT

This paper presents a block-based face-recognition algorithm based on a sparse linear-regression subspace model via locally adaptive dictionary constructed from past observable data (training samples). The local features of the algorithm provide an immediate benefit – the increase in robustness level to various registration errors. Our proposed approach is inspired by the way human beings often compare faces when presented with a tough decision: we analyze a series of local discriminative features (do the eyes match? how about the nose? what about the chin?...) and then make the final classification decision based on the fusion of local recognition results. In other words, our algorithm attempts to represent a block in an incoming test image as a linear combination of only a few atoms in a dictionary consisting of neighboring blocks in the same region across all training samples. The results of a series of these sparse local representations are used directly for recognition via either maximum likelihood fusion or a simple democratic majority voting scheme. Simulation results on standard face databases demonstrate the effectiveness of the proposed algorithm in the presence of multiple mis-registration errors such as translation, rotation, and scaling.

1. INTRODUCTION

Sparse representations have been recently exploited in many pattern recognition applications [1–3]. These approaches are based on the assumption that a test sample approximately lies in a low-dimensional subspace spanned by the training data and thus can be compactly represented by a few training samples. The recovered sparse vector then can be used directly for recognition. This approach is simple and fast since no training stage is needed and the dictionary can be easily expanded by additional training samples. The original sparsity-based face recognition algorithm [1] yields superior recognition performance comparing to existing techniques. However, it suffers from the limitation that the test face must be perfectly aligned to the training data prior to classification. To overcome this problem, various methods have been proposed for simultaneously optimizing the registration parameters and the sparse coefficients [4, 5], leading to even more complicated systems.

In many signal processing applications, local features are more representative and contain more important information than global features. One of such examples is the block-based motion estimation technique successfully employed in multiple video compression standards. In this paper, we investigate the usage of local features in the face recognition task. We propose a robust yet simple approach to deal with the misalignment problem by adopting a local block-based sparsity model. The model is based on the observation that a block in a test image can be sparsely represented by neighboring blocks in the training images and the sparse representation encodes the block identity. Note that in our proposed approach, no explicit registration is required. We propose to use multiple blocks, classify each block individually, and then combine the classification results for all blocks. In this way, instead of making a decision on one single global sparse representation, we rely on a combination of decisions from local sparse representations. This approach exploits the flexibility of the local block-based model and its ability to capture relatively stationary features under uniform and nonuniform variations, leading to a system robust to various types of misalignment.

The remainder of this paper is structured as follows. Details about the proposed method are described in Section 2. Its effectiveness is demonstrated in Section 3 by simulation results. Finally, Section 4 summarizes our work and makes some closing remarks.

2. BLOCK-BASED ROBUST FACE RECOGNITION

We first briefly introduce the original sparsity-based face recognition technique [1]. It is observed that a test sample can be expressed by a sparse linear combination of training samples

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\alpha},$$

where $\boldsymbol{y}$ is the vectorized test sample, columns of $\boldsymbol{D}$ are the vectorized training samples of all classes, and $\boldsymbol{\alpha}$ is a sparse vector (i.e., only few entries in $\boldsymbol{\alpha}$ are nonzero). The classifier seeks the sparsest representation by solving

$$\hat{\boldsymbol{\alpha}}_0 = \arg\min \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \quad \boldsymbol{D}\boldsymbol{\alpha} = \boldsymbol{y}, \tag{1}$$

where $\|\cdot\|_0$ denotes the ℓ_0-norm which is defined as the number of nonzero entries in the vector. Once the sparse vector is recovered, the identity of $\boldsymbol{y}$ is then given by the minimal residual

$$\text{identity}(\boldsymbol{y}) = \arg\min_i \|\boldsymbol{y} - \boldsymbol{D}\delta_i(\hat{\boldsymbol{\alpha}}_0)\|, \tag{2}$$

where $\delta_i(\boldsymbol{\alpha})$ is a vector whose only nonzero entries are the same as those in $\boldsymbol{\alpha}$ associated with class i. With the recently-developed theory of compressed sensing [6], the ℓ_0-norm minimization problem (1) can be efficiently solved by recasting it as a linear programming problem. Alternatively, the problem in (1) can be solved by greedy pursuit algorithms [7, 8].

As previously mentioned, the original technique [1] does not address the problem of registration errors in the test data. In what follows, we propose a robust yet simple approach to deal with misalignment by exploiting the flexibility of the local block-based model. Let K be the number of classes in the training data and N_k be the number of training samples in the kth class. We adopt the inter-frame sparsity model [9] in which a block in a video frame

This work has been supported in part by the National Science Foundation under Grant CCF-0728893.

can be sparsely represented by few neighboring blocks in reference frames. Fig. 1(a) illustrates the proposed method of representing a block in the test face image $\boldsymbol{Y}$ from a locally adaptive dictionary consisting of neighboring blocks in the training images $\{\boldsymbol{X}_t\}_{t=1,\ldots,T}$ in the same physical area, where $T=\sum_{k=1}^{K} N_k$ is the total number of training samples (only one training image is shown in Fig. 1). To be more specific, let $\boldsymbol{y}_{ij}$ be an MN-dimensional vector representing the vectorized $M\times N$ block in the test image with the upper left pixel located at (i,j). Define the search region $\boldsymbol{S}_{ij}^t$ to be the $(M+2\triangle M)\times(N+2\triangle N)$ block in the tth training image $\boldsymbol{X}_t$ as:

$$\boldsymbol{S}_{ij}^t=\begin{bmatrix} x_{i-\triangle M,j-\triangle N}^t & \cdots & x_{i-\triangle M,j+N-1+\triangle N}^t \\ \vdots & \ddots & \vdots \\ x_{i+M-1+\triangle M,j-\triangle N}^t & \cdots & x_{i+M-1+\triangle M,j+N-1+\triangle N}^t \end{bmatrix}.$$

From the search regions of all T training images, we can construct the dictionary $\boldsymbol{D}_{ij}$ for the block $\boldsymbol{y}_{ij}$ as

$$\boldsymbol{D}_{ij}=\left[\boldsymbol{D}_{ij}^1 \quad \boldsymbol{D}_{ij}^2 \quad \cdots \quad \boldsymbol{D}_{ij}^T\right],$$

where each

$$\boldsymbol{D}_{ij}^t=\left[\boldsymbol{d}_{i-\triangle M,j-\triangle N}^t \quad \boldsymbol{d}_{i-\triangle M,j-\triangle N+1}^t \quad \cdots \quad \boldsymbol{d}_{i+\triangle M,j+\triangle N}^t\right]$$

is an $(MN)\times\left((2\triangle M+1)(2\triangle N+1)\right)$ matrix whose columns are the vectorized blocks in the tth training image defined in the same way as $\boldsymbol{y}_{ij}$. The dictionary $\boldsymbol{D}_{ij}$ is locally adaptive and changes from block to block. The size of the dictionary depends on the non-stationary behavior of the data as well as the level of computational complexity we can afford. In the presence of registration error, the test image $\boldsymbol{Y}$ may no longer lie in the subspace spanned by the training samples $\{\boldsymbol{X}_t\}_t$. At the block level, however, $\boldsymbol{y}_{ij}$ can still be approximate by the blocks in the training samples $\left\{\boldsymbol{d}_{ij}^t\right\}_{t,i,j}$. Compared to the original approach, the dictionary $\boldsymbol{D}_{ij}$ better captures the local characteristics. Note that our approach is quite different from patch-based dictionary learning [10] from several angles: *(i)* we emphasize the local adaptivity of the dictionaries; and *(ii)* dictionaries in our approach are directly obtained from the data without any complicated learning process.

We propose that the block $\boldsymbol{y}_{ij}$ in the misaligned image $\boldsymbol{Y}$ can be sparsely approximated by a linear combination of a few atoms in the dictionary $\boldsymbol{D}_{ij}$:

$$\boldsymbol{y}_{ij}=\boldsymbol{D}_{ij}\boldsymbol{\alpha}_{ij}, \tag{3}$$

where $\boldsymbol{\alpha}_{ij}$ is sparse vector, as illustrated in Fig. 1(b). The sparse vector can be recovered by solving the minimal ℓ_0-norm problem

$$\hat{\boldsymbol{\alpha}}_{ij}=\arg\min\left\|\boldsymbol{\alpha}_{ij}\right\|_0 \quad \text{subject to} \quad \boldsymbol{D}_{ij}\boldsymbol{\alpha}_{ij}=\boldsymbol{y}_{ij}. \tag{4}$$

Since our sparse recovery is performed on a small block of data with a modest size dictionary, the resulting complexity of the overall algorithm is manageable. After the sparse vector $\hat{\boldsymbol{\alpha}}_{ij}$ is obtained, the identity of the test block can be determined by the error residuals by

$$\text{identity}(\boldsymbol{y}_{ij})=\arg\min_{k=1,\ldots,K}\left\|\boldsymbol{y}_{ij}-\boldsymbol{D}_{ij}\boldsymbol{\delta}_k\left(\hat{\boldsymbol{\alpha}}_{ij}\right)\right\|_2, \tag{5}$$

where $\boldsymbol{\delta}_k\left(\hat{\boldsymbol{\alpha}}_{ij}\right)$ is as defined in (2).

To improve the robustness, we propose to employ multiple blocks, classify each block individually, and then combine the classification results. The blocks may be chosen completely at random, or manually in the more representative areas (such as the region around eyes) or areas with high SNR, or exhaustively in the entire test image (non-overlapped or overlapped). Note that since each block is handled independently, they can be processed in parallel. Also, since blocks can be overlapped, our proposed algorithm is computationally scalable - more computation delivers better recognition result.

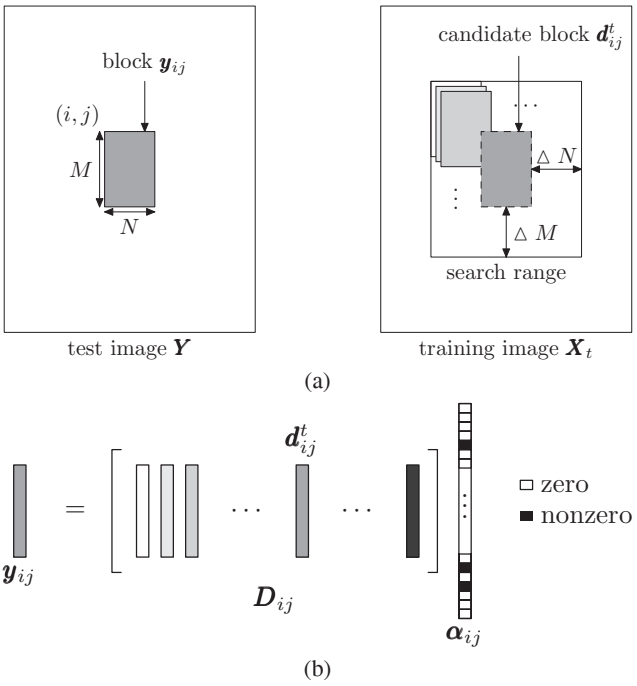

Fig. 1. Representation of a block in the test image from a locally adaptive dictionary. (a) The blocks in the test and training images (only one training sample is displayed). (b) Sparse representation $\boldsymbol{y}_{ij}=\boldsymbol{D}_{ij}\boldsymbol{\alpha}_{ij}$.

Once the recognition results are obtained for all blocks, they can be combined by majority voting. Let L be the number of blocks in the test image $\boldsymbol{Y}$, and $\{\boldsymbol{y}_l\}_{l=1,\ldots,L}$ be the L blocks. Then, by majority voting

$$\text{identity}(\boldsymbol{Y})=\max_{k=1,\ldots,K}\left|\{l=1,\ldots,L:\text{identity}(\boldsymbol{y}_l)=k\}\right|,$$

where $|S|$ denotes the cardinality of a set S and $\text{identity}(\boldsymbol{y}_l)$ is determined by (5).

Maximum likelihood is an alternative way to fuse the classification results from multiple blocks. For a block $\boldsymbol{y}_l$, its sparse representation $\hat{\boldsymbol{\alpha}}_l$ obtained by solving (4), and the local dictionary $\boldsymbol{D}_l$, we define the probability of $\boldsymbol{y}_l$ belonging to the kth class to be inversely proportional to the residual associated with the dictionary atoms in the kth class:

$$p_l^k=P(\text{identity}(\boldsymbol{y}_l)=k)=\frac{1/r_l^k}{\sum_{k=1}^{K}\left(1/r_l^k\right)}, \tag{6}$$

where $r_l^k=\|\boldsymbol{y}_l-\boldsymbol{D}_l\boldsymbol{\delta}_k(\hat{\boldsymbol{\alpha}}_l)\|_2$ is the residual associated with the kth class and the vector $\boldsymbol{\delta}_k(\hat{\boldsymbol{\alpha}}_l)$ is as defined in (5). Then, the identity of the test image $\boldsymbol{Y}$ is given by

$$\text{identity}(\boldsymbol{Y})=\arg\max_{k=1,\ldots,K}\log\left(\prod_{l=1}^{L}p_l^k\right). \tag{7}$$

The maximum likelihood approach can also be used as a measure to reject outliers, as for an outlier the probability of it belonging to some class tends to be uniformly distributed among all classes in the training data.

Fig. 2 illustrates an example of the proposed approach with multiple blocks. The test and training images are taken from the Extended Yale B Database [11] which consists of face images of

38 individuals. More details about this database and the experiment setup will be described in the next section. Fig. 2(a) shows the original (registered) image in the 31st class, and Fig. 2(b) shows the test image to be classified, which is obtained by translating the original one by 3 pixels in each direction, rotating by 4 degrees, and then zooming in by a scaling factor of 1.125 in the vertical direction and 1.143 in the horizontal direction. Due to the misalignment, the original global approach in [1] leads to misclassification, as seen by the residuals in Fig. 2(c) where the 7th class has the minimal residual. Using the proposed approach, we choose 42 blocks of size 8×8 uniformly from the test image in Fig. 2(b). The blocks and classification result for each individual block are displayed in Fig. 2(d). Figs. 2(e) and (f) show the results using majority voting and maximum likelihood, respectively. In both cases, the block-based algorithm yields the correct answer.

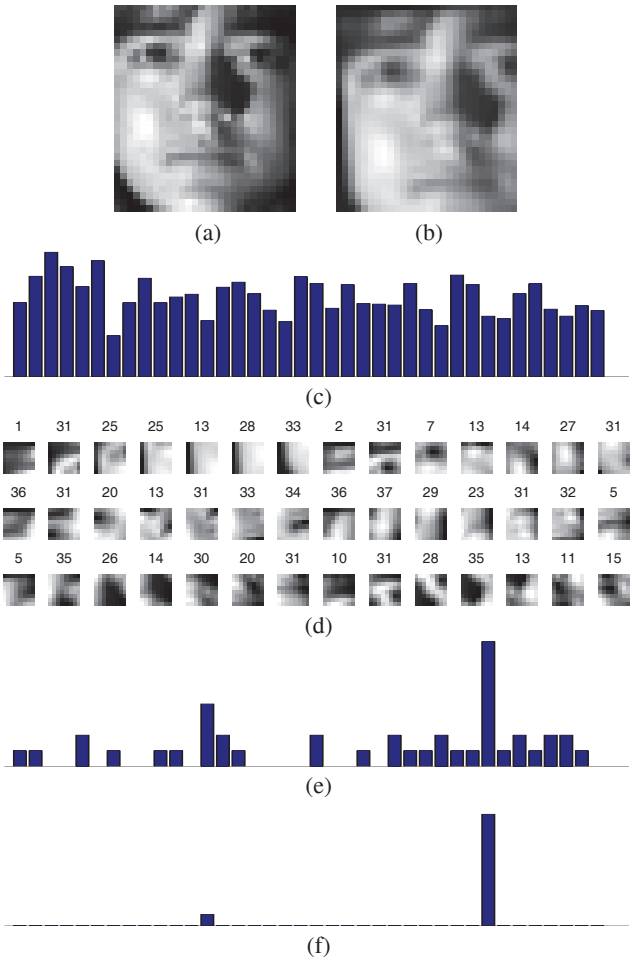

Fig. 2. Example of the proposed sparsity-based approach using multiple blocks. (a) Original image. (b) Distorted test image $\boldsymbol{Y}$. (c) Residuals using the original global approach: identity$(\boldsymbol{Y}) = 7$. (d) Classification results for each of the 42 blocks $\{\boldsymbol{y}_l\}_{l=1,\ldots,42}$. (e) Number of votes for the kth class, $k = 1,\ldots,38$. Identity$(\boldsymbol{Y}) = 31$. (f) Probability of (identity$(\boldsymbol{Y}) = k$), $k = 1,\ldots,38$. Identity$(\boldsymbol{Y}) = 31$.

The above example illustrates the process of the block-based algorithm in the presence of registration errors. When the errors become more significant, we may also augment the local dictionary by including distorted versions of the local blocks in the training data for a better performance, at the cost of higher computational complexity.

3. SIMULATION RESULTS

In this section, we apply the proposed block-based algorithm for identification on a publicly available database - the Extended Yale B Database [11], and compare the performance with the original algorithm in [1]. This database consists of 2414 perfectly-aligned frontal face images of size 192×168 of 38 individuals, 64 images per individual, under various conditions of illumination. In our experiments, for each subject we randomly choose 15 images in Subsets 1 and 2, which were taken under less extreme lighting conditions, as the training data. Then, we randomly choose 500 images from the remaining images as test data. All training and test samples are downsampled to size 32×28. The Subspace Pursuit algorithm [8] is used to solve the sparse recovery problem (4).

To verify the effectiveness of the proposed algorithm under registration errors, we create distorted test images in several ways and keep the training images unchanged. Obviously, the proposed algorithm is robust to image translation by choosing an appropriate search region for each block such that the corresponding blocks in the training images are included in the dictionary. Next, we show experimental results for test images under rotation and scaling operations. In the first set of experiments, the test images are rotated by degrees between -20 and 20, as seen by the example in Fig. 3. We apply the block-based algorithm to 42 blocks of size

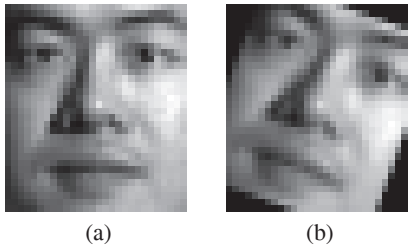

Fig. 3. An example of rotated test images. (a) Original image and (b) the image rotated by 20 degrees clockwise.

8×8 uniformly located on the test image, and the results are combined using the maximum likelihood approach (6). Fig. 4 shows the recognition rate (y-axis) for each rotation degree (x-axis). We see that at a higher level of misalignment, the block-based algorithm (in red circles) outperforms the original algorithm (in blue x-marks) by a large margin.

For the second set of experiments, the test images are stretched in both directions by scaling factors up to 1.313 vertically and 1.357 horizontally. An example of an aligned image in the database and its distorted version to be tested are shown in Fig. 5. Similar to the previous case, for each test image, we apply the algorithm to 42 uniformly-located blocks of size 8×8 and combine the results by (6). Tables 1 and 2 show the percentage of correct identification out of 500 tests with various scaling factors. The first row and the first column in the tables indicate the scaling factors in the horizontal and vertical directions, respectively, and the other entries correspond to the recognition rate in percentage. We see that again when there are large registration errors, the block-based algorithm leads to a better identification performance than the original algorithm.

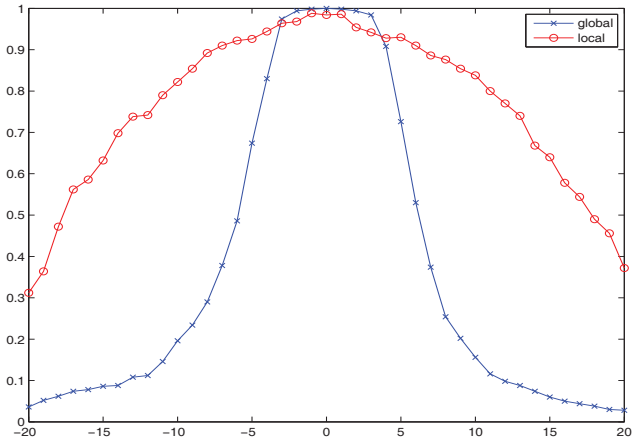

Fig. 4. Recognition rate for rotated test images.

(a) (b)

Fig. 5. An example of scaled test images. (a) Original image and (b) the image scaled by 1.313 vertically and 1.357 horizontally.

Table 1. Recognition rate (in percentage) for scaled test images using the original global approach in [1] under various scaling factors (SF).

SF	1	1.071	1.143	1.214	1.286	1.357
1	100	100	94.8	71.4	51.8	41.4
1.063	99.2	95.0	76.6	51.8	33.8	28.6
1.125	84.6	66.4	42.6	25.2	18.6	14.6
1.188	52	37.2	20.6	15.6	11.6	8
1.25	33.2	26.4	16.8	11.4	9.4	7.6
1.313	33.6	22.6	14.6	10.6	7.4	7.6

Table 2. Recognition rate (in percentage) for scaled test images using the proposed block-based approach under various SF.

SF	1	1.071	1.143	1.214	1.286	1.357
1	98	96.4	97.6	96.4	96.4	95.2
1.063	97.4	96.6	96.6	95.6	92.4	90
1.125	97	95.4	94.6	94.6	92.6	90.2
1.188	95	94	91.8	90.2	85.6	82.2
1.25	93.8	92.4	89	85	79.4	73.6
1.313	88.8	85	79	75.8	67	59.2

In the last experiment, the 500 test images are shifted by 3 pixels downwards and rightwards (about 10% of the side lengths), rotated by 4 degrees counterclockwise, and then zoomed in by 1.125 and 1.143 in vertical and horizontal directions, respectively. One example of the misaligned test images is shown in Figs. 2(a) and (b). In this case of combined misalignment, the original approach only successfully identifies 20 out of 500 test images, while the block-based algorithm yields an identification rate of 82% (i.e., 410 out of 500 are correctly recognized).

4. CONCLUSION

In this paper, we propose a block-based algorithm for face recognition via sparse representation. By constructing locally adaptive dictionaries that capture the relative stationary features in a small neighborhood, the proposed algorithm is robust to various types of misalignment between the test and training data, without explicit computation of the registration parameters. We propose to use multiple blocks in the same test image and combine all classification results to further improve the robustness. As demonstrated by the simulation results on the Extended Yale B Database, the proposed algorithm yields excellent performance in the presence of registration errors.

5. REFERENCES

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

[2] J. K. Pillai, V. M. Patel, and R. Chellappa, "Sparsity inspired selection and recognition of iris images," in *Proc. IEEE Third International Conference on Biometrics: Theory, Applications and Systems*, Sept. 2009, pp. 1–6.

[3] X. Hang and F.-X. Wu, "Sparse representation for classification of tumors using gene expression data," *Journal of Biomedicine and Biotechnology*, vol. 2009, 2009, doi:10.1155/2009/403689.

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

[5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Towards a practical face recognition system: Robust registration and illumination by sparse representation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 597–604.

[6] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[7] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[8] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. on Information Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[9] T. T. Do, Y. Chen, D. T. Nguyen, N. H. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proc. of IEEE International Conference on Image Processing*, Nov. 2009.

[10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[11] 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 Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, June 2001.