Sparse Representation for Face Recognition and Image Alignment

Ahana Gangopadhyay
ahana@wustl.edu

Abstract

Sparse representation has gained a large momentum in many areas of computer vision, particularly for pattern recognition tasks like image classification, because of its ability to naturally discriminate between classes. Here, we review its application in face recognition by analyzing and implementing a sparse coding algorithm known as Sparse Representation based Classification (SRC). We also examine whether the choice of features is critical for sparse representation using the public Labeled Faces in the Wild (LFW) dataset. It is seen that eigenfaces perform much better than merely down-sampling the images, contrary to the theory of sparse representation. Finally, we analyze that SRC gives a poor accuracy on the images because it cannot handle misalignments between the training and test images, and briefly describe an alignment algorithm called Robust Alignemnt by Sparse Representation (RASR) that can be merged seamlessly with SRC for performing transformation-invariant face recognition.

1 Introduction

Human faces are the most extensively studied object in image-based recognition. This partly stems from an attempt to replicate the extraordinary face recognition ability of the human visual system, and partly from the innumerable important applications for face recognition technology spanning from access control to multimedia processing. Face recognition has unique advantages over other biometric modalities in that it is universally accepted, well-understood and can be acquired passively, and acquisition devices are cheap and heavily commercialized.

As a computer vision problem, face recognition essentially consists of two components: a) feature extraction, and b) Classifier construction and label prediction. Commonly, Nearest Neighbor (NN) [5] or Nearest Subspace (NS) [10] methods are used for classification. However, NN predicts the label of the test image by only using its nearest neighbor in the training data, which is extremely vulnerable to noise. NS, on the other hand, approximates the test image by using all images belonging to the same class, and assigns it to the class which minimizes the reconstruction error. This may not work well when the classes are highly correlated to each other. To overcome these problems, Wright et. al. proposed a face recognition algorithm in [21] that uses a sparse coding framework for classifying face images that is naturally discriminative.

The idea behind this lies in the observation that natural images can be sparsely encoded by structural primitives. On the other hand, in statistical signal processing, representing a signal w.r.t. an over-complete dictionary of base elements or signal atoms have seen a recent surge of interest, further spurred on by the discovery that when the optimal representation is sparse enough, it can be efficiently computed by convex optimization. Combining these two ideas, Wright et.al. exploited the discriminative nature of sparse representation to perform classification of face images in their classical paper [21]. The test sample was represented as a sparse linear combination of the training samples themselves, eliminating the need for a generic dictionary, and offering natural discrimination between classes by seeking the sparsest representation. In this work, I implemented the SRC algorithm for face recognition using the Labeled Faces in the Wild (LFW) dataset.

A challenging issue in face recognition is how to address pose variations. Although SRC has achieved impressive recognition performances with public databases, it does not deal with misalignments between training and testing images. The problem of face recognition across pose variations can be solved by using approaches that can extract pose-tolerant features, or real-view based matching that capture and store multiple real views to exhaust all possible poses for the face recognition tool. However, these methods are often intractable or impractical. A more feasible alternative is to synthesize virtual views to substitute the demand of real views from a limited number of images (or even a single one) through pose transformation in the 2-D space or face reconstruction and projection in the 3-D space. Several approaches have been used in literature to deal with 2-D pose alignment methods. These include parallel deformation to generate virtual views covering a set of possible poses from a single example view using feature-based 2-D warping [2], Active Shape Models (ASM) [4] to represent faces by connected point distributions of locations of facial components which are then adjusted to the incoming images, and Active Appearance

Models (AAM) [3] to simultaneously model shape variations represented by point distributions and textures represented by pixel intensities. While these model-based techniques are advantageous in dealing with variations in expressions and pose, they often add unnecessary complexity in applications where inter-image variations can be described reasonably well by a finite set of deformations. [19] takes this approach by focusing on deformations having few degrees of freedom, e.g. similarity transformations, using the training images themselves as the appearance model.

The paper is organized as follows. Section 2 summarizes the related work in computer vision community in the domains of face recognition and image alignment which exploit sparsity in image representation. Section 3 describes the proposed approach by first outlining the face recognition algorithm using Sparse Representation based Classification (SRC). It also describes the role of features in sparse representation and analyzes why SRC is a poor choice when the face images are not aligned well. An elegant image alignment algorithm called Robust Alignment via Sparse Representation (RASR) is then presented, that can be merged seamlessly with SRC and exploits similar sparsity constraints as SRC to find optimal transformations to align a test image to the training set. Section 4 briefly describes the LFW dataset and presents the results I obtained using two different feature extraction techniques.

2 Related work

2.1 Sparse Representation in Face Recognition

Sparsity provides a powerful prior in a wide range of computer vision problems for inference with high-dimensional visual data that have intricate low-dimensional structures embedded in them [20]. Sparse representation has been exploited in harnessing the semantics of the data in vision tasks including face recognition [21], image super-resolution [22], motion and data segmentation [7], supervised de-noising and image classification [14] among others.

The work proposed in [20] boosted research in sparsity based face recognition. Gao et. al. [8] proposed the Kernel Sparse Representation (KSR) that uses the sparse coding technique in a high-dimensional feature space where non-linearly separable features can be linearly separated, leading to a better accuracy and lower reconstruction error. Yang et. al. in [23] used Gabor features for SRC with a learned Gabor occlusion dictionary that results in a compressed dictionary, reducing computational cost in coding occluded face images while improving the accuracy. [6] discussed the l^1 graph for classification by encoding each datum as a sparse representation of the remaining samples, and automatically selecting the most informative neighbors for each datum. [24] argued that SRC needs a considerable number of training samples to make the dictionary over-complete, instead proposing the Collaborative Representation based Classification (CRC) where training samples belonging to all possible classes contribute to the reconstruction process.

2.2 Sparse Representation in Image alignment

Sparse coding demonstrates a surprising robustness to noise and occlusions, but is highly sensitive to misalignment between the training and test sets. The first face alignment algorithm proposed in the context of sparse coding and SRC was [12], which projected an unaligned target image to random-projection manifolds defined by the model images and an affine transformation model. Each projection was then separated into the aligned projection target and a residue, and the aligned target was iteratively optimized by minimizing the residue. However, the test face was aligned to all the subjects simultaneously, making the algorithm more prone to local minima when the number of subjects increases. [19] eradicates this problem by finding the optimal transformation for aligning the test image to each training subset separately, followed by sparse representation based classification. Moreover unlike [12] which computes the deformation step by minimizing the l^2 norm of the error at each step of registration, [19] seeks a deformation step by enforcing sparsity even to the registration error, so that the transformed test image differs from the training subsets only for a small number of pixels.

There have been numerous other efforts in the domain of image alignment which exploit sparse coding in one way or another. [15] proposed Robust Alignement by Sparse and Low-rank decomposition (RASL) for a batch of linearly correlated images which seeks an optimal set of image domain transformations such that the matrix of transformed images can be decomposed as the sum of a sparse matrix of errors and a low-rank matrix of aligned images. However, to align a new test image to previously aligned ones, RASL has to adjust all the previous transformations to seek the matrix rank minimization. On the other hand, [13] arranged the input images into a 3-D tensor, claiming that optimally registered images can be deeply sparsified in the in the gradient domain and the frequency domain, with the separation of a sparse tensor of errors.

All the approaches outlined above are offline techniques, that are memory- and time-consuming when it comes to aligning increasing number of images. Online image alignment is gradually gaining popularity in order to address the problem of aligning dynamically increasing images. Methods like ORIA (Online Robust Image Alignment) and t-GRASTA (transformed Grassmannian Robust Adaptive Subspace Tracking Algorithm) are online image alignment approaches proposed to improve the scalability of RASL. [17] integrated the geometric transformation into the online Robust Principal Component Analysis (RPCA) approach for image alignment. However, all these methods assume that large errors like occlusion and corruption are sparse and separable with respect to image intensity. Real-world images are often corrupted by spatially varying intensity distortions, which can bias the subspace estimated from image intensities. [25] attempted to eradicate this problem by learning a low-dimensional subspace from image gradient orientations instead of pixel intensities and seeking alignment in the IGO domain.

3 Proposed approach

3.1 Sparse Representation based Classification

A classical method for solving face recognition problems, Sparse Representation based Classification (SRC) [21] attempts to project a face image as a sparse linear representation with respect to an over-complete dictionary of base images. The idea behind this is that the sparsest representation is naturally discriminative; among all subsets of basis vectors, it selects the subset that most compactly represents the test sample and rejects all other possible but less compact representations. This discriminative nature of sparse representation can be exploited to perform classification by representing the test sample with respect to an over-complete dictionary whose elements are the training samples themselves. If sufficient training samples are available from each class, the test sample can be well represented by only the training samples from its true class. The general framework for classification using sparse representation is outlined below.

In the context of face recognition, let us consider a pool of labeled training samples, each belonging to one of k different subjects. Each sample is a grayscale image of size $w \times h$, whose columns can be stacked into a vector $v \in \mathcal{R}^m,\ m = wh$. We can then arrange the n_i given training samples from the i-th class as columns of a matrix $A = [v_{i,1}, v_{i,2}, ..., v_{i,n_i} \in \mathcal{R}^{m \times n_i}]$. Assuming that samples from a single class lie on a linear subspace, and given sufficient training samples of the i-th object class, any test sample $y \in \mathcal{R}^m$ will approximately lie in the linear span of the training samples associated with the i-th class:

$$y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + ... + \alpha_{i,n_i}v_{i,n_i} \tag{1}$$

Since the membership of the test sample y is initially unknown, we define a new matrix A for the entire training set by concatenating the entire set of n training samples from all k subjects as

$$A = [A_1, A_2, ..., A_k] = [v_{1,1}, v_{1,2}, ..., v_{k,n_k}] \tag{2}$$

Then, the ideal linear representation of y can be written in terms of all the training samples as

$$y = Ax_0, \tag{3}$$

where $x_0 = [0, ..., 0, \alpha_{i,1}, ..., \alpha_{i,n_i}, 0, ..., 0]$ is a coefficient vector whose entries are zero except in the locations corresponding to the i-th class. In face recognition problems, usually $m < n$, so that the sn $y = Ax$ is underdetermined, and so, its solution is not unique. Conventionally, this difficulty can be solved by choosing the minimum l^2-norm solution

$$\hat{x_2} = argmin||x||_2, \text{ subject to } y = Ax \tag{4}$$

l^2 minimization, however, typically produces coefficient vectors that are dense or non-sparse, with large non-zero entries corresponding to training samples from many different classes. For a sparsest representation, we can instead minimize the l^0 norm, which counts the number of non-zero entries in a vector. However, this is an NP-hard problem difficult even to approximate. Instead, we can exploit the fact that if the solution sought is sparse enough, the solution to an l^0 minimization problem is equal to the solution of an l^1 minimization problem as follows

$$\hat{x_1} = argmin||x||_1, \text{ subject to } y = Ax. \tag{5}$$

For the current work, I solved an equivalent minimization problem of the following form

$$\hat{x_1} = \arg\min_x ||y - Ax||_2^2 + \lambda||x||_1, \tag{6}$$

where λ is the regularization parameter. In the ideal case, non-zero entries in the optimal estimate will all be associated with the columns of A belonging to a single object class i, to which the test sample can then be assigned. However, noise and modeling errors may lead to small non-zero entries associated with multiple object classes. In order to harness the subspace structure associated with images in face recognition, a classification scheme can then be designed that assigns the test sample to the class whose training samples can reconstruct the test image with the minimum residual error. For this, we define $\delta_i \in \mathcal{R}^n \to \mathcal{R}^n$ to be the characteristic function that selects the coefficients of the coefficient vector $\hat{x_1}$ associated with the i-th class. For each class i, we can then approximate the given test sample y as $\hat{y_i} = A\delta_i(\hat{x_1})$, and classify y by assigning it to the class that minimizes the residual between y and $\hat{y_i}$

$$class(y) = \arg\min_i ||y - A\delta_i(\hat{x_1})||_2. \tag{7}$$

3.2 The Role of Features

Conventionally, in computer vision literature, different feature extraction schemes are used in conjunction with classifiers primarily for the purpose of finding projections that better separate different object classes in a lower dimensional space. For face recognition tasks, these are often holistic face features such as Eigenfaces [18], Fisherfaces [1] and Laplacianfaces [9], or partial facial features, like patches around the eyes and nose [16]. According to [21], if sparsity in the recognition problem is properly harnessed, the choice of features is no longer critical, provided the number of features is sufficiently large and the sparse representation is correctly computed. It was shown using a number of experiments with the Extended Yale B database that even random projections or down-sampled images performed as well as carefully engineered features like Eigenfaces and Fisherfaces.

For this work, I experimented with two types of features: downsampled images and Eigenfaces. Eigenfaces are nothing but the eigenvectors derived from the covariance matrix of the high-dimensional vector space of face images, and themselves form a basis set of all images used to construct the covariance matrix. Principal component analysis shows that only a small number of eigenfaces can capture most of the variance in the faces, and hence can approximate most faces reasonably well. Eigenfaces are therefore a standard method in face recognition to reduce the dimensionality of face images. A number of experiments were conducted by projecting both the training and test sets on different number of eigenfaces.

3.3 Image Alignment

Classic face recognition algorithms like SRC work well on public databases, but cannot handle real recognition tasks which involve variations in illumination, misalignment and occlusions. Face image differences caused by variations in pose are often larger than differences across subjects in a database, leading to a considerable degradation in accuracy. In this section, I will analyze a simple face alignment algorithm that can be merged seamlessly with SRC, called Robust Alignment by Sparse Representation (RASR) [19].

RASR assumes that the training images are well-aligned, while test images are only loosely controlled. It also considers that test images are corrupted by deformations with very few degrees of freedom, i.e., similarity transformations, and poses the image alignment problem as the search for an optimal set of transformations that minimizes the l^1 norm of the representation error. If a test image y_0 is well-aligned to the training set, it can be represented as a sparse linear combination Ax of all the images in the database, plus a sparse error e due to corrupted pixels. The sparse representation can then be recovered by the following l^1 minimization problem

$$min_{x,e}||x||_1 + ||e||_1, \text{ subject to } y_0 = Ax + e \tag{8}$$

If y_0 is subjected to some pose or misalignment, we actually observe $y = y_0 o \tau^{-1}$, for some transformation $\tau \in T$, where T is a finite-dimensional group of transformations acting on the image domain. The transformed image can then no longer be sparsely represented by images belonging to the true class,so that SRC is not valid. If the true deformation τ^{-1} can be determined, its inverse can be applied to retrieve the unwarped image and find its sparse representation again. The idea behind finding the optimal transformation τ is therefore to seek a transformation that allows the sparsest representation of the corresponding aligned image

$$\hat{\tau} = \underset{x,e,\tau \in T}{\arg\min} ||x||_1 + ||e||_1, \text{ subject to } yo\tau = Ax + e \tag{9}$$

This simultaneous optimization over the coefficients x, the error e and transformation τ is a non-convex optimization problem, rendered difficult by the presence of many local minima that correspond to aligning y to different subjects. A more appropriate choice is to find the best alignment of the test image for each subject i

$$\hat{\tau_i} = \underset{x,e,\tau_i \in T}{\arg\min} ||e||_1, \text{ subject to } yo\tau_i = A_i x + e, \tag{10}$$

where A_i is the sub-dictionary of face images belonging only to the i-th subject. It is to be noted that sparsity w.r.t. x need not be enforced any more in the cost function, since A_i consists of face images from a single subject. Although Eq. 10 is still non-convex, we can start from a good initial guess of transformations (e.g. from the output of a face detector), and repeatedly refine this initialization by linearizing about the current estimate of τ as follows

$$yo\tau + J\Delta\tau = A_i x + e, \tag{11}$$

where $J = \frac{\partial}{\partial\tau} yo\tau$ is the Jacobian of $yo\tau$ w.r.t. the transformation parameters τ, and $\Delta\tau$ is the step in τ. When the test image is correctly aligned to the i-th training set, it is expected to differ from $A_i x$ only for a minority of pixels, and hence at each step, we seek a deformation step $\Delta\tau$ that best sparsifies the registration error e (instead of minimizing it via l^2 norm) as follows

$$\Delta\hat{\tau}_1 = \underset{x,e,\Delta\tau_i \in T}{\arg\min} ||e||_1, \text{ subject to } yo\tau + J\Delta\tau = A_i x + e, \tag{12}$$

Once the best transformation τ_i has been computed for each subject i, each training set can be aligned to y by mapping the dictionary $A \leftarrow [A_1 o\tau_1^{-1}|A_2 o\tau_2^{-1}|...|A_k o\tau_k^{-1}]$, and a global sparse representation of the form in Eq. 6 can be solved as before to assign the test image to the class that can reconstruct the

Figure 1: Sample faces from the LFW database after face detection.

sample with the minimum residual error. Additionally, the alignment problem produces per-subject alignment residuals $||e||_1$ which can be used to prune unpromising candidates from the global optimization problem, leading to a more tractable set of labels to choose from. Usually, 2D similarity or projective transformations are taken as T. An important point to keep in mind is that, apart from normalizing the training images in the beginning, the warped testing images should be normalized in every iteration of the alignment algorithm as $y \leftarrow \frac{y o \tau}{||y o \tau||_2}$, to steer away from degenerate global minima.

4 Experimental Results

4.1 Labeled Faces in the Wild Dataset

Labeled Faces in the Wild [11] is a publicly available database of face photographs designed to study the problem of unconstrained face recognition. It offers unique challenges for face recognition problems because it exhibits great variations in pose, and changes in illumination and occlusion. For this work, I chose 19 subjects from the database, each having 35 images, and randomly selected 20 for training and the remaining 15 for testing. Faces were detected from all the images using the Viola-Jones detector and cropped to a size of 60x80 pixels. A few sample face images belonging to different subjects, after face detection and cropping, are given in Fig. 1.

SRC was conducted with two sets of features: down-sampled images and eigenfaces. For the latter, eigenfaces were constructed from the pool of training images and both training and testing sets were projected on different number of eigenfaces to compute features of different sizes. The first 24 eigenfaces determined from the LFW database are given in Fig. 2. The results on LFW database for both sets of features, followed by Sparse Representation based Classification, for different dimensionalities of extracted features, corresponding to $\lambda = 0.001$, are given in Table 1. We can see that in all cases, eigenfaces give a much more discriminative set of features compared to merely down-sampling the images. Thus, contrary to [21], we can deduce that for difficult face databases with misalignments, occlusions and variations in pose and illumination, the choice of features is indeed critical even in the case where sparsity in the recognition problem is properly harnessed.

Feature set	160-D	200-D	240-D
Down-sampled images	36.84%	39.29%	29.82%
Eigenfaces	55.78%	56.14%	58.59%

Table 1: Recognition accuracy on the FFW dataset using two different feature sets for different feature dimensionalities.

5 Conclusion

This paper briefly summarizes the ongoing research efforts using sparse coding to solve face recognition and image alignment tasks and in particular, analyzes the papers [21] and [19]. It also implements the SRC algorithm for face recognition with the LFW dataset using down-sampled images and eigenfaces. It is seen that SRC, while being a powerful face recognition tool, do not perform well with LFW because of misaligned images, although eigenfaces produce more discriminative features compared to down-sampling. Lastly, we see that image

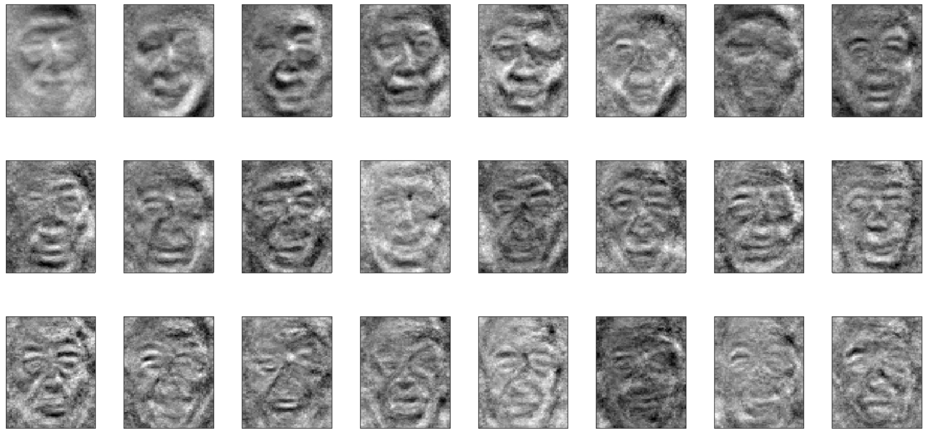

Figure 2: The first 24 eigenfaces from the LFW dataset.

alignment prior to performing SRC has the potential to improve the classification accuracy.

Acknowledgments

I thank the course instructor Dr. Ayan Chakrabarti and the TAs Abby Stylianou and Jarett Gross for their constant help and support in compiling this work and throughout the course.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997.

[2] D. Beymer and T. Poggio. Face recognition from one example view. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 500–507. IEEE, 1995.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. John Wiley & Sons, 2012.

[6] D. C. Duro, S. E. Franklin, and M. G. Dubé. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using spot-5 hrg imagery. *Remote Sensing of Environment*, 118:259–272, 2012.

[7] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2790–2797. IEEE, 2009.

[8] S. Gao, I. W.-H. Tsang, and L.-T. Chia. Kernel sparse representation for image classification and face recognition. In *European Conference on Computer Vision*, pages 1–14. Springer, 2010.

[9] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. Face recognition using laplacianfaces. *IEEE transactions on pattern analysis and machine intelligence*, 27(3):328–340, 2005.

[10] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *Computer vision and pattern recognition, 2003. Proceedings. 2003 IEEE computer society conference on*, volume 1, pages I–I. IEEE, 2003.

[11] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[12] J. Huang, X. Huang, and D. Metaxas. Simultaneous image transformation and sparse representation recovery. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[13] Y. Li, C. Chen, F. Yang, and J. Huang. Deep sparse representation for robust image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4894–4901, 2015.

[14] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[15] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2233–2246, 2012.

[16] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B. Vijayakumar. Partial & holistic face recognition on frgc-ii data using support vector machine. In *Computer Vi-*

sion and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on, pages 48–48. IEEE, 2006.

[17] W. Song, J. Zhu, Y. Li, and C. Chen. Image alignment by online robust pca via stochastic gradient descent. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(7):1241–1250, 2016.

[18] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

[19] 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 Transactions on Pattern Analysis and Machine Intelligence*, 34(2):372–386, 2012.

[20] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, 2010.

[21] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2009.

[22] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[23] M. Yang and L. Zhang. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. *Computer Vision–ECCV 2010*, pages 448–461, 2010.

[24] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang. Collaborative representation based classification for face recognition. *arXiv preprint arXiv:1204.2358*, 2012.

[25] Q. Zheng, Y. Wang, and P. A. Heng. Online robust image alignment via subspace learning from gradient orientations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1753–1762, 2017.