Blind Image Deblurring Using Spectral Properties of Convolution Operators

Guangcan Liu, *Member, IEEE*, Shiyu Chang, and Yi Ma, *Fellow, IEEE*

***Abstract*— Blind deconvolution is to recover a sharp version of a given blurry image or signal when the blur kernel is unknown. Because this problem is ill-conditioned in nature, effectual criteria pertaining to both the sharp image and blur kernel are required to constrain the space of candidate solutions. While the problem has been extensively studied for long, it is still unclear how to regularize the blur kernel in an elegant, effective fashion. In this paper, we show that the blurry image itself actually encodes rich information about the blur kernel, and such information can indeed be found by exploring and utilizing a well-known phenomenon, that is, sharp images are often high pass, whereas blurry images are usually low pass. More precisely, we shall show that the blur kernel can be retrieved through analyzing and comparing how the spectrum of an image as a convolution operator changes before and after blurring. Subsequently, we establish a convex kernel regularizer, which depends only on the given blurry image. Interestingly, the minimizer of this regularizer guarantees to give a good estimate to the desired blur kernel if the original image is sharp enough. By combining this powerful regularizer with the prevalent nonblind devonvolution techniques, we show how we could significantly improve the deblurring results through simulations on synthetic images and experiments on realistic images.**

***Index Terms*— Image deblurring, blind deconvolution, blur kernel estimation, point spread function, spectral methods.**

I. Introduction

FIGURE 1 shows a very common result of a very blurry image of a car taken by a moving camera — or similar blurring effect can be observed in many surveillance photos where the camera is static but the car is moving. In many situations, ones need to recover the sharp version of the blurry image so that the details (such as the numbers on the license plate) become recognizable to human eyes. It is in general difficult to correctly deblur the whole image if both the camera motion and the scene geometry are both entirely unknown.

Manuscript received April 2, 2014; revised July 13, 2014; accepted September 24, 2014. Date of publication October 8, 2014; date of current version October 21, 2014. This work was supported in part by the Start-Up Fund, Nanjing University of Information Science and Technology, Nanjing, China, in part by the National Science Foundation of Jiangsu Province under Grant 2012045, and in part by the National Natural Science Foundation of China under Grant 61272223.

G. Liu is with the Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: gcliu@nuist.edu.cn).

S. Chang is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA (e-mail: chang87@illinois.edu).

Y. Ma is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 200031, China (e-mail: mayi@shanghaitech.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2362055

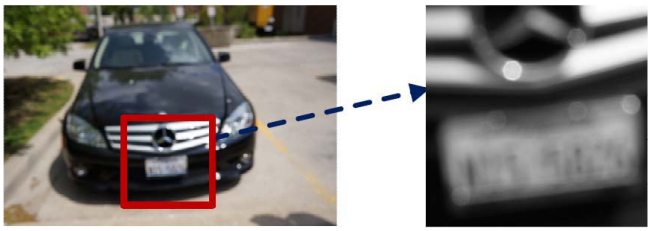


Fig. 1. A scenario of image deblurring. We take a picture of size 1624×2448 by using a NEX-5N camera. The picture is blurry due to an uncontrolled camera motion. We want to process the picture such that human eyes are able to recognize its details (e.g., the plate number).

Fortunately, while ones focus on a small image region (e.g., the license plate area shown in Figure 1), it is reasonable to believe that the blurry image, denoted as $B \in \mathbb{R}^{n_1 \times n_2}$ in this paper, is approximately generated by the convolution of a sharp image, denoted as I_0, and a generic blur kernel, denoted as $K_0 \in \mathbb{R}^{m_1 \times m_2}$:

$$B \approx I_0 \star K_0, \quad \text{s.t. } K_0 \in \mathcal{S},$$

where $\star$ denotes the discrete 2D convolution operator, S is the *simplex* (i.e., nonnegative and sums to one) of all possible blur kernels, $\{n_1, n_2\}$ are the image sizes, and $\{m_1, m_2\}$ are the sizes of the blur kernel. In this way, the blind image deblurring problem is mathematically formulated as the problem of *blind deconvolution* [1]–[3], which is to recover the latent sharp image I_0 when the blur kernel K_0 is unknown.[1]

The blind decovolution problem has been being widely investigated for several decades in optical society, image processing, computer graphics and computer vision [1], [5]–[10]. Yet, as will be elucidated later, this problem is still far from being well solved in more general cases even with the notable progresses made recently (see [11]–[14]). A straightforward approach for blind deconvolution is to jointly seek the latent sharp image I_0 and the blur kernel K_0 by

$$\min_{I,K} \|B - I \star K\|_F^2, \quad \text{s.t. } K \in \mathcal{S},$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. Unfortunately, the above optimization problem is severely ill-conditioned, as it can be perfectly minimized by infinite number of pairs (I, K). For example, the *no-blur explanation* $(I = B, K = \delta)$ is a perfect solution, because $B = B \star \delta$, where δ is a delta kernel. Indeed, even if the blur kernel K_0 is given, the non-blind deconvolution problem is still

[1] In contrast, *non-blind deconvolution* [4] is the restoration of I_0 when K_0 is given in an exact or approximate fashion.

ill-conditioned [4], [15]. So, in general, it is necessary to regularize the desired solution for the image variable I [16]:

$$\min_{I,K} \|B - I \star K\|_F^2 + \lambda f(I), \quad \text{s.t. } K \in \mathcal{S}, \tag{1}$$

where $\lambda \geq 0$ is a parameter. The regularizer $f(I)$ is usually chosen as the *total variation* [17] or its variations [4], [15], [18]. However, as pointed out in [8], such image gradient based regularizers generally favor a blurry solution over a sharp one. More precisely, whenever $f(I)$ is convex and shift-invariant,[2] it can actually be proven that (see Section V)

$$f(I \star K) \leq f(I), \quad \forall I \in \mathbb{R}^{n_1 \times n_2}, \ \forall K \in \mathcal{S}, \tag{2}$$

which gives $f(B) \leq f(I_0)$ (assume $B = I_0 \star K_0$). Since $B = B \star \delta$, the no-blur explanation (i.e., $I = B$ and $K = \delta$) is still a solution favored by (1). Hence, it is also critical and in fact indispensable to regularize the kernel variable K, i.e., it is necessary to consider an extended version of (1):

$$\min_{I,K} \|B - I \star K\|_F^2 + \lambda f(I) + \alpha h(K), \ \text{s.t. } K \in \mathcal{S}, \tag{3}$$

where $\lambda \geq 0$ and $\alpha \geq 0$ are two parameters.

While it is well-recognized that the regularizer $h(K)$ is important and indispensable for blind deconvolution, the existing proposals for $h(K)$, e.g., the Gaussian function $h(K) = \|K\|_F^2$ [7], [19], the sparse regularizer $h(K) = \|K\|_1$ [14], [20] and the Bayesian prior [8], in fact contain no real information about the desired blur kernel K_0. Therefore, in order to generate solutions of useful, existing deblurring approaches often require some heuristic regularizers such as salient structure selection [19], [20]. However, those heuristic regularizers can work well only on some simple cases where there is no serious blur, but could not handle more difficult deblurring tasks where the images are severely blurred (e.g., Figure 1). Furthermore, a specific kind of prior actually may not fit well a real world blur kernel, which is usually the combination of a sparse curve-like kernel (which corresponds to the camera motion) and a dense Gaussian-like kernel (which models effects such as out of focus).

In this work, we derive a much more effective regularizer for the blur kernel that can generally and significantly benefit the solution of the blind deconvolution problem. Our regularizer is based on a well-known observation; that is, *blurry images are usually low-pass and sharp images are often high-pass, or in other words, blurring can largely decrease the image frequencies in Fourier domain*. While well-recognized, to our knowledge, this classical observation has not been utilized to estimate the blur kernel. This paper will take a step in this direction. Our idea is to consider the spectral properties of an image as a convolution operator: For a given image (i.e., matrix), consider its convolution with any other matrix. The convolution defines a linear operator. *Then the classical observation suggests that the spectrum (the set of eigenvalues, i.e., the Fourier frequencies) of the linear operator for a blurry image should be significantly smaller than that for its sharp counter part*, as will be further confirmed by the investigations of this paper. Based on this classical observation, interestingly, we devise a convex regularizer which tends to be minimal at the blur kernel of desired, K_0. Namely, given an observed image B represented by a certain image feature $\mathcal{L}$, we deduce a convex function, denoted as $h^{\mathcal{L}(B)}(K)$:

$$h^{\mathcal{L}(B)}(K) : \mathbb{R}^{m_1 \times m_2} \to \mathbb{R}. \tag{4}$$

Unlike most previous work (see [7], [14], [19], [20]) where the regularizer $h(K)$ is independent of the observed image B, our regularizer $h^{\mathcal{L}(B)}(K)$ explicitly depends on the given blurry image and encodes rich information about how the blurry image B is related to the sharp image I_0. It is hence somewhat natural to anticipate that this regularizer $h^{\mathcal{L}(B)}(K)$ would depend on the unknown sharp image I_0 too. But rather surprisingly, as we will show, under fairly broad conditions, we can come up with a very effective regularizer $h^{\mathcal{L}(B)}(K)$ that does not depend on any information about the sharp image I_0 at all, and thus the desired kernel K_0 can be approximately retrieved by minimizing $h^{\mathcal{L}(B)}(K)$ directly.

Equipped with such a strong regularizer $h^{\mathcal{L}(B)}(K)$, we can jointly seek the sharp image I_0 and the blur kernel K_0 by solving the optimization problem (3). Experimental performance of our algorithm is a bit surprising: Even when the observed image is blurred to the extent where human eyes cannot recognize its details (e.g., Figure 1), it is still possible for our algorithm to restore a sharp version with recognizable details. In addition to empirical evaluations, theoretical results presented in this paper could also help to understand the blind deconvolution problem: It is known that there are infinite number of ways to decompose a given blurry image B into the convolution of I and K. Nevertheless, while I (or its edge map) is assumed to be very sharp, our theory implies that such decomposition is actually unique or approximately so. Moreover, our analysis provides an explanation for why sharp edges are good features for deblurring: The spectrum of the images in the edge domain is more sensitive to blurring than in the raw pixel domain, and thus the blur kernel K_0 is easier to restore by using edge features than using raw pixel values. In summary, the contributions of this paper include:

- ◇ While it has been known for long that sharp images often have high frequencies and blurry images are usually of low-pass, this observation has not been well utilized to solve the blind deconvolution problem. For the first time, we show that such a classical observation indeed gives an accurate estimate to the desired blur kernel, K_0.
- ◇ We establish a novel kernel regularizer that can be effectual for various blurs such as motion blur and de-focus. Unlike previous regularizers which in fact contains no real information about the blur kernel, our regularizer has a strong effect and can even directly retrieve the blur kernel K_0 without knowing the latent sharp image I_0. Using our regularizer, the generation of meaningful solutions for blind deconvolution does not require delicate balancing any more.
- ◇ Our studies are helpful for understanding the blind deconvolution problem and explaining the widely accepted

[2] $f(I)$ is shift-invariant means that $f(\mathcal{T}(I)) = f(I)$ holds for any shift operator $\mathcal{T}$. It is easy to see that image gradients based regularizers (e.g., total variation) are shift-invariant.

doctrine that the sharp edges of images are good features for deblurring [11], [14], [16], [19], [20].

II. Blind Deconvolution by Spectral Properties of Image Convolution

In this section, we present a rigorous derivation that elucidates how the proposed regularizer $h^{\mathcal{L}(B)}(K)$ is established. The algorithm for blind deconvolution will be given at the end of this section.

A. Spectrum of a Natural Image as a Convolution Operator

1) Preliminaries: The concept of *convolution* is well-known. We briefly introduce it for the completeness of presentation. Suppose X and Y are functions of two discrete variables, i.e., X and Y are matrices. Then the 2D convolution procedure of X and Y is given by

$$(X \star Y)(i, j) = \sum_{u,v} X(i-u, j-v)Y(u, v),$$

where $(\cdot)(i, j)$ denotes the (i, j)th entry of a matrix. The convolution operator is linear and can be converted into matrix multiplication. Let $v(\cdot)$ be the vectorization of a matrix. Then it can be manipulated that

$$v(X \star Y) = \mathcal{A}_{k_1,k_2}(X)v(Y), \quad \forall Y \in \mathbb{R}^{k_1 \times k_2}, \tag{5}$$

where $\mathcal{A}_{k_1,k_2}(\cdot)$ is called the *convolution matrix* of a matrix, and $\{k_1, k_2\}$ are token as parameters. For an ℓ_1-by-ℓ_2 matrix X, its convolution matrix, denoted as $\mathcal{A}_{k_1,k_2}(X)$ in this paper, is of size $(l_1 + k_1 - 1)(l_2 + k_2 - 1)$-by-$k_1k_2$.

2) Convolution Eigenvalues and Convolution Eigenvectors of Images: In this work, an image I is treated as a matrix associated with a certain feature filter $\mathcal{L}$:

$$\mathcal{L}(I) = \mathcal{L} \star I. \tag{6}$$

Typically choices for the feature filter include $\mathcal{L} = \delta$ (i.e., using original pixel values as features) or $\mathcal{L} =$ "Laplacian of Gaussian (LoG)" (i.e., using edge features).

To formally characterize the procedure of transforming sharp images into blurry images, we define the so-called *convolution eigenvalues* and *eigenvectors* as follows:

Definition 1: For an image $I \in \mathbb{R}^{l_1 \times l_2}$ represented by a feature filter $\mathcal{L}$, its first convolution eigenvalue, denoted as $\sigma_1^{\mathcal{L}}(I)$ or $\sigma_{max}^{\mathcal{L}}(I)$, is defined as

$$\sigma_1^{\mathcal{L}}(I) = \max_{X \in \mathbb{R}^{s_1 \times s_2}} \|\mathcal{L}(I) \star X\|_F, \text{ s.t. } \|X\|_F = 1,$$

where $\{s_1, s_2\}$ are called "sampling sizes" in this paper. The maximizer to above problem is called the first convolution eigenvector, denoted as $\kappa_1^{\mathcal{L}}(I)$.

Similarly, the ith $(i = 2, \ldots, s_1s_2)$ convolution eigenvalue, denoted as $\sigma_i^{\mathcal{L}}(I)$, is defined by

$$\sigma_i^{\mathcal{L}}(I) = \max_{X \in \mathbb{R}^{s_1 \times s_2}} \|\mathcal{L}(I) \star X\|_F,$$
$$\text{s.t.} \|X\|_F = 1, \langle X, \kappa_j^{\mathcal{L}}(I)\rangle = 0, \quad \forall j < i,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner production between two matrices. The maximizer to above problem is the ith convolution eigenvector, denoted as $\kappa_i^{\mathcal{L}}(I)$.

Classical Fourier theory suggests that the convolution eigenvalues (i.e., $\{\sigma_i^{\mathcal{L}}(I)\}$) of an image are indeed linked to the frequency components of the image. However, it is still unknown how to interpret in a Fourier sense the convolution eigenvectors (i.e., $\{\kappa_i^{\mathcal{L}}(I)\}$), which play a central role in our approach. In the following, we will summarize some properties of the convolution eigenvalues and eigenvectors:

- From (5), it can be seen that the convolution eigenvectors/eigenvalues are exactly the right singular vectors/values of the convolution matrix. So, for an image I with the associated convolution matrix $\mathcal{A}_{s_1,s_2}(\mathcal{L}(I))$, its all s_1s_2 convolution eigenvalues and eigenvectors can be found by computing the Singular Value Decomposition (SVD) of the matrix, $(\mathcal{A}_{s_1,s_2}(\mathcal{L}(I)))^T\mathcal{A}_{s_1,s_2}(\mathcal{L}(I))$, which is of size s_1s_2-by-s_1s_2.
- Let $\sigma_{min}^{\mathcal{L}}(\cdot)$ denote the smallest convolution eigenvalue of an image. Then

$$\|\mathcal{L}(I) \star X\|_F \geq \sigma_{min}^{\mathcal{L}}(I) > 0, \quad \forall \|X\|_F = 1, \tag{7}$$

  where the inequality $\sigma_{min}^{\mathcal{L}}(I) > 0$ follows from the assertion that the convolution matrix is of full column rank except the extreme case $I = 0$ [21].
- It has been known that the blurry images are usually low-pass and the sharp images are often high-pass, i.e., blurring could decrease dramatically the Fourier frequencies of an image. In other words, blurring may significantly reduce the convolution eigenvalues, which essentially correspond to the image frequencies in Fourier domain. In fact, this classical observation can be proven to be true to some extent. More precisely, if $B = I \star K$ and $K \in \mathcal{S}$, then we have (detailed proofs are in Section V)

$$\sigma_i^{\mathcal{L}}(B) \leq \sigma_i^{\mathcal{L}}(I), \quad \forall i = 1, \ldots, s_1s_2. \tag{8}$$

  When the original image is sharp, the reduction amount can be very significant, as exemplified in Figure 2. In particular, when the edge features are used, Figure 2(f) shows that it is even possible to reach $\sigma_{min}^{\mathcal{L}}(I) \gg \sigma_{max}^{\mathcal{L}}(B)$.

The properties in inequality (8) suggest that a suitable way to formally define the concept of *sharp image*, which appears frequently in the articles related to deconvolution, would be the following: An image I is called τ*-sharp* if and only if $\sigma_{min}^{\mathcal{L}}(I) \geq \tau$, where $\tau > 0$ is a parameter. That is, in general, the convolution eigenvalues of sharp images are large, while those of blurry images are relatively smaller. This definition is consistent with the classical observation which states that sharp images are often of high-pass.

B. A Convex Blur Kernel Regularizer

By making use of the well-recognized observation that blurring could significantly reduce the image frequencies or equivalently the convolution eigenvalues (as shown in Figure 2), we can derive a novel convex regularizer, denoted as $h^{\mathcal{L}(B)}(K)$. Interestingly, this regularizer tends to be minimal at the desired blur kernel, K_0.

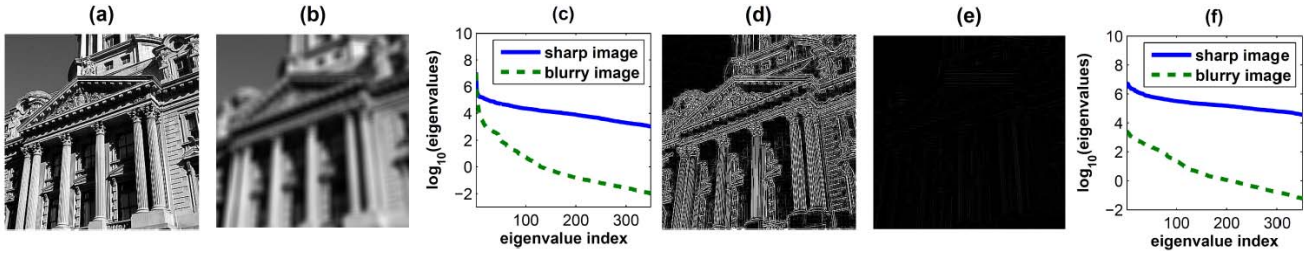


Fig. 2. Exemplifying that blurring can significantly decrease the convolution eigenvalues of a sharp image. (a) A sharp image. (b) The blurry image created by convoluting the sharp image with a Gaussian kernel. (c) Plots of all convolution eigenvalues ($s_1 = s_2 = 18$) of the sharp and blurry images, using $\mathcal{L} = \delta$. (d) The edges of the sharp image, using $\mathcal{L} =$ LoG. (d) The edges of the blurry image (note that the black area contains negative values). (e) Plots of convolution eigenvalues of the sharp and blurry images, choosing $\mathcal{L} =$ LoG.

1) Derivation: For ease of exploration, we begin with the simple case where the blurry image B is exactly generated by the convolution of I_0 and K_0, i.e., $B = I_0 \star K_0$. As convolution operators are associative and commutative, the effect of feature extraction is

$$\mathcal{L}(B) = \mathcal{L}(I_0) \star K_0, \tag{9}$$

where $\mathcal{L}$ is a predefined feature filter. For the rest of this paper, we consistently choose $\mathcal{L} =$ LoG, i.e., we use edge features by default (but our analysis applies to any feature filters).

By Definition 1,

$$\|\mathcal{L}(B) \star \kappa_i^{\mathcal{L}}(B)\|_F = \sigma_i^{\mathcal{L}}(B), \quad \forall 1 \le i \le s_1 s_2,$$

where $\kappa_i^{\mathcal{L}}(B) \in \mathbb{R}^{s_1 \times s_2}$ is computed from the given blurry image B by SVD, and the sampling sizes $\{s_1, s_2\}$ are taken as parameters. By (9), we have

$$\|(\mathcal{L}(I_0) \star K_0) \star \kappa_i^{\mathcal{L}}(B)\|_F = \sigma_i^{\mathcal{L}}(B), \quad \forall 1 \le i \le s_1 s_2.$$

Since the convolution operator is linear, we further have

$$\left\| \mathcal{L}(I_0) \star \frac{K_0 \star \kappa_i^{\mathcal{L}}(B)}{\|K_0 \star \kappa_i^{\mathcal{L}}(B)\|_F} \right\|_F = \frac{\sigma_i^{\mathcal{L}}(B)}{\|K_0 \star \kappa_i^{\mathcal{L}}(B)\|_F}.$$

Notice that $\left\| \frac{K_0 \star \kappa_i^{\mathcal{L}}(B)}{\|K_0 \star \kappa_i^{\mathcal{L}}(B)\|_F} \right\|_F = 1$. By (7), $\|\mathcal{L}(I_0) \star X\|_F \ge \sigma_{min}^{\mathcal{L}}(I_0), \forall \|X\|_F = 1$. Thus we have the following necessary conditions for constraining possible K_0:

$$\|K_0 \star \kappa_i^{\mathcal{L}}(B)\|_F \le \frac{\sigma_i^{\mathcal{L}}(B)}{\sigma_{min}^{\mathcal{L}}(I_0)}, \quad \forall 1 \le i \le s_1 s_2. \tag{10}$$

Now consider the bottom convolution eigenvectors of the blurry image B, i.e., the convolution eigenvectors that correspond to the low frequency components of B. Due to the fact that the sharp image I_0 is high-pass and has a (relatively) large $\sigma_{min}^{\mathcal{L}}(I_0)$ (see Figure 2), the ratio $\sigma_i^{\mathcal{L}}(B)/\sigma_{min}^{\mathcal{L}}(I_0)$ should be very small while $\sigma_i^{\mathcal{L}}(B) < \eta$ for some threshold $\eta > 0$. That is to say, $\|K_0 \star \kappa_i^{\mathcal{L}}(B)\|_F \approx 0$ for $\sigma_i^{\mathcal{L}}(B) < \eta$. Hence, it is natural to try recovering the desired blur kernel K_0 by finding a solution which is approximately orthogonal (in a convolution sense) to some bottom convolution eigenvectors of B:

$$\min_K \sum_{\sigma_i^{\mathcal{L}}(B) < \eta} \|K \star \kappa_i^{\mathcal{L}}(B)\|_F^2, \quad \text{s.t. } K \in \mathcal{S}, \tag{11}$$

where $\eta > 0$ is a parameter. For the reference of readers, we present in Figure 3 some examples of the bottom convolution eigenvectors. Roughly, the bottom eigenvectors could be regarded as some kind of "phases" corresponding to the low frequency components of a burry image.

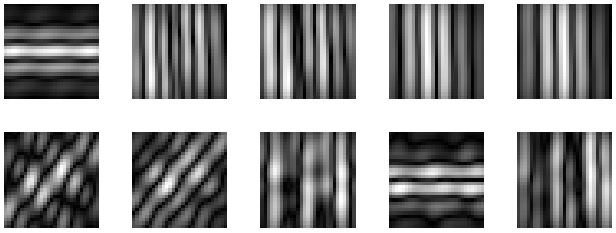


Fig. 3. Visualization of ten bottom convolution eigenvectors of a blurry image. In this experiment, I_0 is the Lena image, K_0 is a 9×9 Gaussian kernel, and the sampling sizes are chose as $s_1 = s_2 = 29$.

We have experimentally confirmed that the criterion (11) indeed works well in estimating the blur kernel. Nevertheless, there is an unknown parameter η to select. To mitigate this issue, we introduce below an η-free criterion that is approximately equivalent to (11). Define a function of K as

$$h^{\mathcal{L}(B)}(K) \doteq \sum_{i=1}^{s_1 s_2} \frac{\|K \star \kappa_i^{\mathcal{L}}(B)\|_F^2}{(\sigma_i^{\mathcal{L}}(B))^2}, \tag{12}$$

which is indeed a soft-threshold version of the objective defined in (11): The weight $1/(\sigma_i^{\mathcal{L}}(B))^2$ drops fast while $\sigma_i^{\mathcal{L}}(B)$ goes large, and thus the terms $\{\|K \star \kappa_i^{\mathcal{L}}(B)\|_F^2\}$ corresponding to small convolution eigenvalues will dominate the whole function. As discussed above (11), the desired blur kernel K_0 should have a small value for $h^{\mathcal{L}(B)}(K)$, and, in contrast, an incorrect blur kernel may produce a relatively larger objective value. In fact, we have experimentally found that $h^{\mathcal{L}(B)}(K_0) \approx 0$, which is consistent with the discussions above (11). For the example used in Figure 3, it is computed that $h^{\mathcal{L}(B)}(K_0) = 0.12$, $\min_K h^{\mathcal{L}(B)}(K) = 0.09$, and, in sharp contrast, $h^{\mathcal{L}(B)}(\delta) = 8.5 \times 10^6$. Actually, the assertion that $h^{\mathcal{L}(B)}(K_0)$ is small can be proven to be true to some extent: By (10), we have that

$$h^{\mathcal{L}(B)}(K_0) \le \frac{s_1 s_2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2}. \tag{13}$$

Hence, we could try to obtain an approximate estimate of the desired blur kernel K_0 by

$$\hat{K}_0 = \arg\min_K h^{\mathcal{L}(B)}(K), \quad \text{s.t. } K \in \mathcal{S}, \tag{14}$$

where S denotes the $(m_1 m_2 - 1)$-dimensional simplex, and $\{m_1, m_2\}$ are the sizes of the blur kernel. Comparing to

Algorithm 1 Computing the Hessian Matrix H Defined in (15)

Input: blurry image B.
Parameters: kernel sizes $\{k_1, k_2\}$; sampling sizes $\{s_1, s_2\}$.
1. Compute the edge map of B by $\mathcal{L}(B) = B \star \text{LoG}$.
2. Compute the convolution matrix $\mathcal{A}_{s_1,s_2}(\mathcal{L}(B))$ according to the definition in (5).
3. Let $M = (\mathcal{A}_{s_1,s_2}(\mathcal{L}(B)))^T \mathcal{A}_{s_1,s_2}(\mathcal{L}(B))$. Then obtain the convolution eigenvalues $\{\sigma_i^{\mathcal{L}}(B)\}$ and eigenvectors $\{\kappa_i^{\mathcal{L}}(B)\}$ by performing SVD on M.
4. For each $\kappa_i^{\mathcal{L}}(B)$, compute its convolution matrix $\mathcal{A}_{k_1,k_2}(\kappa_i^{\mathcal{L}}(B))$ according to (5).
5. Compute the Hessian matrix H by (15).
Output: H.

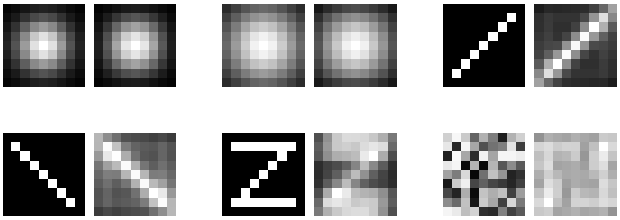


Fig. 4. Demonstrating the effectiveness of the regularizer $h^{\mathcal{L}(B)}(K)$, using six synthetical examples. Left: The blur kernel with size 9×9. Right: The kernel estimated by solving (14).

the criterion (11), this new criterion avoids choosing the threshold η and is therefore more brief.

It is easy to see that $h^{\mathcal{L}(B)}(K)$ is a quadratical (hence convex) function. Namely, $h^{\mathcal{L}(B)}(K) = (\nu(K))^T H \nu(K)$ with the Hessian matrix H given by

$$H = \sum_{i=1}^{s_1 s_2} \frac{(\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B)))^T \mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B))}{(\sigma_i^{\mathcal{L}}(B))^2}, \qquad (15)$$

where $\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B))$ is the convolution matrix of the ith convolution eigenvector of B. Algorithm 1 summarizes the whole procedure of computing the above Hessian matrix H.

2) Analysis: We now show how effective the regularizer $h^{\mathcal{L}(B)}(K)$ is by evaluating how close its minimizer $\hat{K}_0$ is to the true kernel K_0 (proofs to all the theories below can be found in Section V).

Theorem 1 (Noiseless): Suppose $B = I_0 \star K_0$, $K_0 \in \mathcal{S}$, and $I_0 \neq 0$. For the kernel $\hat{K}_0$ estimated by (14), we have that

$$\|K_0 - \hat{K}_0\|_F \leq \sqrt{2} \frac{\sigma_{max}^{\mathcal{L}}(B)}{\sigma_{min}^{\mathcal{L}}(I_0)}.$$

The above theorem argues that the edges are better than the raw pixels as a feature for the recovery of blur kernel, because the edge features can achieve a smaller bound for the estimate error. This to some extent corroborates the previous studies (see [11], [14], [16], [19], [20]) which observe that the edge feature is a good choice for image deblurring. Note that in general $\sigma_{min}^{\mathcal{L}}(I_0)$ can be significantly larger than $\sigma_{max}^{\mathcal{L}}(B)$. For the example shown in Figure 2(f), we have computed that the ratio $\sigma_{max}^{\mathcal{L}}(B)/\sigma_{min}^{\mathcal{L}}(I_0) = 0.15$, and thus the above theorem gives that the estimation error is upper bounded by 0.21. Figure 4 shows some more simulated examples of blur kernels directly estimated from various blurred Lena images, compared with the true kernels. Notice, that these kernels are obtained without any knowledge about the original image I_0 at all! These simulated results clearly demonstrate the effectiveness of the proposed regularizer $h^{\mathcal{L}(B)}(K)$: Our regularizer puts a strongly correct prior on the desired kernel entirely based on the given blurry image.

Without any restrictions, it is known that there are infinite number of ways to decompose a given blurry image B into the convolution of an image I and a blur kernel K. Interestingly, when the image I (or its edge image) is assumed to be very sharp (i.e., its smallest convolution eigenvalue is large), Theorem 1 indeed implies that the decomposition for given image B tends to be unique in a sense that the allowable kernels should be very close to each other. The following corollary makes this statement more precise: We call an image I is τ-sharp if $\sigma_{min}^{\mathcal{L}}(I) \geq \tau$ with $\tau = 2\sqrt{2}\sigma_{max}^{\mathcal{L}}(B)/\varepsilon$ for some small number ε. Then we have the following:

Corollary 1: Let $\tau = 2\sqrt{2}\sigma_{max}^{\mathcal{L}}(B)/\varepsilon$. Denote all possible τ-sharp decompositions of a blurry image B as $\Omega_B^{\varepsilon} = \{(K, I) | I \star K = B, K \in \mathcal{S}, \sigma_{min}^{\mathcal{L}}(I) \geq \tau\}$. For any two pairs $(K_0', I_0'), (K_0'', I_0'') \in \Omega_B^{\varepsilon}$, we have

$$\|K_0' - K_0''\|_F \leq \varepsilon,$$

where $\varepsilon > 0$ is any small parameter.

Theorem 1 is based on the assumption that there is no noise, i.e., $B = I_0 \star K_0$. Due to the fact the blur in reality may not exactly follow the convolution procedure, a more appropriate model is that $B = I_0 \star K_0 + N$, where N denotes some unknown errors or noises. In this case, we have the following theorem to bound the estimate error:

Theorem 2 (Noisy): Suppose $B = I_0 \star K_0 + N$, $K_0 \in \mathcal{S}$, $\|\mathcal{L}(N)\|_F \leq \epsilon$, and $I_0 \neq 0$. For the kernel $\hat{K}_0$ estimated by (14), we have that

$$\|K_0 - \hat{K}_0\|_F \leq \sqrt{2} \frac{\sigma_{max}^{\mathcal{L}}(B) + ccond(B)\sqrt{s_1 s_2}\epsilon}{\sigma_{min}^{\mathcal{L}}(I_0)},$$

where $ccond(B) = \sigma_{max}^{\mathcal{L}}(B)/\sigma_{min}^{\mathcal{L}}(B)$ is the "convolution condition number" of B.

C. Reliable Deconvolution via Alternating Minimization

While brief and effectual, as can be seen from Figure 4, the convex program (14) alone may not produce a satisfactory kernel which is good enough for restoring a sharp, natural image. We have indeed experimentally found that the blur kernel from (14) often leads to very sharp, but unnatural images full of artifacts. For reliable deconvolution, instead of directly using (14), we utilize the proposed regularizer $h^{\mathcal{L}(B)}(K)$ as an additional term to constrain the deconvolution problem. Namely, we jointly seek the sharp image I_0 and the blur kernel K_0 by

$$\min_{I,K} \|B - I \star K\|_F^2 + \lambda \|\nabla I\|_1 + \alpha h^{\mathcal{L}(B)}(K), \quad \text{s.t. } K \in \mathcal{S},$$

which follows from (3) by choosing $f(I) = \|\nabla I\|_1$ (i.e., total variation) and $h(K) = h^{\mathcal{L}(B)}(K)$.

Although the above problem is nonconvex in nature and might heavily depend on the initial solution chosen in advance, it is now equipped with our strong kernel regularizer, $h^{\mathcal{L}(B)}(K)$, and thus the initialization step is no long crucial. As a consequence, unlike the previous algorithms (see [7], [11], [14], [19], [20]) which need to carefully chose the initial solution, we simply choose the observed blurry image B as the initial condition for I. Then blind deconvolution is carried out by iterating the following two procedures until convergence:

- While fixing the variable I, the blur kernel K is updated by solving $\min_K \|B - I \star K\|_F^2 + \alpha h^{\mathcal{L}(B)}(K)$, s.t. $K \in \mathcal{S}$, which is equal to the following quadratical program:

$$\min_K \|v(B) - \mathcal{A}_{m_1,m_2}(I)v(K)\|_2^2 + \alpha(v(K))^T H v(K), \quad \text{s.t. } K \in \mathcal{S},$$

  where the Hessian matrix H is computed by Algorithm 1, $\|\cdot\|_2$ is the ℓ_2-norm of a vector, and $v(\cdot)$ denotes the vectorization of a matrix.
- While fixing the blur kernel K, the estimate of the image I is updated by

$$\min_I \|B - I \star K\|_F^2 + \lambda\|\nabla I\|_1,$$

  which can be solved by any of the many non-blind deconvolution algorithms developed in the literature (see [4], [22], [23]). In this work, we simply use the fast method introduced by Krishnan and Fergus [4].

In addition, unlike previous blind deblurring methods that need to carefully control the number of iterations, we run the iterations until convergence. Usually, our algorithm converges within 200 iterations.

1) On Choosing the Parameters: There are six parameters in our blind deblurring algorithm: The kernel sizes $\{m_1, m_2\}$ (usually $m_1 = m_2$), the sampling sizes $\{s_1, s_2\}$, and the trade-off parameters $\{\alpha, \lambda\}$. The kernel sizes $\{m_1, m_2\}$ are better to be a bit larger than the true blur size, which requires trying the algorithm several times to determine. When the kernel sizes have been determined, the sampling sizes could be simply set as $s_1 = 1.5m_1$ and $s_2 = 1.5m_2$. The parameter λ plays the role of suppressing possible artifacts arising from the non-blind deconvolution procedure. Usually, it is moderately good to choose this parameter from the range of 0.001 to 0.002. The parameter α also needs to be set properly. When α is too small, our algorithm generally converges to the no-blur explanation (i.e., $I = B, K = \delta$). If α is set to be too large, the recovered image could be very sharp, but contains artifacts (see Figure 5). In the environments where it is infeasible to try the algorithm multiple times for determining the parameters, we recommend $\alpha = 1000$.

2) Computational Complexity: Thanks to the fast non-blind deconvolution algorithms recently established in the literature (see [4]), the procedure for updating the image variable I is already very fast. So we only analyze the cost for computing the Hessian matrix H and updating the kernel variable K. For simplicity, we assume that $s_1 = m_1$ and $s_2 = m_2$. Then the complexity of computing the Hessian matrix H is $O(m_1^2 m_2^2 n_1 n_2 + m_1^3 m_2^3)$, where $\{n_1, n_2\}$ are image sizes. In each iteration, the convolution matrix $\mathcal{A}_{m_1,m_2}(I)$ need be updated. So, in summary, the overall complexity of our blind deblurring algorithm is $O(n_s(m_1 m_2 n_1 n_2) + m_1^2 m_2^2 n_1 n_2 + m_1^3 m_2^3)$, where n_s is the number of iterations. Generally, our current algorithm is not fast while dealing with large images and/or large kernels. Nevertheless, it is possible to speed up the algorithm by fast Fourier transforms. We leave this as future work.

Fig. 5. Exemplifying the influences of the parameter α. The blurry image is generated by convoluting the Lena image (resized to 300×300) with a 9×9 Gaussian kernel. The parameter λ is set as 0.0015.

3) Implementation Details: The Matlab codes of this work have been available at https://sites.google.com/site/guangcanliu/. The employed non-blind deconvolution tool [4] is downloaded from http://cs.nyu.edu/~dilip/research/fast-deconvolution.

III. Experiments

A. Results

We test with five examples (two synthetical, three real): The synthetical images are the convolution of 300×300 natural images and 13×13 synthetical blur kernels. The real images are subregions (about 350×350) selected from large pictures captured by a NEX-5N camera (one may also refer to Figure 1). To show the advantages of the proposed approach, we also test five state-of-the-art blind deconvolution algorithms, including [11], [14], [16], [19], and [20].

Figure 6 shows the comparison results. For the clarity of viewing, we only show the best result of the previous algorithms for comparison. On the simple examples with easy blur kernels (the first three examples in Figure 6), our algorithm performs as well as the most effective baseline. While dealing with the challenging examples (the last two examples in Figure 6), where the blur kernels are complicated,

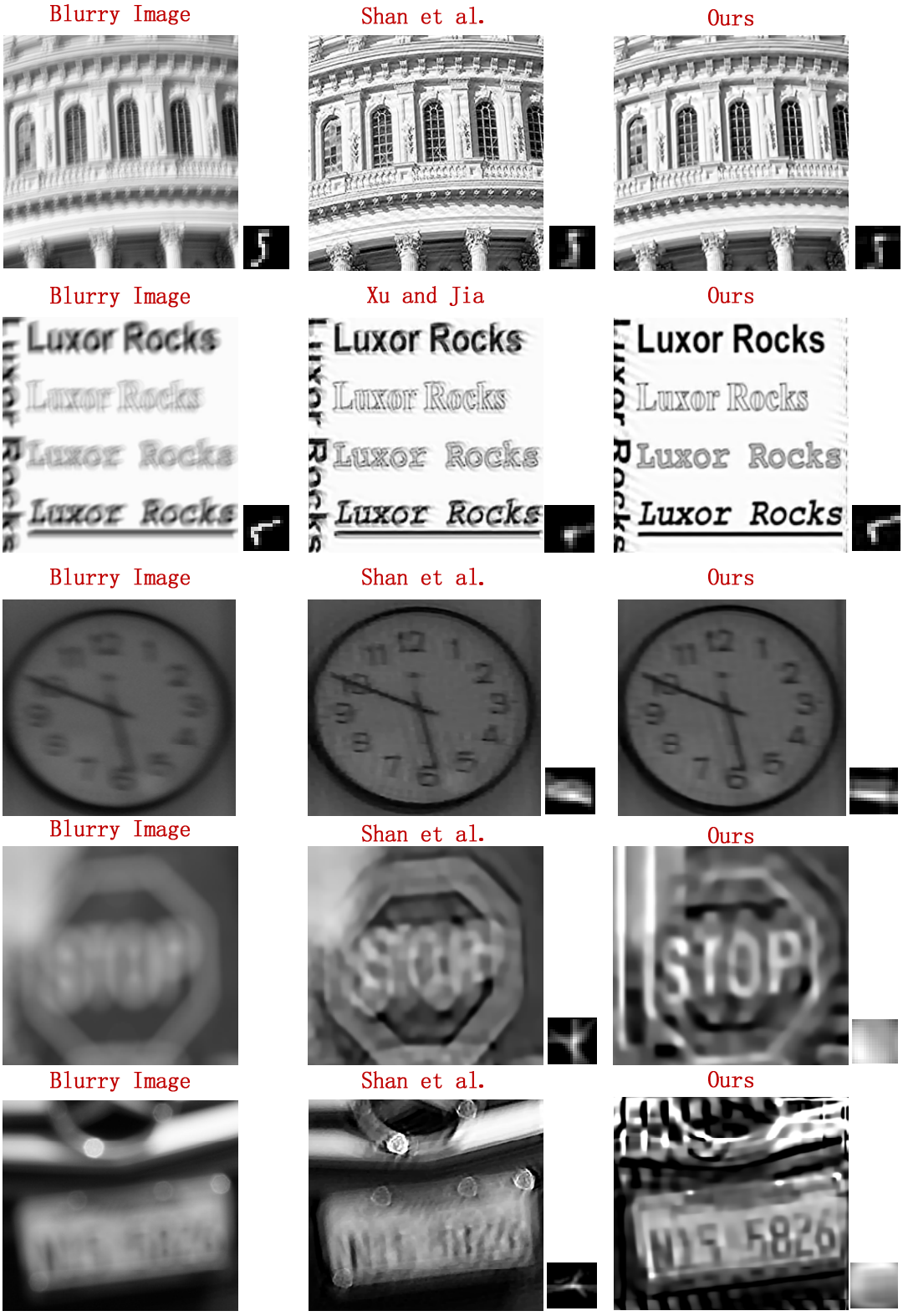


Fig. 6. Comparison results, using two synthetical and three real examples. Left: The observed blurry image. Middle: The best result among Fergus et al. [16], Shan et al. [20], Cho and Lee [19], Xu and Jia [14], and Cho et al. [11]. Right: Our result. For these five examples, the parameter α is set as 5050, 1050, 70, 4000, and 27000, respectively. The kernel sizes are, 13×13, 13×13, 17×17, 17×17, and 31×31, respectively. The parameters of the baselines are also manually tuned to best.

it can be seen that our algorithm works distinctly better than the most competitive baseline. This is because that our regularizer $h^{\mathcal{L}(B)}(K)$ contains strong and accurate priors about the blur kernels, whereas the kernel priors adopted by previous algorithms are too weak to handle such difficult deblurring tasks. In particular, the forth and fifth examples illustrate that it is even possible for our algorithm to successfully handle some extremely difficult cases, where the blurry images are

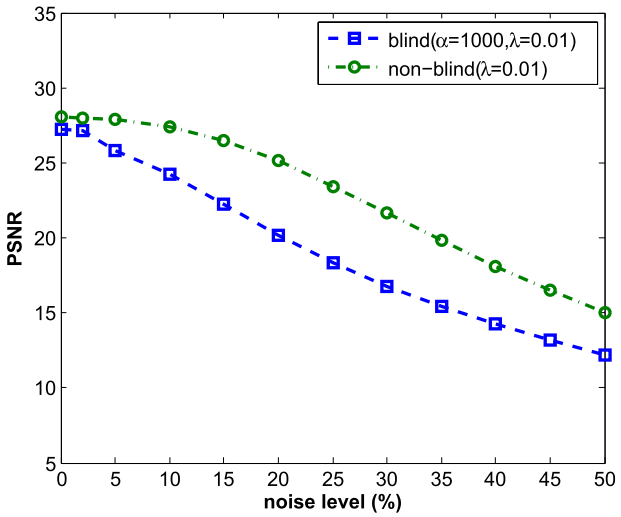


Fig. 7. Investigating the influences of white noise. The noised blurry image is generated according to $B = I_0 \star K_0 + N$, where N is of i.i.d. Gaussian. The used I_0 and K_0 are the same as in Figure 5. The label "blind" refers to our blind deblurring algorithm, and "non-blind" refers to the non-blind deconvolution algorithm [4], which assumes the true blur kernel. The "noise level" is computed by $(\|N\|_F/\|B - N\|_F)\times 100\%$.

very unclear such that human eyes are unable to recognize their contents.

B. Diagnosis

1) On Handling White Noise: Figure 7 shows that the our blind deblurring algorithm can only handle a small amount (about 2%) of white noise, which is dense. The reasons are two-fold: 1) The kernel regularizer $h^{\mathcal{L}(B)}(K)$ is less effective in the presence of noises. This can be seen from Theorem 2, which states that the effect of noises might be amplified by a factor of $ccond(B)\sqrt{s_1 s_2}/\sigma_{min}(I_0)$ (which could be fairly large). 2) The employed total variation prior is not very good at denoising. This is revealed by the phenomenon that even the non-blind deconvolution procedure (which uses K_0) drops notably while the noise level is higher than 10%. Hence, whenever the image is known to be contaminated by a considerable amount of white noise, we suggest to perform denoising before deblurring.

2) Results on the Images With Few Edges: Our approach prefers the cases where the edges of a sharp image are significantly different before and after blurring. So the approach would be less effective while the images contain fewer edges. Figure 8 shows three examples of the results obtained from the images lacking sharp edges. Particularly, the third example of Figure 8 shows that the true blur kernel cannot be accurately recovered if the original image has very few edges. Yet, it is indeed not very critical to consider deblurring such images, as there is almost no difference between their sharp and blurred versions.

3) Results on the Images With Non-Uniform Blur Kernels: Note that the forth example of Figure 6 may have non-uniform blur kernels, since the deblurred result favors the characters "O,P" over "S,T". This is true, as verified in Figure 9: While the parameter α is varying, the shape of the estimated blur kernel changes as well. These results also reflex that our proposed regularizer (which is derived with a uniform setting) could be stable with respect to some modeling errors, which are often sparse (this is different from the white noise).

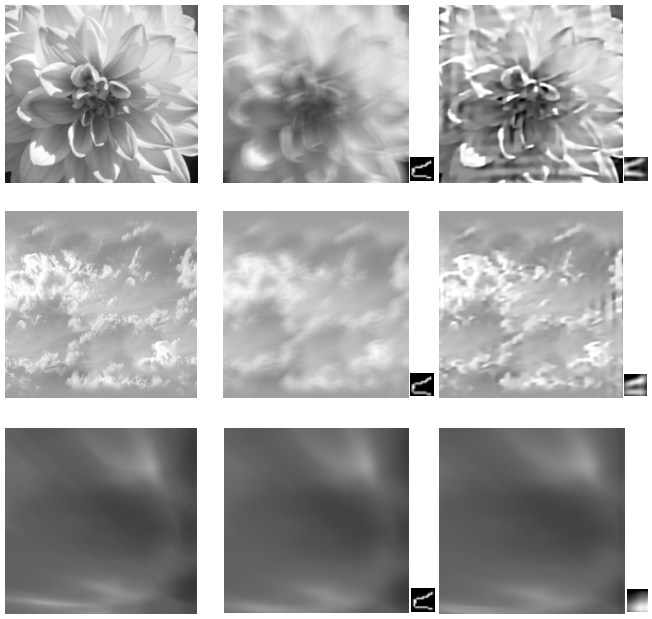


Fig. 8. Some results obtained from the images with few edges. Left: The original sharp image. Middle: The blurred version of the sharp image. Right: The deblurred version produced by our algorithm.

Fig. 9. Non-uniform blur kernels. Left: The observed blurry image. Middle: The deblurred version produced by our algorithm with $\alpha = 1000$. Right: The deblurred version produced by our algorithm with $\alpha = 4000$.

IV. Conclusions and Discussions

By studying how the spectrum of an image as a convolution operator changes before and after blurring, we have derived a convex regularizer on the blur kernel that depends on the given blurry image. For the blind deconvolution problem, we showed that this convex regularizer can deal with various types of realistic and challenging blurs (e.g., a combination of Gaussian and motion blurs). Both theoretical and experimental results have verified the validity and effectiveness of the proposed regularizer. Notice that our regularizer harnesses certain necessary conditions on the blur kernel, but not sufficient. So it is entirely possible there might exist other, potentially more effective image-dependent regularizers for the blur kernel.

Theorem 1 and the empirical results in Figure 4 even suggest that in theory, for certain class of sharp images, it might be possible to solve the blind deconvolution problem by avoiding the dependence between I_0 and K_0. Namely, the problem might be well solved by using only two computationally stable procedures: 1) Firstly estimate the blur kernel without estimating the sharp image at all, 2) and then recover the sharp

image by performing non-blind deconvolution. However, we leave such investigations for future.

V. Proofs

A. Proof of the Inequality (2)

Proof: Let $\mathcal{T}_{u,v}(\cdot)$ denote the transformation operator that shifts a 2D function $I(\cdot,\cdot)$ from $I(x,y)$ to $I(x-u,y-v)$. By the convexity of $f(\cdot)$,

$$\begin{aligned} f(I\star K) &= f(\sum_{u,v} K(u,v)\mathcal{T}_{u,v}(I)) \\ &\le \sum_{u,v} K(u,v) f(\mathcal{T}_{u,v}(I)) = f(I), \end{aligned}$$

where the last equality is due to the assumption that f is shift-invariant, i.e., $f(\mathcal{T}_{u,v}(I)) = f(I)$. Note that the shift-invariant property can be naturally satisfied, because $\mathcal{T}_{u,v}(I)$ and I are essentially the same image. ∎

B. Proof of the Inequality (8)

Proof: Using the "min-max" half of the Courant-Fisher theorem, we have that

$$\begin{aligned} \sigma_i^{\mathcal{L}}(B) &= \min_{\substack{Y_j,\\ j=1,\cdots,i-1.}} \max_{\substack{\|X\|_F=1,\\ \langle X,Y_j\rangle=0,\\ j=1,\cdots,i-1.}} \|\mathcal{L}(B)\star X\|_F \\ &\le \min_{\substack{Y_j=\kappa_j^{\mathcal{L}}(I),\\ j=1,\cdots,i-1.}} \max_{\substack{\|X\|_F=1,\\ \langle X,Y_j\rangle=0,\\ j=1,\cdots,i-1.}} \|\mathcal{L}(B)\star X\|_F \\ &= \max_{\substack{\|X\|_F=1,\\ \langle X,\kappa_j^{\mathcal{L}}(I)\rangle=0,\\ j=1,\cdots,i-1.}} \|\mathcal{L}(B)\star X\|_F. \end{aligned}$$

By (2), we have $\|\mathcal{L}(B)\star X\|_F \le \|\mathcal{L}(I)\star X\|_F$ and thus

$$\begin{aligned} \sigma_i^{\mathcal{L}}(B) &\le \max_{\substack{\|X\|_F=1,\\ \langle X,\kappa_j^{\mathcal{L}}(I)\rangle=0,\\ j=1,\cdots,i-1.}} \|\mathcal{L}(B)\star X\|_F \\ &\le \max_{\substack{\|X\|_F=1,\\ \langle X,\kappa_j^{\mathcal{L}}(I)\rangle=0,\\ j=1,\cdots,i-1.}} \|\mathcal{L}(I)\star X\|_F = \sigma_i^{\mathcal{L}}(I), \end{aligned}$$

where the last equality is due to the definition of the convolution eigenvalues. ∎

C. Proof of Theorem 1

We first establish the following lemma:

Lemma 1: If $B \neq 0$, then the Hessian matrix H defined by (13) is positive definite and obeys

$$h^{\mathcal{L}(B)}(X) \ge \frac{s_1 s_2}{(\sigma_{max}^{\mathcal{L}}(B))^2}\|X\|_F^2, \quad \forall X \in \mathbb{R}^{m_1\times m_2}.$$

Proof: It could be calculated that

$$\begin{aligned} h^{\mathcal{L}(B)}(X) &= \sum_{i=1}^{s_1 s_2} \frac{\|X\star\kappa_i^{\mathcal{L}}(B)\|_F^2}{(\sigma_i^{\mathcal{L}}(B))^2} \\ &\ge \frac{1}{(\sigma_{max}^{\mathcal{L}}(B))^2}\sum_{i=1}^{s_1 s_2}\|X\star\kappa_i^{\mathcal{L}}(B)\|_F^2 \\ &\doteq \frac{(\nu(X))^T H_1 \nu(X)}{(\sigma_{max}^{\mathcal{L}}(B))^2}, \end{aligned}$$

where $H_1 = \sum_{i=1}^{s_1 s_2}(\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B)))^T\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B))$. The convolution matrix $\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B))$ is a linear operator of $\kappa_i^{\mathcal{L}}(B)$ and can be explicitly written as

$$\mathcal{A}_{m_1,m_2}(\kappa_i^{\mathcal{L}}(B)) = [(\nu(\kappa_i^{\mathcal{L}}(B)))^T\Omega_1;\ (\nu(\kappa_i^{\mathcal{L}}(B)))^T\Omega_2;\ \cdots],$$

where $\{\Omega_j\}_{j=1}^{(m_1+s_1-1)(m_2+s_2-1)}$ is a set of selection matrices (taking value 0 or 1) independent of $\{\kappa_i^{\mathcal{L}}(B)\}$.

In the following, we shall show that $H_1 = s_1 s_2\mathcal{I}$, where $\mathcal{I}$ is the $m_1\times m_2$ identity matrix. We have

$$\begin{aligned} H_1 &= \sum_i\sum_j(\Omega_j)^T\nu(\kappa_i^{\mathcal{L}}(B))(\nu(\kappa_i^{\mathcal{L}}(B)))^T\Omega_j \\ &= \sum_j(\Omega_j)^T(\sum_i\nu(\kappa_i^{\mathcal{L}}(B))(\nu(\kappa_i^{\mathcal{L}}(B)))^T)\Omega_j \\ &= \sum_j(\Omega_j)^T\Omega_j, \end{aligned}$$

which reveals the fact that H_1 is actually independent of $\{\kappa_i^{\mathcal{L}}(B)\}$, as long as $\{\nu(\kappa_i^{\mathcal{L}}(B))\}$ are orthogonal to each other. So, $H_1 = s_1 s_2\mathcal{I}$ could be proved by assigning a special matrix to each $\kappa_i^{\mathcal{L}}(B)$. Let δ_{ij} denote an $s_1\times s_2$ matrix with the (i,j)th entry being 1 and everything else being 0 (i.e., delta kernel). Consider the special case where $\kappa_1^{\mathcal{L}}(B) = \delta_{11}, \kappa_2^{\mathcal{L}}(B) = \delta_{12},\cdots,\kappa_{s_1 s_2}^{\mathcal{L}}(B) = \delta_{s_1 s_2}$. Then for any $X\in\mathbb{R}^{m_1\times m_2}$, we have

$$\sum_{i=1}^{s_1 s_2}\|X\star\kappa_i^{\mathcal{L}}(B)\|_F^2 = s_1 s_2\|X\|_F^2 = s_1 s_2(\nu(X))^T\nu(X).$$

But we also have $\sum_{i=1}^{s_1 s_2}\|X\star\kappa_i^{\mathcal{L}}(B)\|_F^2 = (\nu(X))^T H_1\nu(X)$. So, the equality $(\nu(X))^T H_1\nu(X) = s_1 s_2(\nu(X))^T\nu(X)$ holds for any $X\in\mathbb{R}^{m_1\times m_2}$, i.e., $H_1 = s_1 s_2\mathcal{I}$. ∎

Proof of Theorem 1: By (13) and the convexity of the function $h^{\mathcal{L}(B)}(\cdot)$,

$$\begin{aligned} h^{\mathcal{L}(B)}(K_0-\hat{K}_0) &\le h^{\mathcal{L}(B)}(K_0) + h^{\mathcal{L}(B)}(\hat{K}_0) \\ &\le 2h^{\mathcal{L}(B)}(K_0) \le \frac{2s_1 s_2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2}. \end{aligned}$$

By Lemma 1, we also have

$$h^{\mathcal{L}(B)}(K_0-\hat{K}_0) \ge \frac{s_1 s_2}{(\sigma_{max}^{\mathcal{L}}(B))^2}\|K_0-\hat{K}_0\|_F^2.$$

Hence, $\|K_0-\hat{K}_0\|_F^2 \le 2(\sigma_{max}^{\mathcal{L}}(B))^2/(\sigma_{min}^{\mathcal{L}}(I_0))^2$. ∎

D. Proof of Corollary 1

Proof: Since $\hat{K}_0$ is deterministic, we have

$$\begin{aligned} \|K_0'-K_0''\|_F &= \|(K_0'-\hat{K}_0)+(\hat{K}_0-K_0'')\|_F \\ &\le \|K_0'-\hat{K}_0\|_F + \|\hat{K}_0-K_0''\|_F \le \sqrt{2}\frac{\sigma_{max}^{\mathcal{L}}(B)}{\sigma_{min}^{\mathcal{L}}(I_0')} \\ &\quad + \sqrt{2}\frac{\sigma_{max}^{\mathcal{L}}(B)}{\sigma_{min}^{\mathcal{L}}(I_0'')} \le \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

∎

E. Proof of Theorem 2

We need the following lemma to accomplish the proof.

Lemma 2: For any two matrices X and Y, we have

$$\|X \star Y\|_F \leq \|X\|_F \|Y\|_1,$$

where $\|\cdot\|_1$ is the ℓ_1-norm of a matrix.

Proof: Decompose Y into the sum of its negative part (denoted as Y^-) and nonnegative part (denoted as Y^+). By (2) and the convexity of matrix norms,

$$\begin{aligned}\|X \star Y\|_F &= \|X \star Y^+ + X \star Y^-\|_F \\ &\leq \|X \star Y^+\|_F + \|X \star Y^-\|_F \\ &\leq \|X\|_F \|Y^+\|_1 + \|X\|_F \|Y^-\|_1 = \|X\|_F \|Y\|_1.\end{aligned}$$

■

Proof of Theorem 2: In noisy case, the inequality (10) needs be changed to

$$\begin{aligned}\|K_0 \star \kappa_i^{\mathcal{L}}(B)\|_F &\leq \frac{\sigma_i^{\mathcal{L}}(B) + \|\mathcal{L}(N) \star \kappa_i^{\mathcal{L}}(B)\|_F}{\sigma_{min}^{\mathcal{L}}(I_0)} \\ &\leq \frac{\sigma_i^{\mathcal{L}}(B) + \epsilon\|\kappa_i^{\mathcal{L}}(B)\|_1}{\sigma_{min}^{\mathcal{L}}(I_0)} \leq \frac{\sigma_i^{\mathcal{L}}(B) + \sqrt{s_1 s_2}\epsilon}{\sigma_{min}^{\mathcal{L}}(I_0)}.\end{aligned}$$

Thus the inequality (13) becomes

$$h^{\mathcal{L}(B)}(K_0) \leq \sum_i \frac{(1 + \frac{\sqrt{s_1 s_2}\epsilon}{\sigma_i^{\mathcal{L}}(B)})^2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2} \leq \frac{s_1 s_2 (1 + \frac{\sqrt{s_1 s_2}\epsilon}{\sigma_{min}^{\mathcal{L}}(B)})^2}{(\sigma_{min}^{\mathcal{L}}(I_0))^2}.$$

Then by following the proof procedure of Theorem 1, this proof can be finished. ■

REFERENCES

[1] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, pp. 55–59, 1972.

[2] G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Opt. Lett.*, vol. 13, no. 7, pp. 547–549, 1988.

[3] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, May 1996.

[4] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2009, pp. 1033–1041.

[5] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Sparse representation based blind image deblurring," in *Proc. ICME*, Jul. 2011, pp. 1–6.

[6] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 104–111.

[7] N. Joshi, R. Szeliski, and D. Kriegman, "PSF estimation using sharp edge prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, Dec. 2011.

[9] C. L. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, Aug. 2004.

[10] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA, USA: MIT Press, 1964.

[11] T. S. Cho, S. Paris, B. K. P. Horn, and W. T. Freeman, "Blur kernel estimation using the radon transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 241–248.

[12] J. Jia, "Single image motion deblurring using transparency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[13] A. Levin, "Blind motion deblurring using image statistics," in *Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2006, pp. 841–848.

[14] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 157–170.

[15] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.

[16] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, 2006.

[17] C. Jordan, "Sur la série de Fourier," *Comp. Rendus Hebdomadaires Séances L'Académie Sci.*, vol. 92, pp. 228–230, 1881.

[18] Y.-L. You and M. Kaveh, "Fourth-order partial differential equations for noise removal," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1723–1730, Oct. 2000.

[19] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, p. 145, 2009.

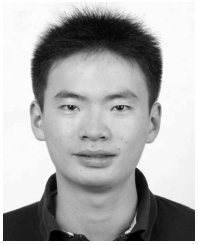
[20] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, no. 3, p. 73, 2008.

[21] F. Sroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 874–883, Jul. 2005.

[22] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Progressive inter-scale and intra-scale non-blind image deconvolution," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, 2008.

[23] S. Cho, J. Wang, and S. Lee, "Handling outliers in non-blind image deconvolution," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 495–502.

Guangcan Liu (M'11) received the bachelor's degree in mathematics and the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2004 and 2010, respectively. He was a Post-Doctoral Researcher with the National University of Singapore, Singapore, from 2011 to 2012, the University of Illinois at Urbana-Champaign, Champaign, IL, USA, from 2012 to 2013, Cornell University, Ithaca, NY, USA, from 2013 to 2014, and Rutgers University, Piscataway, NJ, USA, in 2014. Since 2014, he has been a Professor with the School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, China. His research interests mainly include machine learning, computer vision, and image processing.



Shiyu Chang received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2011 and 2014, respectively, where he is currently pursuing the Ph.D. degree. His research interests include pattern recognition, machine learning, computer vision, and data mining. He has served as a Program Committee Member and reviewer in many academic conferences and journals in the related areas.

Yi Ma (F'12) received the bachelor's degree in automation and applied mathematics from Tsinghua University, Beijing, China, in 1995, and the M.S. degree in electrical engineering and computer science, the M.A. degree in mathematics, and the Ph.D. degree in electrical engineering and computer science from the University of California at Berkeley, Berkeley, CA, USA, in 1997, 2000, and 2000, respectively. From 2000 to 2011, he was an Associate Professor with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA, and a Principal Researcher and the Group Manager of the Visual Computing Group with Microsoft Research Asia, Beijing, China, from 2009 to 2014. He is currently a Professor and the Associate Dean of the School of Information Science and Technology with ShanghaiTech University, Shanghai, China. His main research interests include computer vision and data science. He was a recipient of the David Marr Best Paper Prize at the International Conference on Computer Vision in 1999, the Honorable Mention for the Longuet-Higgins Best Paper Award at the European Conference on Computer Vision in 2004, the CAREER Award from the U.S. National Science Foundation in 2004, and the Young Investigator Program Award from the U.S. Office of Naval Research in 2005.