Close the Loop: Joint Blind Image Restoration and Recognition with Sparse Representation Prior

Haichao Zhang[†‡], Jianchao Yang[‡], Yanning Zhang[†], Nasser M. Nasrabadi[§] and Thomas S. Huang[‡]
[†] School of Computer Science, Northwestern Polytechnical University, Xi'an China
[‡] Beckman Institute, University of Illinois at Urbana-Champaign, IL USA
[§] U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD USA
[‡]{hczhang,jyang29,huang}@ifp.uiuc.edu [†]ynzhang@nwpu.edu.cn [§]nasser.m.nasrabadi.civ@mail.mil

Abstract

Most previous visual recognition systems simply assume ideal inputs without real-world degradations, such as low resolution, motion blur and out-of-focus blur. In presence of such unknown degradations, the conventional approach first resorts to blind image restoration and then feeds the restored image into a classifier. Treating restoration and recognition separately, such a straightforward approach, however, suffers greatly from the defective output of the ill-posed blind image restoration. In this paper, we present a joint blind image restoration and recognition method based on the sparse representation prior to handle the challenging problem of face recognition from low-quality images, where the degradation model is realistic and totally unknown. The sparse representation prior states that the degraded input image, if correctly restored, will have a good sparse representation in terms of the training set, which indicates the identity of the test image. The proposed algorithm achieves simultaneous restoration and recognition by iteratively solving the blind image restoration in pursuit of the sparest representation for recognition. Based on such a sparse representation prior, we demonstrate that the image restoration task and the recognition task can benefit greatly from each other. Extensive experiments on face datasets under various degradations are carried out and the results of our joint model shows significant improvements over conventional methods of treating the two tasks independently.

1. Introduction

In many real world applications, such as video surveillance, the target of interest in the captured image usually suffers from low qualities, such as low resolution due to the long distance of the target, motion blur due to the relative motion between the target and the camera, and out-of-focus blur if the the target is not in the focus of the capture device, or even some complex combinations of these factors. In such practical scenarios, it will present a big challenge to perform many high level vision tasks such as recognition.

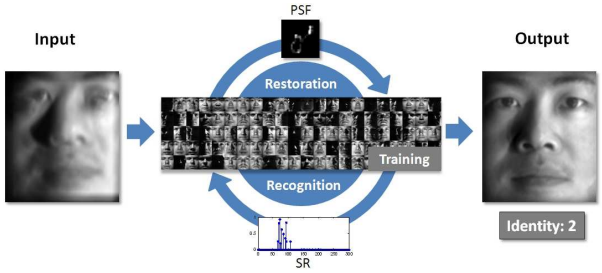

Figure 1. Sparse Representation based Joint Blind Restoration and Recognition (JRR) framework. Given a blurry observation, JRR iteratively estimates the PSF and the underlying identity based on the sparse representation prior. The algorithm will output the estimated PSF, a deblurred image, and the identity of the observation.

A natural solution to this problem would be to first perform image restorations to obtain an image with better quality [6, 14, 19], and then feed the restored result into a recognition system. Such a straightforward approach has the problem that many restoration algorithms are designed for improving human visual perception only, rather than machine perception, thus there is no guarantee of recognition improvements. Even worse, when the degradation model is unknown, general purpose restoration schemes, such as deblurring, do not perform well on some realistic images that do not exhibit strong edge structures, such as faces, and will typically introduce severe artifacts that actually deteriorate the recognition performance. Instead of restoring the test image, another approach could be to estimate the degradation model first, use it to transform the training images, and then compare the input test image with the synthetically generated training set. This method generally works better than the previous one. But for many realistic data whose degradation model is very complex, it may easily fail.

Image deblurring is a long-standing restoration prob-

lem in image processing and computer vision communities [6, 14]. Recent works have shown that it is possible to estimate both the blur kernel and the latent sharp image with high quality from a single blurry observation [6, 14, 11, 1, 2]. However, these methods rely on the key assumption of the existence of strong edge structures in the latent image, which facilitates the algorithms to find a meaningful local minimum [11]. In situations of few strong edge structures, *e.g.*, face images, these methods may fail. Although much progress has been made on pure image restoration, only few works have studied the impacts of restoration on recognition, or vice versa, the effects of recognition on restoration. The method in [3] alternated between recognition and restoration to change the patch sampling prior using non-parametric belief propagation for digit recognition, with the assumption of a known image blur model. Hennings-Yeomans *et al*. [9] proposed a method to extract features from both the low-resolution faces and their super-resolved ones within a single energy minimization framework. Nishiyama *et al*. [12] proposed to improve the recognition of blurry faces with a pre-defined finite set of Point Spread Function (PSF) kernels. However, these methods only deal with some simple image degradations.

We present in this paper a Joint image Restoration and Recognition (JRR) approach based on the sparse representation prior for face images, to handle the challenging task of face recognition from low-quality images in a blind setting, *i.e.*, with no *a priori* knowledge on the blur kernels, which can be non-parametric and very complex (Figure 1). We assume that we have sharp and clean training face images for all the test subjects, and the degraded test image, if correctly restored, can be well represented as a linear combination of the training faces from the same subject up to some sparse errors, thus leading to a sparse representation in terms of all the training faces. When the test subject is not present in the gallery, it will violate our sparse representation assumption, and in principle the test subject can be rejected via a similar approach as in [18], which is not considered in the current paper. With such a sparse representation prior, the proposed method connects restoration and recognition in a unified framework by seeking sparse representations over the training faces via ℓ_1-norm minimization. On one hand, a better restored image can be better represented by the images from the same class, leading to a sparser representation in terms of the training set, thus facilitating recognition; on the other hand, a better resolved sparse representation, which implies better recognition ability, can give a more meaningful regularization in the solution space for blind restoration. Our approach iteratively restores the input image by searching for the sparsest representation, which can correct the initial possibly erroneous recognition decision and recognize the person's identity with increasing confidence.

The rest of the paper is organized as follows. Section 2 summarizes the role of sparsity in both image restoration and recognition. Section 3 proposes our JRR framework and presents an efficient optimization procedure to find the solution. Experiments on face datasets under various realistic degradations are carried out in Section 4. Finally, we make some discussions and conclude the paper in Section 5.

2. Sparse Representation in Restoration and Recognition

In this section, we briefly introduce the basics of sparse representation and summarize its applications on both ill-posed inverse image restoration and pattern recognition.

2.1. Sparse Representation

Sparse representation modeling of data assumes an ability to describe signals as a linear combination of a few atoms from a pre-specified dictionary. Formally, given a signal $\mathbf{x} \in \mathbb{R}^m$ and a dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_n] \in \mathbb{R}^{m \times n}$, where typically $m \leq n$, we can recover a sparse representation ($\epsilon = 0$) or sparse approximation ($\epsilon > 0$) $\hat{\boldsymbol{\alpha}}$ for $\mathbf{x}$ by:

$$\begin{aligned} &\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \\ &\text{s.t. } \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \leq \epsilon. \end{aligned} \tag{1}$$

The model tries to seek the most compact representation for the signal $\mathbf{x}$ given the dictionary $\mathbf{D}$, which can be orthogonal basis ($m = n$), over-complete basis ($m < n$) [1] or dictionary learned from the training data [19]. For orthonormal basis, solution to (1) is merely the inner products of the signal with the basis. However, for general dictionary (non-orthogonal and over-complete), the optimization for (1) is combinatorially NP-hard. Recent works show that, this NP-hard problem can be tackled by replacing the non-convex ℓ_0-norm with ℓ_1-norm under some mild conditions [4], which makes the objective function convex while exact solution can still be guaranteed. Using the Lagrange multiplier, we can reformulate the relaxed ℓ_1-problem as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{x}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1. \tag{2}$$

Sparsity plays an important or even crucial role in many fields, such as image restoration [5, 1, 19], compressive sensing, and recognition [17, 18]. In the following, we will make a brief discussion on the role of sparsity in both image restoration and pattern recognition.

2.2. Sparsity in Image Restoration

A close inspection of the progress made in the field of image processing in the past decades reveals that much of it can be attributed to better modeling of the image content [5]. Sparsity is arguably the most widely used prior

for image restoration, such as image denoising, inpainting, super-resolution and deblurring [5]. Among these, we specifically focus the discussion on image deblurring.

Image blurring is a widely existing degradation factor in the real life imaging process (*e.g.*, surveillance), possibly resulting from defocusing, relative motion between the object and the camera, to name a few [6, 14], which may bring severe adverse impacts on both human perception and machine perception (*e.g.*, classification). Assuming convolutional blur model and additive white Gaussian noise, the low quality image observation process can be modeled as [6]:

$$\mathbf{y} = \mathbf{k} * \mathbf{x} + \varepsilon, \tag{3}$$

where $\mathbf{k}$ is the PSF (blur kernel) and $*$ denotes the convolution operator. The problem of (blind) deblurring is to estimate both the latent sharp image $\mathbf{x}$ and the blur kernel $\mathbf{k}$ from the blurry and noisy observation $\mathbf{y}$. With more unknowns than knowns, this is a typical ill-posed inverse problem, thus requiring regularization to stabilize the solution:

$$\{\hat{\mathbf{x}}, \hat{\mathbf{k}}\} = \arg\min_{\mathbf{x},\mathbf{k}} \|\mathbf{k} * \mathbf{x} - \mathbf{y}\|_2^2 + \lambda\rho(\mathbf{x}) + \gamma\varrho(\mathbf{k}), \tag{4}$$

where $\rho(\mathbf{x})$ is a regularization term on the desired image, and $\varrho(\mathbf{k})$ regularizes the possible blur kernels, typically an ℓ_2-norm penalty [2]. Most of the current restoration methods can be cast into such a regularization framework where the regularization terms based on image prior are crucial for obtaining better restoration results and are related somehow with the sparse property of natural images [6, 14, 10, 19, 1, 5]. With the sparsity prior as regularization, we can arrive at the following formulation:

$$\{\hat{\mathbf{x}}, \hat{\mathbf{k}}\} = \arg\min_{\mathbf{x},\mathbf{k}} \|\mathbf{k} * \mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{D}^\top\mathbf{x}\|_1 + \gamma\|\mathbf{k}\|_2, \tag{5}$$

where $\mathbf{D}^\top$ is some sparse transformation (such as Wavelet, Curvelet, among others [1]) or sparsity inducing operator (such as handcrafted derivative filters or filters learned from training images [6, 14, 10, 13]). When $\mathbf{D}$ is orthonormal, we have $\boldsymbol{\alpha} = \mathbf{D}^\top\mathbf{x}$ as the transform coefficients, and thus we can rewrite Eqn. (5) as:

$$\{\hat{\boldsymbol{\alpha}}, \hat{\mathbf{k}}\} = \arg\min_{\boldsymbol{\alpha},\mathbf{k}} \|\mathbf{k} * \mathbf{D}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 + \gamma\|\mathbf{k}\|_2^2. \tag{6}$$

To achieve better sparsity for the representation $\boldsymbol{\alpha}$, $\mathbf{D}$ can be generalized to be non-orthogonal and over-complete, by either combining different orthonormal basis or learning from the data [19]. In this paper, we model $\mathbf{D}$ as the training data directly, which closely relates the solution $\hat{\boldsymbol{\alpha}}$ with recognition as we will discuss in the following.

2.3. Sparse Representation for Recognition

The application of sparse representation for classification is based on the assumption that data samples belonging to the same class live in the same subspace of a much lower dimension, thus a new test sample can be well represented by the training samples of the same class, which leads to a natural sparse representation over the whole training set. Casting the recognition problem as one of finding a sparse representation of the test image in terms of the training set as a whole up to some sparse errors due to occlusion, Wright *et al*. [18] showed that such a simple sparse representation based approach is robust to partial occlusions and can achieve promising recognition accuracy on public face datasets. This idea is further extended in their later work [15] to handle face misalignment. Formally, given a set of training samples for the c-th class $\mathbf{D}_c = [\mathbf{d}_{c,1}, \mathbf{d}_{c,2}, \cdots]$, a test sample $\mathbf{x}$ from class c can be well represented by $\mathbf{D}_c$ with coefficients $\boldsymbol{\alpha}_c$. As the label for $\mathbf{x}$ is unknown, it is assumed that $\boldsymbol{\alpha}_c$ can be recovered from the sparse representation of $\mathbf{x}$ in terms of the dictionary constructed from training samples of all C classes by

$$\begin{aligned} \hat{\boldsymbol{\alpha}} = &\arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \\ &\text{s.t. } \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{x}\|_2^2 \le \epsilon, \end{aligned} \tag{7}$$

where $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_C], \boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top, \cdots, \boldsymbol{\alpha}_C^\top]^\top$. Then the label for the test sample $\mathbf{x}$ is determined as the class which gives the minimum reconstruction error:

$$\hat{c} = \arg\min_c \|\mathbf{D}\delta_c(\hat{\boldsymbol{\alpha}}) - \mathbf{x}\|_2^2 = \arg\min_c \|\mathbf{D}_c\hat{\boldsymbol{\alpha}}_c - \mathbf{x}\|_2^2.$$

$\delta_c(\cdot)$ is an indicator function keeping the elements corresponding to the c-th class while setting the rest to be zero.

3. Joint Blind Restoration and Recognition with Sparse Representation Prior

In this section, we present our joint restoration and recognition framework in the blind situation, *i.e.*, no *a priori* information on the image degradation process about the blurry query image is available, and develop an efficient minimization algorithm to solve the problem.

3.1. Problem Formulation

In conventional recognition works, the test image $\mathbf{y}$ is often assumed to be captured under ideal condition without any degradation, *i.e.* $\mathbf{y} = \mathbf{x}$. Some simple environmental variations, such as illumination and mild misalignment, can be fairly well handled given enough training samples [15]. In reality, however, we may only get observation $\mathbf{y}$ for $\mathbf{x}$ with degradations, *e.g.*, blur as in (3), which are hard to model beforehand and can bring serious problems to the recognition task. Therefore, recognition from a single blurry observation is a very challenging task, especially in the case of blind situation (dubbed as *blind recognition*), *i.e.*, no *a priori* information is available for the observation process. As far as we know, few works have been done on

this challenging blind recognition problem. In this work, we aim to address the task of blind recognition by exploiting the interactions between restoration and recognition with the sparse representation prior. Formally, given the blurry observation $\mathbf{y}$, and the sharp training image set $\mathbf{D}$, we want to estimate the latent sharp image $\mathbf{x}$, blur kernel $\mathbf{k}$, as well as the class label c simultaneously by

$$\{\hat{\mathbf{x}}, \hat{\mathbf{k}}, \hat{c}\} = \arg\min_{\mathbf{x},\mathbf{k},c} \mathbf{E}(\mathbf{x}, \mathbf{k}, c), \tag{8}$$

where

$$\begin{aligned}\mathbf{E}(\mathbf{x}, \mathbf{k}, c) = &\|\mathbf{k} * \mathbf{x} - \mathbf{y}\|_2^2 + \eta\|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \\ &+ \tau\sum_{l=1}^{L}|\mathbf{e}_l * \mathbf{x}|^s + \gamma\|\mathbf{k}\|_2^2.\end{aligned} \tag{9}$$

We explain each term of the model in detail as follows.

1. The first term is the conventional reconstruction constraint, *i.e.*, the restored image should be consistent with the observation with respect to the estimated degradation model.
2. The second term means the recovered sharp image can be well represented by the clean training set.
3. The third term enforces that the representation of the recovered image in terms of the training set should be sparse. In other words, the algorithm favors a solution $\mathbf{x}$ that can be sparsely represented by the training set. Meanwhile, this sparse representation also recognizes the identity of the observation.
4. The fourth term is a general sparse prior for natural images using sparse exponential of the responses of derivative filters to further stabilize the solution, where typically $0.5 \leq s \leq 0.8$.
5. The last term is merely a ℓ_2-norm stable regularization for the blur kernel.

The basic idea of the model is that the restored image should have a sparse representation in terms of the training images if the blur kernel is correctly estimated, and meanwhile the sparse representation itself identifies the observed target. On one hand, the sparse representation prior effectively regularizes the solution space of the possible latent images and blur kernels; on the other hand, better estimated blur kernel will promote better sparse representations for recognition. As shown by Eqn. (9), our model unifies the restoration (6) and recognition (7) in a unified framework based on the sparse representation prior. Note that the proposed model is a general framework which can handle different kinds of image degradations, *e.g.*, out-of-focus blur, various motion blurs, translation misalignment, and etc., which can be modeled by a linear operator.

3.2. Optimization Procedure

The proposed model (9) involves multiple variables and is hard to minimize directly. We adopt the alternating minimization scheme advocated by recent sparse optimization and image deblurring works [16, 14, 10, 2], which reduces the original problem into several simpler subproblems. Following this scheme, we address the subproblems for each of the optimization variables in an alternating fashion and present an overall efficient optimization algorithm. In each step, our algorithm reduces the objective function value, and thus will converge to a local minima. To start, we initialize the sparse representation $\hat{\boldsymbol{\alpha}}$ as that recovered from $\mathbf{y}$ with respect to $\mathbf{D}$, and the latent sharp image $\hat{\mathbf{x}}$ as $\mathbf{D}\hat{\boldsymbol{\alpha}}$.

3.2.1 Blur Kernel Estimation: Optimizing for $\mathbf{k}$

In this subproblem, we fix all other variables and and optimize the image blur kernel $\mathbf{k}$ by

$$\hat{\mathbf{k}} = \arg\min_{\mathbf{k}} \|\hat{\mathbf{x}} * \mathbf{k} - \mathbf{y}\|_2^2 + \gamma\|\mathbf{k}\|_2^2. \tag{10}$$

This is a least square problem with Tikhonov regularization, which leads to a close-form solution for $\mathbf{k}$:

$$\hat{\mathbf{k}} = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\hat{\mathbf{x}})} \circ \mathcal{F}(\mathbf{y})}{\overline{\mathcal{F}(\hat{\mathbf{x}})} \circ \mathcal{F}(\hat{\mathbf{x}}) + \gamma\mathbf{I}}\right),$$

where $\mathcal{F}(\cdot)$ denotes Fast Fourier Transform (FFT), $\mathcal{F}^{-1}(\cdot)$ denotes inverse FFT, $\overline{\mathcal{F}(\cdot)}$ denotes the complex conjugate of $\mathcal{F}(\cdot)$, and "$\circ$" denotes element-wise multiplication.

3.2.2 Latent Image Recovery: Optimizing for $\mathbf{x}$

Given the current kernel estimation $\hat{\mathbf{k}}$ and sparse representation $\hat{\boldsymbol{\alpha}}$, we want to update the estimation for the latent sharp image $\mathbf{x}$. The optimization problem (9) becomes

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x} * \hat{\mathbf{k}} - \mathbf{y}\|_2^2 + \eta\|\mathbf{x} - \mathbf{D}\hat{\boldsymbol{\alpha}}\|_2^2 + \tau\sum_{l=1}^{L}|\mathbf{e}_l * \mathbf{x}|^s. \tag{11}$$

This optimization problem can be solved efficiently with variable substitution and FFT [16, 14, 10, 2]. Introducing new auxiliary variables $\mathbf{u}_l (l \in 1, 2, \cdots, L)$, we can rewrite the energy function in (11) as:

$$\begin{aligned}\mathbf{E}(\mathbf{x}, \mathbf{u}) = &\|\mathbf{x} * \hat{\mathbf{k}} - \mathbf{y}\|_2^2 + \eta\|\mathbf{x} - \mathbf{D}\hat{\boldsymbol{\alpha}}\|_2^2 \\ &+ \tau\sum_{l=1}^{L}|\mathbf{u}_l|^s + \beta\sum_{l=1}^{L}\|\mathbf{u}_l - \mathbf{e}_l * \mathbf{x}\|_2^2,\end{aligned} \tag{12}$$

which can be divided into two sub-problems: $\mathbf{x}$-subproblem and $\mathbf{u}$-subproblem. In the $\mathbf{x}$-subproblem, the energy function to be minimized becomes

$$\mathbf{E}(\mathbf{x}) = \|\mathbf{x} * \hat{\mathbf{k}} - \mathbf{y}\|_2^2 + \eta\|\mathbf{x} - \mathbf{D}\hat{\boldsymbol{\alpha}}\|_2^2 + \beta\sum_{l=1}^{L}\|\mathbf{e}_l * \mathbf{x} - \mathbf{u}_l\|_2^2$$

which can be solved efficiently using FFT as:

$$\hat{\mathbf{x}} = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\hat{\mathbf{k}})} \circ \mathcal{F}(\mathbf{y}) + \eta\mathcal{F}(\mathbf{D}\hat{\boldsymbol{\alpha}}) + \beta\sum_{l=1}^{L}\overline{\mathcal{F}(\mathbf{e}_l)} \circ \mathcal{F}(\mathbf{u}_l)}{\overline{\mathcal{F}(\hat{\mathbf{k}})} \circ \mathcal{F}(\hat{\mathbf{k}}) + \eta\mathbf{I} + \beta\sum_{l=1}^{L}\overline{\mathcal{F}(\mathbf{e}_l)} \circ \mathcal{F}(\mathbf{e}_l)}\right).$$

In the $\mathbf{u}$-subproblem, $\mathbf{u}_l$ can be estimated by solving the following problem given fixed $\mathbf{x}$:

$$\hat{\mathbf{u}}_l = \arg\min_{\mathbf{u}_l} \tau|\mathbf{u}_l|^s + \beta\|\mathbf{u}_l - \mathbf{e}_l * \mathbf{x}\|_2^2, \tag{13}$$

which can be solved efficiently over each dimension separately [10]. In practice, we use first-order derivative filters $\{\mathbf{e}_1 = [1, -1], \mathbf{e}_2 = [1, -1]^\top\}$ and set $s = 0.5$ as [10]. We follow the multi-scale estimation scheme for stable estimations of the blur kernel $\mathbf{k}$ and latent sharp image $\mathbf{x}$ as in [6, 14, 2]. Conventional schemes such as structure prediction have also been incorporated into optimization [2].

3.2.3 Sparse Projection: Optimizing for $\boldsymbol{\alpha}$

With the recovered kernel $\hat{\mathbf{k}}$ and sharp training set $\mathbf{D}$, we can generate the corresponding blurry dictionary $\mathbf{D}_\mathrm{b}$ via

$$\mathbf{D}_\mathrm{b} = \mathbf{D} * \hat{\mathbf{k}}, \tag{14}$$

where the convolution $*$ is performed on each column of $\mathbf{D}$ with $\hat{\mathbf{k}}$. Then the sparse representation vector $\boldsymbol{\alpha}$ can be updated by

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{D}_\mathrm{b}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1, \tag{15}$$

from which the classification decision is made using

$$\hat{c} = \arg\min_{c} \|\mathbf{D}_\mathrm{b}\delta_c(\hat{\boldsymbol{\alpha}}) - \mathbf{y}\|_2^2. \tag{16}$$

We do not use the deblurred image and the sharp training set to compute the sparse representation $\boldsymbol{\alpha}$ because the deblurring process may introduce artifacts which is disadvantageous for recovery and recognition. Based on compressive sensing theory, we can recover the sparse representation using the blurry observation $\mathbf{y}$ and the blurry dictionary $\mathbf{D}_\mathrm{b}$ and thus circumvent the above problem. The overall algorithm optimizes over blur kernel $\mathbf{k}$, latent sharp image $\mathbf{x}$, sparse representation $\boldsymbol{\alpha}$ and class label c alternatively. Algorithm 1 describes the procedures of our joint blind restoration and recognition algorithm.

Algorithm 1: Joint Blind Image Restoration and Recognition with Sparse Representation Prior.

Input: a blurry image $\mathbf{y}$, training image set $\mathbf{D}$
Output: estimated blur kernel $\hat{\mathbf{k}}$, restored image $\hat{\mathbf{x}}$, and the class label $\hat{c}$
Initialization: sparse vector $\hat{\boldsymbol{\alpha}}$ recovered from $\mathbf{y}$ in terms of $\mathbf{D}$, and $\hat{\mathbf{x}} = \mathbf{D}\hat{\boldsymbol{\alpha}}$.
for $t = 1, 2, \cdots, T$ **do**
- **Kernel Estimation:** update kernel $\mathbf{k}$ by minimizing Eqn.(10);
- **Image Estimation:** update the latent image $\mathbf{x}$ estimation via minimizing Eqn.(11);
- **Sparse Projection:** recovering the sparse coefficients $\boldsymbol{\alpha}$ by minimizing Eqn.(15);
- **Classification:** estimate the class label c from Eqn.(16).

4. Experiments and Results

In this section, we present several experiments to demonstrate the effectiveness of the proposed JRR method in terms of both restoration accuracy and recognition accuracy. The Extended Yale B [7](48×42) and CMU Multi-PIE [8](80×60) datasets are used for evaluation in this work. The Extended Yale B dataset contains 38 individuals, each with 64 near frontal view images under different illuminations. For CMU Multi-PIE dataset, We use the frontal images with neutral expression under varying illuminations from session 1 for computational considerations.

For restoration, we compare our algorithm with the fast deblurring method in [2], one of the *state-of-the-art* blind deblurring algorithms. Root Mean Square Error (RMSE) is employed to compare the estimation accuracy for both the blur kernel and the restored image. For classification, we compare our JRR algorithm with the following methods: (1) SVM: classification with linear SVM trained on the sharp training set; (2) SRC: directly feed the blurry observation into the sparse representation based classification algorithm [18]; and (3) SRC-B: first estimate the kernel and then generate a blurred training set for SRC.[1]

4.1. An Illustrative Example

We illustrate the proposed method with a simple example in Figure 2. Given a blurry observation, we jointly recover the blur kernel, the latent sharp image, and the class label in an iterative way. Figure 2 shows that, as the optimization iteration increases, the latent representation becomes sparser and sparser as indicated by the increase of Sparsity Concentration Index (SCI) measure[2], which implies that the underlying class label of the test image can be determined with increasing confidence. At the same time, the restored image resembles more and more to the ground truth as indicated by the decrease of the restoration error, which means

[1] Another approach is first to deblur the test image and then use the deblurred image for recognition. Empirically, we observe that this method may perform even worse than using the original blurry image directly, mainly due to the artifacts induced by the deblurring step (Figure 4), and thus we do not compare with this method in the sequel.

[2] SCI is defined as $\mathrm{SCI}(\mathbf{x}) = \frac{C \cdot \max_i \|\delta_i(\mathbf{x})\|_1 / \|\mathbf{x}\|_1 - 1}{C-1}$, where C is the total number of classes [18].

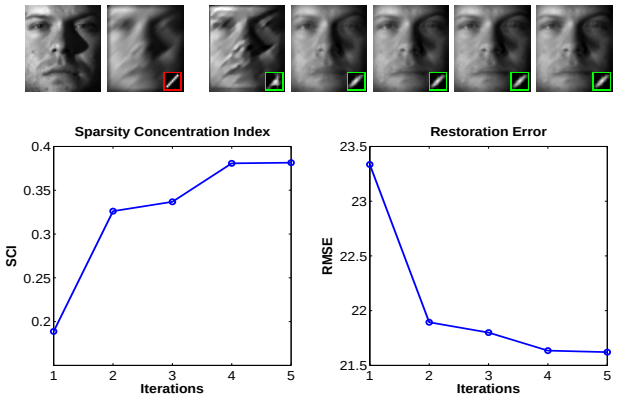

Figure 2. The joint blind restoration and recognition optimization process for 5 iterations. Top row, left to right: ground-truth sharp image, blurry test image, and the restored images from iteration 1 to 5. The ground truth and estimated PSFs are framed in red and green borders respectively. Bottom row, left: sparsity of the recovered sparse coefficients in terms of SCI; right: restoration errors in terms of RMSE.

that the estimated blur kernel gets more and more accurate. Actually, in the first iteration, the blurry input is wrongly assigned with class label of subject 4, while the ground truth label is subject 1. After the second iteration, with better restored image and kernel, the algorithm can correctly finds the true class label. This illustrates that our approach can effectively regularize the ill-posed blind image restoration in pursuit of the sparsest representation for recognition. On one hand, a better recovered image will have a more meaningful sparse representation for recognition; on the other hand, the updated sparse representation, tightly connected with recognition, will provides a powerful regularization for the followed blind image restoration. In practice, we notice that the joint optimization proces converges very quickly, typically in no more than 4 iterations. Therefore, we fix the iteration number as 4 in all the following experiments.

4.2. Joint Blind Image Restoration and Recognition

In this subsection, we conduct experiments on joint image restoration and recognition for face images under various blind degradation settings. In our JRR algorithm, the tasks of image restoration and recognition are tightly coupled. However, to facilitate the comparisons with conventional restoration and recognition approaches respectively, we will present the results for restoration and recognition separately in the sequel.

4.2.1 Blind Image Restoration

We first quantitatively evaluate the kernel estimation and image restoration accuracy on Extended Yale B face dataset. To be consistent with the recognition evaluation, we randomly select half of the images for each subject as the training set. We then randomly choose 10 images from the rest as our testing examples for restoration. For each test image, we generate its blurry images using the 8 realistic non-parametric complex blur kernels proposed by Levin *et al.* in [11], shown in the first row of Table 2. Given a blurry input, our JRR algorithm estimates the unknown blur kernel without any prior knowledge and recovers the underlying sharp latent image, which are then evaluated in terms of RMSE with respect to the ground truth. We compare our JRR algorithm with the fast deblurring method in [2].

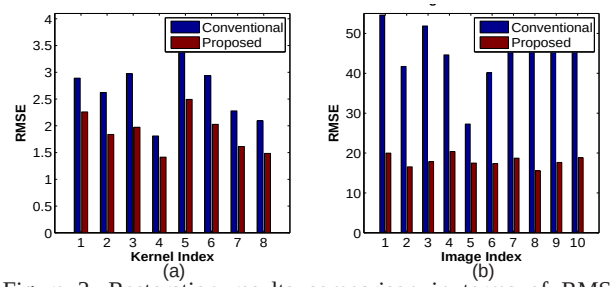

Figure 3. Restoration results comparison in terms of RMSE. (a) kernel estimation; (b) image estimation.

Figure 3 (a) shows the average RMSEs for each estimated kernels given the blurry inputs, where our JRR method improves the kernel estimation accuracy substantially compared with the fast deblurring algorithm. This can be explained by the fact that face images are lack of strong edge structures, especially in the case of blurry observation, which presents a great challenge to the existing blind deblurring methods. With the sparse representation prior, however, our method demonstrates much more robustness in estimating the complex blur kernels. Figure 3 (b) shows the comparisons of average restoration RMSEs for the 10 images under the 8 complex kernels. Due to the incorporation of the sparse representation prior, our algorithm improves the restoration accuracy significantly over the fast deblurring method for all the test images. By exploiting the sparse representation prior, the restored image has more details and less artifacts (Figure 4), implying a more accurate sparse representation, thus facilitating recognition, as shown in the following.

4.2.2 Blind Image Recognition

For recognition, we first evaluate the recognition performance of the proposed method on Extended Yale B dataset. We randomly select half of the images for each subject for training, and use the rest for testing. To generate the blurry inputs, we also add two more simple parametric blur kernels, *i.e.*, linear motion kernel and Gaussian blur kernel, in addition to the eight complex blur kernels [11]. For each blur kernel, we generate a set of blurred testing images, leading to in total 10 testing sets. Table 1 summarizes the recognition results for a simple motion blur (10 pixel-length with 45 degree) and a Gaussian kernel (with standard deviation 3), where the kernel size is 9×9. Our JRR algorithm outperforms SRC remarkably, while slightly better

Table 1. Recognition rate (%) on Extend Yale B under simple parametric blur kernels.

Kernel Type	SVM	SRC	SRC-B	JRR
Motion	40.0	68.7	85.3	**86.0**
Gaussian	29.9	57.7	**84.8**	**84.8**

Table 2. Recognition accuracy (%) on Extend Yale B set under complex non-parametric blur kernels.

Kernels								
Sizes	19	17	15	27	13	21	23	23
SVM	45.9	27.2	45.8	11.2	43.5	48.4	20.9	16.9
SRC	79.8	54.1	74.9	21.3	65.5	83.5	36.6	30.3
SRC-B	80.6	**79.3**	73.4	33.0	70.1	76.8	51.9	51.9
JRR	**86.2**	**79.3**	**85.7**	**43.1**	**81.9**	**86.4**	**64.7**	**54.8**

than SRC-B. This is because the conventional blind deblurring method can estimate the blur kernel reasonably well in simple blur model case. Table 2 presents the recognition results under the complex non-parametric blur kernels. In this case, conventional blur kernel estimation methods fail easily due to the complexity of the kernels and lack of strong structures in the face images, and as a result, the recognition results of our JRR algorithm outperform those of SRC-B and SRC by a large margin in most cases.

We then evaluate our algorithm on Multi-PIE [8] dataset, with 15 images from each subject of Session 1 for training and the rest of Session 1 for testing. Due to space limitation, we only report the results for the third complex kernel as shown in Table 3. Again, our algorithm performs much better than other methods. Note that as the conventional kernel estimation method is not robust enough in this case, SRC-B performs even worse than SRC. We further evaluate our algorithm in a more realistic scenario, where the blur kernel for generating a blurry image is not fixed but randomly chosen from $\{$*Linear Motion kernel, Gaussian kernel, Nonparametric Complex kernel, Delta (no blur)*$\}$. The recognition results for this case are shown in Table 4, and our proposed JRR method outperforms all the other methods with large margins on both datasets.

Finally, to visually demonstrate the effectiveness of our JRR algorithm, we compare the estimated kernels, deblurred images, and the top-10 selected atoms with the largest absolute coefficients from sparse representations under two different kernels, shown in Figure 4. Top row shows the results of SRC; middle row shows the results of conventional blind deblur followed by SRC; and bottom row shows our results. The blur kernels framed in red denote the ground truth kernels, and those framed in green are the estimated kernels. In both cases, our algorithm can accurately estimate the unknown blur kernels and can output sharp images close to the ground truth, while the fast deblurring method is not robust and fails drastically for the

Table 3. Recognition rate (%) on Multi-PIE with the third complex blur kernel.

Algorithm	SVM	SRC	SRC-B	JRR
Accuracy	84.8	85.2	79.1	**91.4**

Table 4. Recognition rate (%) with randomly blur kernels on both Extended Yale B and Multi-PIE.

Algorithm	SVM	SRC	SRC-B	JRR
Extended Yale B	57.0	68.8	66.3	**73.7**
Multi-PIE	49.4	53.6	54.9	**61.3**

complex kernel. To the right of each restored image, top-10 atoms from the sharp training set are selected by the largest absolute sparse representation coefficients, where red numbers denote atoms chosen from the same class (correct) and blue numbers denote otherwise (wrong). It is clear that our JRR algorithm can select more atoms from the same class with more concentrated large coefficients, indicating better recognition ability.

However, a challenging situation is when the blurry test image suffers from extreme illuminations, as in Figure 5, where little information about the facial structures is kept for deblurring. In this case, the deblurring task becomes extremely challenging and the blur kernel may not be correctly estimated even with our algorithm, which will lead to incorrect classification decisions. In both datasets we use, there are in fact a notable amount of such kind of images, which pose great challenges to the task of blind recognition on these datasets. Yet, with the sparse representation prior, the deblurring result of our algorithm looks much more reasonable than that of the fast deblurring method.

5. Conclusion and Future Work

We propose in this paper a joint restoration and recognition method with the sparse representation prior, and demonstrate its application on face recognition from a single blurry image. By combining these two interactive tasks, our algorithm demonstrates significant improvements over that of treating them separately. In the current model, mild translation misalignment between test and training images can be captured and compensated by the blur kernel. For future work, more complex alignment models, *e.g.*, affine transformation, can be incorporated into our framework to further handle more challenging misalignment between the blurry test image and sharp training images with techniques similar to [15] and [20]. Moreover, using learned dictionary rather than the training images directly in our model is also interesting and worthy of investigation in the future.

Acknowledgements This work is supported by NSF (60872145, 60903126), National High-Tech.(2009AA01Z315), Postdoctoral Science Foundation (20090451397, 201003685) and Cultivation Fund from Ministry of Education (708085) of China. This work is also supported by U.S. ARL and ARO under grant number W911NF-09-1-0383.

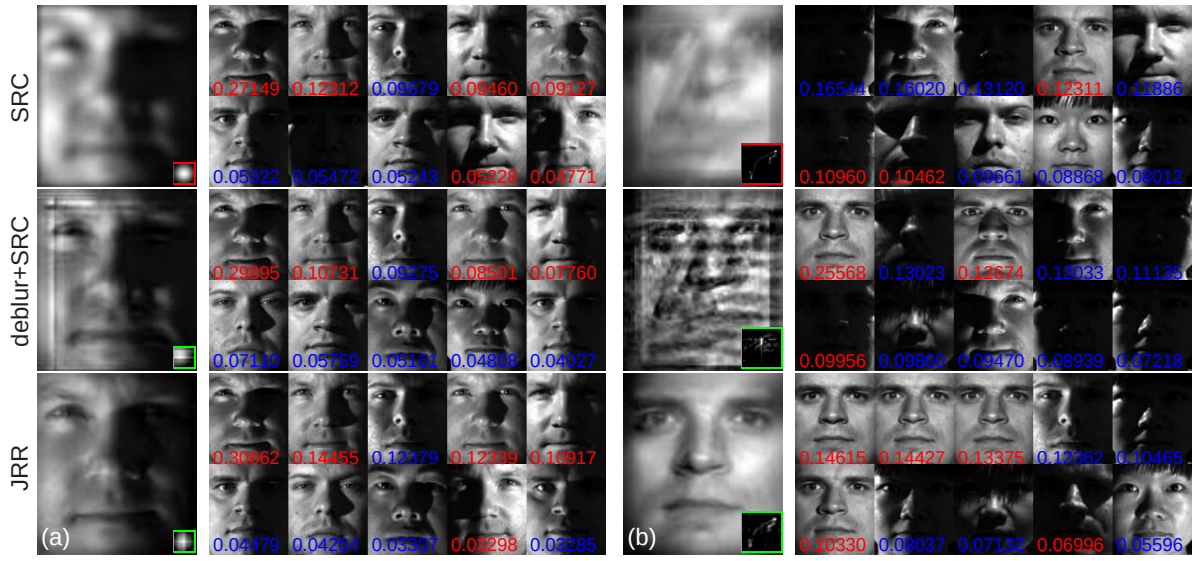

Figure 4. Image restoration results under (a) parametric PSF (Gaussian blur) and (b) realistic non-parametric PSF (27×27 non-parametric motion blur). Top: SRC; Middle: conventional deblur + SRC; Bottom: JRR. The PSF kernels framed in red denote the ground-truth kernels while those in green are estimated kernels. Atoms corresponding to the top-10 largest absolute coefficient values are shown together with the absolute values for each method, with red indicating atoms selected from the same class as the test image.

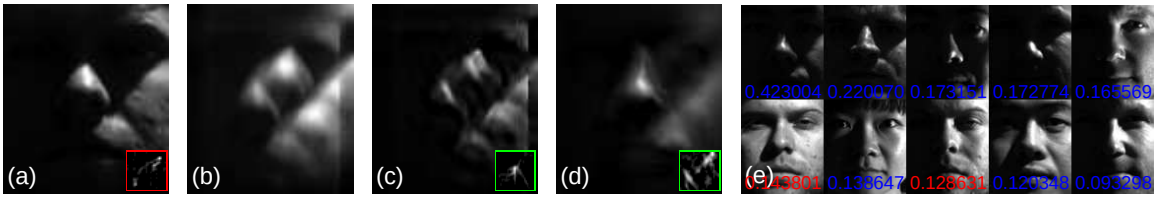

Figure 5. Failure case analysis. (a) ground truth image and kernel; (b) blurry input; estimated image and kernel using (c) conventional deblurring method [2] and (d) the proposed JRR method; (e) top-10 selected atoms with the JRR method. Kernel estimation is very challenging due to the extreme illumination.

References

[1] J.-F. Cai, H. Ji, C. Liu, and Z. Shen. Blind motion deblurring from a single image using sparse approximation. In *CVPR*, 2009. 2, 3

[2] S. Cho and S. Lee. Fast motion deblurring. In *SIGGRAPH ASIA*, 2009. 2, 3, 4, 5, 6, 8

[3] M. Das Gupta, S. Rajaram, N. Petrovic, and T. S. Huang. Restoration and recognition in a loop. In *CVPR*, 2005. 2

[4] D. L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math*, 59:797–829, 2004. 2

[5] M. Elad, M. A. T. Figueiredo, and Y. Ma. On the role of sparse and redundant representations in image processing. *Proc. of IEEE*, 98(6):972–982, 2010. 2, 3

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *SIGGRAPH*, 2006. 1, 2, 3, 5

[7] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE TPAMI*, 23(6):643–660, 2001. 5

[8] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *IEEE Intl. Conf. Automatic Face and Gesture Recog.*, 2008. 5, 7

[9] P. H. Hennings-Yeomans, S. Baker, and B. V. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *CVPR*, 2008. 2

[10] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *NIPS*, 2009. 3, 4, 5

[11] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009. 2, 6

[12] M. Nishiyama, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi. Facial deblur inference to improve recognition of blurred faces. In *CVPR*, 2009. 2

[13] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *CVPR*, 2005. 3

[14] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *SIGGRAPH*, 2008. 1, 2, 3, 4, 5

[15] 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 *CVPR*, 2009. 3, 7

[16] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Img. Sci.*, 1(3):248–272, 2008. 4

[17] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proc. of IEEE*, 98(6):1031–1044, 2010. 2

[18] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE TPAMI*, 2009. 2, 3, 5

[19] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super resolution as sparse representation of raw image patches. In *CVPR*, 2008. 1, 2, 3

[20] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Blurred/non-blurred image alignment using sparseness prior. In *ICCV*, 2007. 7