SPARSE REPRESENTATION AND LEAST SQUARES-BASED CLASSIFICATION IN FACE RECOGNITION

Michael Iliadis[⋆], *Leonidas Spinoulas*[⋆], *Albert S. Berahas*[†], *Haohong Wang*[‡], *Aggelos K. Katsaggelos*[⋆]

[⋆] Dept. of Electrical Engineering and Comp. Sc., Northwestern University, Evanston, IL 60208, USA
[†] Dept. of Eng. Sciences and Appl. Mathematics, Northwestern University, Evanston, IL 60208, USA
[‡] TCL Research America, San Jose, CA 95134, USA

ABSTRACT

In this paper we present a novel approach to face recognition. We propose an adaptation and extension to the state-of-the-art methods in face recognition, such as sparse representation-based classification and its extensions. Effectively, our method combines the sparsity-based approaches with additional least-squares steps and exhitbits robustness to outliers achieving significant performance improvement with little additional cost. This approach also mitigates the need for a large number of training images since it proves robust to varying number of training samples.

Index Terms— Face recognition, sparse representation, classification.

1. INTRODUCTION

One of the most popular problems in computer vision is face recognition (FR) which focuses on deducing a subject's identity through a provided image [1]. Over the last decade, researchers have been focusing on addressing practical large-scale FR systems in uncontrolled environments [2].

Recently, face recognition via sparse representation-based classification (SRC) [3] and its extensions [4, 5] have proven to provide state-of-the-art performance. The main idea is that a subject's face sample can be represented as a sparse linear combination of available images of the same subject captured under different conditions (e.g., poses, lighting conditions, etc.). The same principle can also be applied when a face image itself is represented in a lower dimensional space describing important and easily identifiable features. In order to enforce sparsity, ℓ_1 optimization algorithms [6, 7] can be employed. Then, the face class that yields the minimum reconstruction error is selected in order to classify or identify the subject, whose test image or sample is available.

Sparse coding has also been proposed to jointly address the problems of blurred face recognition and blind image recovery [8]. Furthermore, it has been utilized in [9] to deduce a transformation-invariant face recognition algorithm as well as in [10] to extend the SRC framework for handling misalignment, pose and illumination invariance.

Despite their success, the necessity of ℓ_1 optimization methods for improved face recognition rates has been recently criticized [11, 12]. Zhang *et. al.* argued that FR performance improvement stems from the collaborative representation of one image using multiple similar images rather than the usage of the ℓ_1 norm in the optimization procedure, and proposed a simpler regularized least squares formulation to solve the FR problem. Furthermore, the authors in [11, 13] state that ℓ_1 norm techniques can only be successful under certain conditions. Specifically, Wright *et. al.* state in [13]: "*The sparse representation based face recognition assumes that the training images have been carefully controlled and that the number of samples per class is sufficiently large. Outside these operating conditions, it should not be expected to perform well.*"

In order to overcome the limitation of requiring large amounts of samples per class, Deng et. al. [4, 5] proposed the use of additional dictionaries, constructed using the available data, to refine FR performance. The motivation for this work comes from the fact that different images of the same subject share a lot of similarities; hence, an additional intra-class generic dictionary of subject samples could effectively model face variations. In this method, which is called Extended SRC (ESRC), the test sample is represented as a sparse linear combination of the training and intra-class dictionary samples.

In this work, we show that sparsity-based approaches can be effectively combined with additional least-squares steps to provide significant performance improvement with little additional cost. Moreover, the proposed approach can overcome the need for a large number of training images since it proves robust to varying number of training samples as we will show in the experimental section.

This paper is organized as follows. Section 2 overviews existing face recognition classifiers and presents their modeling. The proposed face recognition algorithm is analyzed in Section 3. Finally, experimental results and discussion about the performance of the proposed approach are presented in Section 4 and conclusions are drawn in Section 5.

2. RELATED WORK

In this section we survey face recognition algorithms based on sparse representation and regularized least squares. Let $\mathbf{y} \in \mathbb{R}^d$ denote the face test sample, where d is the dimensionality of a selected face feature and $\mathbf{T} = [T_i, \ldots, T_c] \in \mathbb{R}^{d \times n}$ denote the matrix (dictionary) with the set of samples of c subjects stacked in columns. $T_i \in \mathbb{R}^{d \times n_i}$ denotes the n_i set of samples of the i^{th} subject, such that, $\sum_i n_i = n$.

Sparse Representation-based Classification (SRC) [3]: In SRC, the test sample $\mathbf{y}$ is represented by,

$$\mathbf{y} = \mathbf{T}\mathbf{a} + \mathbf{e}, \tag{1}$$

where $\mathbf{e} \in \mathbb{R}^d$ is dense noise and $\mathbf{a} \in \mathbb{R}^n$ is a sparse vector with nonzero elements corresponding to few samples in $\mathbf{T}$. Thus, the test sample can be represented as a sparse linear combination of the samples in $\mathbf{T}$. The coefficients of $\mathbf{a}$ can be estimated solving the optimization problem,

$$\hat{\mathbf{a}} = \underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{T}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1 . \tag{2}$$

The complete steps of the SRC algorithm are presented in Algorithm 1.

Extended SRC (ESRC) [4]: In ESRC, the test sample $\mathbf{y}$ is represented by,

$$\mathbf{y} = \mathbf{T}\mathbf{a} + \mathbf{V}\mathbf{b} + \mathbf{e}, \tag{3}$$

where $\mathbf{V} \in \mathbb{R}^{d \times n}$ is a variation dictionary that models *intra-class* variant bases, such as, lighting changes, exaggerated expressions, or occlusions, for the representation of each subject i, while $\mathbf{a} \in \mathbb{R}^n$ is a sparse vector, as in SRC. Different types of variations, that cannot be captured by $\mathbf{V}$, are represented by the dense noise term $\mathbf{e} \in \mathbb{R}^d$. Vector $\mathbf{b} \in \mathbb{R}^n$ is also considered to be sparse and its coefficients can effectively capture the contribution of uncontrolled viewing conditions in the final image and are, hence, not informative about the subject's identity. Thus, the test sample is represented as the linear combination of $\mathbf{T}\mathbf{a}$, capturing the subject's identity, and $\mathbf{V}\mathbf{b}$, capturing sparse noise terms. The variation matrix $\mathbf{V}$ can be constructed by the differences of each sample to its corresponding's class centroids, as suggested in [4],

$$\mathbf{V} = [T_1 - \mathbf{m}_1\mathbf{r}_1^T, \ldots, T_c - \mathbf{m}_c\mathbf{r}_c^T], \tag{4}$$

where $\mathbf{m}_i = \frac{1}{n_i} T_i \mathbf{r}_i$ is the centroid of class i, and $\mathbf{r}_i = [1, \ldots, 1]^T \in \mathbb{R}^{n_i}$.

In ESRC, the sparse vectors $\mathbf{a}$ and $\mathbf{b}$ can be obtained by,

$$\hat{\mathbf{a}}, \hat{\mathbf{b}} = \underset{\mathbf{a},\mathbf{b}}{\operatorname{argmin}} \left\| \mathbf{y} - [\mathbf{T}, \mathbf{V}] \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right\|_2^2 - \lambda \left\| \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right\|_1 . \tag{5}$$

Similar to SRC, classification (or else subject identification) is performed by selecting the class i that provides the smallest residual. The difference is that in computing the residual of each class, the term $\mathbf{V}\mathbf{b}$ is also subtracted from the test sample.

Algorithm 1 The SRC Algorithm

Inputs: Vector $\mathbf{y}$ and matrix $\mathbf{T}$.

1. Normalize the columns of $\mathbf{T}$ to have unit ℓ_2-norm.
2. Estimate the sparse vector $\hat{\mathbf{a}}$ solving the problem in (2).
3. Compute the residuals for each class i as,

$$e_i(\mathbf{y}) = \|\mathbf{y} - T_i\hat{\mathbf{a}}_i\|_2 ,$$

where $\hat{\mathbf{a}}_i$ is the segment of $\mathbf{a}$ associated with class i.

Output: Identity of $\mathbf{y}$ as, Identity$(\mathbf{y}) = \operatorname{argmin}_i\{e_i\}$

Regularized least squares (CR-RLS) [12]: In CR-RLS a regularized least squares method is proposed in order to collaboratively represent the test sample without imposing sparsity constraints on the unknown variable $\mathbf{a}$. Again, classification is performed by minimizing the reconstruction term for each class. The optimization problem, of this very efficient method, is given by,

$$\hat{\mathbf{a}} = \underset{\mathbf{a}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{T}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_2^2, \tag{6}$$

which can be easily solved in closed form.

3. SPARSE REPRESENTATION AND REGULARIZED LEAST SQUARES

Let the linear representation of the test sample, $\mathbf{y}$, be given by (3). In [4], $\mathbf{a}$ and $\mathbf{b}$ were regularized using the ℓ_1 norm. As Zhang *et. al.* showed in [12], when increasingly many training samples are used for the representation of a test sample, the discriminating ability between classes reduces, leading to comparably small reconstruction errors for all classes. Thus, sparsity of coefficients should be considered so that only a few training samples represent the test sample. Similarly, if there are redundant and overcomplete facial variant bases in $\mathbf{V}$, the combination coefficients in $\mathbf{b}$ are naturally sparse.

In the proposed Sparse Representation and Regularized Least Squares (SR+RLS) method, an initial estimation of $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ is obtained by solving the ESRC problem in (5). Ideally, this initial estimation will provide us with the largest coefficients at locations of vector $\mathbf{a}$ corresponding to the class that the test sample belongs to. In other words, we expect this class to exhibit the smallest residual compared to other classes. However, due to face variations this may not always be the case, and the coefficient values in $\mathbf{a}$ could be noisy. Having estimated $\hat{\mathbf{a}}$ through the optimization problem in (5), we expect that the test sample is best represented as a linear combination of training samples of its true class. Nevertheless, the corresponding mixing coefficients need not be the largest.

The decision of the correct identity in SRC is dictated by the minimum residual. In this work we account for the possibility of the correct identity hiding under slightly higher residuals than the minimum, due to face variations. Thus, we propose a novel face recognition algorithm that is solved in four separate steps.

1. We first estimate $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ solving the optimization problem in (5).
2. Based on the initial estimation of $\hat{\mathbf{a}}$, we construct a new face dictionary that consists of the training samples of the classes whose corresponding coefficients in $\hat{\mathbf{a}}$ are nonzero, while the remaining sets of training samples for all other classes are nullified (set to zero).
3. Having constructed the new smaller dictionary we can estimate the new coefficients by solving a regularized least squares problem.
4. Finally, the face identity will be chosen based on the minimum class residual provided by the updated coefficients.

Next, we discuss the aforementioned steps in more detail.

3.1. The Sparse Representation (SR) step

In order to obtain the initial estimation of $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ we solve the problem in (5). This is a standard sparse coding problem and can be solved using any ℓ_1 minimization algorithm, such as Homotopy [14].

3.2. Dictionary construction

Let the function $f(\hat{\mathbf{a}}_i)$, where $\hat{\mathbf{a}}_i$ is the segment of $\hat{\mathbf{a}}$ associated with class i, be given as,

$$f(\hat{\mathbf{a}}_i) = \begin{cases} 0, & \text{if } \hat{\mathbf{a}}_i = \mathbf{0} \\ 1, & \text{otherwise} \end{cases}. \tag{7}$$

Then the new dictionary $\tilde{\mathbf{T}}$ is constructed as follows,

$$\tilde{\mathbf{T}} = [f(\hat{\mathbf{a}}_i) \odot T_i, \ldots, f(\hat{\mathbf{a}}_c) \odot T_c] \in \mathbb{R}^{d \times n} \tag{8}$$

where $\odot$ denotes the convolution operator.

3.3. The Regularized Least Squares (RLS) step

Having constructed the new dictionary with most training samples suppressed to zero, we can obtain a new estimation vector by solving the regularized least squares problem,

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}} \left\| \mathbf{y} - \tilde{\mathbf{T}}\mathbf{f} \right\|_2^2 + \lambda \|\mathbf{f}\|_2^2. \tag{9}$$

The problem in (9) has the closed form solution,

$$\hat{\mathbf{f}} = \left(\tilde{\mathbf{T}}^T \tilde{\mathbf{T}} + \lambda \mathbf{I} \right)^{-1} \tilde{\mathbf{T}}^T \mathbf{y}, \tag{10}$$

where $\hat{\mathbf{f}} \in \mathbb{R}^n$ is the vector with nonzero coefficients only at locations where the training samples are not zero, and $\lambda > 0$ is a constant.

Algorithm 2 Classification based on SR+RLS Algorithm

Inputs: Vector $\mathbf{y}$ and matrix $\mathbf{T}$.

1. Construct the variation matrix $\mathbf{V}$ using (4).
2. Apply PCA on the training samples $\mathbf{T}$ and project $\mathbf{T}$ and $\mathbf{V}$ onto a d dimensional space.
3. Normalize the columns of $\mathbf{T}$ and $\mathbf{V}$ to have unit ℓ_2-norm.
4. Estimate $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ solving the problem in (5).
5. Construct dictionary $\tilde{\mathbf{T}}$ using the estimated coefficients in $\hat{\mathbf{a}}$ and estimate $\hat{\mathbf{f}}$ using the problem in (9).
6. Compute the residuals for each class i as,

$$e_i(\mathbf{y}) = \left\| \mathbf{y} - \mathbf{T}_i \hat{\mathbf{f}}_i \right\|_2,$$

where $\hat{\mathbf{f}}_i$ is the coding coefficient vector associated with class i.

Output: Identity of $\mathbf{y}$ as, $\text{Identity}(\mathbf{y}) = \text{argmin}_i \{e_i\}$.

The solution to problem (9) has the following properties,

- The RLS step is more likely to provide the true identity since we reconstruct for fewer classes and thus less noise.
- We expect $\hat{\mathbf{f}}$ to have larger coefficient values corresponding to the true identity's training samples compared to the initial estimate $\hat{\mathbf{a}}$ through (5).
- The problem in (9) is well-defined since $\tilde{\mathbf{T}}$ is expected to consist of fewer (nonzero) columns than rows. Thus, RLS is appropriate for solving such problem.
- We do not add significant complexity to the solution since the least squares step in (10) can be solved very efficiently.

3.4. The SR+RLS classification

The SR+RLS algorithm is summarized in Algorithm 2. An example of our method is presented in Figure 2.

4. EXPERIMENTAL RESULTS

In this section we present experiments on publicly available databases, AR [15] and Extended Yale B [16], to show the efficacy of the proposed method. In order to solve the ℓ_1 minimization problem we use the Homotopy method [14]. We compare our method with SRC [3], ESRC [4] and CR-RLS[1] [12]. For SRC and ESRC we set the regularization parameter $\lambda = 0.005$ as proposed in [4] and for CR-LRS we set $\lambda = 0.001$ as suggested in [12]. For fair comparisons, we set the same parameters in our algorithm as the other methods: $\lambda = 0.005$ for the ℓ_1 problem and $\lambda = 0.001$ for the regularized least squares problem. All parameters remain the same for both databases. As input facial features we use Eigenfaces with $d = 300$.

[1] Source code obtained from author's website.

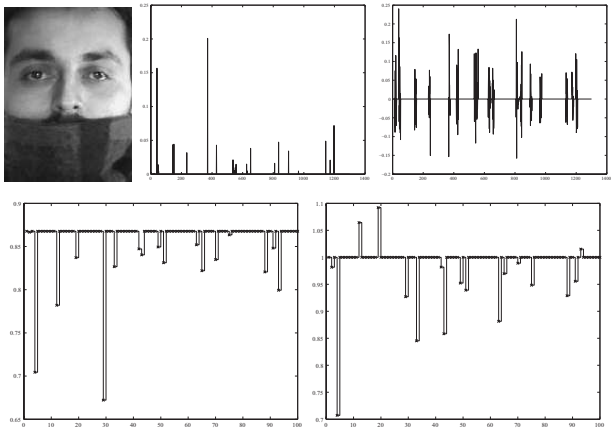

Fig. 1: The top left figure shows an example face, belonging to identity 4, occluded with a scarf. The upper left plot shows the coefficient values of the estimated $\hat{\mathbf{a}}$ using the problem in (5). The upper right plot shows the coefficient values of $\hat{\mathbf{f}}$ estimated by the problem in (9). The lower two plots show the estimated residuals after solving problems (5) and (9) from left to right, respectively. ESRC would classify this face as identity 29 since the lowest residual is at index 29 while the re-estimation of coefficients, as described in Section 3.3, results in the correct classification of the image providing the minimum residual at index 4.

Fig. 2: Example images from the AR database. This database is more challenging than the Yale, since more subjects need to be recognized while there exhibits large face variations within each class.

4.1. AR database

This experiment is a reproduction of that in section 5 of [11][2]. The AR database consists of over 3000 frontal images of 126 individuals. There are 26 images of each individual, and each person participated in two sessions, separated by two weeks. The faces in AR contain variations such as changes in illumination, expressions and facial disguises (e.g., sunglasses or scarfs). Example images of this database are shown in Figure 2.

In our experiments, 100 subjects were randomly chosen (50 male and 50 female). For each subject, we randomly permute the 26 images, and then take the first half for training and the rest for testing. Thus, we have 1300 training and 1300 testing samples. For statistical stability, we generate 10 different training and testing dataset pairs by randomly permuting 10 times. The images are cropped to have dimensions 165×120 pixels and converted to gray-scale.

[2]The method in [11] is very similar to the CR-LRS [12]. The difference is that in [12] a regularization ℓ_2 norm term is incorporated. However, with $\lambda = 0.001$ set to a relative small number the results were the same

Table 1: Recognition rate on the AR database.

Algorithms	Recognition rate	Time
SRC [3]	92.49 $\pm$0.82	0.0266s
ESRC [4]	96.98 $\pm$0.46	0.0982
CR-RLS [12]	96.80 $\pm$0.54	**0.0002s**
SR+RLS	**98.07 $\pm$0.38**	0.1040s

Table 2: Recognition rate on the Extended Yale B database.

Algorithms	Recognition rate	Time
SRC [3]	97.17 $\pm$0.62	0.0339s
ESRC [4]	96.65 $\pm$0.76	0.0630s
CR-RLS [12]	97.78 $\pm$0.49	**0.0002s**
SR+RLS	**98.11 $\pm$0.23**	0.0747s

Table 1 shows the recognition rates for this experiment. SR+RLS achieves the highest performance with 98.07% while ESRC has the second best recognition rate, slightly better than CR-RLS.

4.2. Extended Yale B database

The extended Yale B dataset consists of 2414 frontal face images of 38 subjects. They are captured under various lighting conditions. They are cropped to have dimensions 192×168 pixels and normalized. For each subject, we randomly select half of the images for training (i.e., about 32 images per subject) and the other half for testing. Again for each subject, we randomly permute the images per subject and take the first half for training and the rest for testing. For statistical stability, we generate 10 different training and testing dataset pairs.

In Table 2 we report the performance for this experiment. Again, SR+RLS achieves the highest recognition rate with 98.11%. CR-RLS achieves the second best, while ESRC has even lower performance than SRC.

4.3. Recognition from fewer training samples

In order to show the robustness of our proposed method we evaluate AR and Extended Yale B databases with fewer training samples per subject keeping the same number of test samples. For the AR database, we perform two experiments, where we randomly choose 8 (AR_8) and 4 (AR_4) training samples per subject. In the Yale database we randomly choose 8 training samples per subject. Again, for statistical stability, we generate 10 different training and test dataset pairs.

The results are presented in Table 3. It is apparent that, with fewer training samples, SRC's performance reduces dramatically, which is consistent with the observations in [11, 12]. On the other hand, our method proves its robustness even with fewer training samples per subject outperforming the state-of-the-art methods.

Table 3: Recognition rate on the AR and Extended Yale B databases using reduced training samples.

Algorithms	Yale_8	AR_8	AR_4
SRC [3]	85.15 $\pm$1.02	84.52 $\pm$0.79	69.12 $\pm$1.58
ESRC [4]	84.49 $\pm$1.29	93.30 $\pm$1.23	83.73 $\pm$0.91
CR-RLS [12]	88.51 $\pm$1.06	93.99 $\pm$0.64	85.06 $\pm$0.90
SR+RLS	**88.87 $\pm$0.65**	**95.48 $\pm$0.58**	**85.12 $\pm$1.04**

4.4. Discussion

The following conclusions are drawn based on the results on the two tested databases:

1. For both databases, SR+RLS outperforms SRC, ESRC and CR-RLS either with all training samples or few training samples.
2. ESRC has lower recognition rate for the Yale B database than the SRC when multiple training samples are considered. Instead, SR+RLS is consistent in performance in both databases for all cases.
3. Our method has the lowest time performance compared to the other methods. However, the overhead of the regularized least squares step is insignificant compared to the initial ℓ_1 estimation. All experiments were conducted in MATLAB on a 3.0 GHz PC with 8 GB RAM.

5. CONCLUSIONS

In this work, we presented a novel approach to face recognition effectively combining sparse representation and regularized least squares-based classification. We show that a simple additional least squares step in the optimization procedure can provide noticeable performance improvement while being robust to varying numbers of training samples in the dictionary. Hence, we manage to improve the face recognition performance of l_1 minimization schemes even with low availability of training data samples, despite the criticism that such methods are not beneficial under lack of numerous samples per class.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[2] E. G. Ortiz and B. C. Becker, "Face recognition for web-scale datasets," *Comput. Vis. Image Underst.*, vol. 118, pp. 153–170, Jan. 2014.

[3] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[4] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sept. 2012.

[5] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," in *IEEE Conf. Comp. Vision Pattern Recognition*, Jun. 2013, pp. 399–406.

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

[7] Y. Tsaig and D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

[8] H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," in *IEEE Int. Conf. Comp. Vision*, Nov. 2011, pp. 770–777.

[9] J. Huang, X. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," in *IEEE Conf. Comp. Vision Pattern Recognition*, Jun. 2008, pp. 1–8.

[10] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 372–386, Feb. 2012.

[11] Q. Shi, A. Eriksson, A. Van den Hengel, and C. Shen, "Is face recognition really a compressive sensing problem?," *IEEE Conf. Comp. Vision Pattern Recognition*, vol. 0, pp. 553–560, 2011.

[12] D. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *IEEE Int. Conf. Comp. Vision*, Nov. 2011, pp. 471–478.

[13] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[14] D. L. Donoho and Y. Tsaig, "Fast solution of l1-norm minimization problems when the solution may be sparse.," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, 2008.

[15] Aleix Martínez and Robert Benavente, "The AR face database," Tech. Rep. 24, Computer Vision Center, Bellatera, Jun. 1998.

[16] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.