

Text Image Deblurring Using Kernel Sparsity Prior

Xianyong Fang, *Member, IEEE*, Qiang Zhou, Jianbing Shen^{ID}, *Senior Member, IEEE*, Christian Jacquemin, and Ling Shao^{ID}, *Senior Member, IEEE*

Abstract—Previous methods on text image motion deblurring seldom consider the sparse characteristics of the blur kernel. This paper proposes a new text image motion deblurring method by exploiting the sparse properties of both text image itself and kernel. It incorporates the L_0 -norm for regularizing the blur kernel in the deblurring model, besides the L_0 sparse priors for the text image and its gradient. Such a L_0 -norm-based model is efficiently optimized by half-quadratic splitting coupled with the fast conjugate descent method. To further improve the quality of the recovered kernel, a structure-preserving kernel denoising method is also developed to filter out the noisy pixels, yielding a clean kernel curve. Experimental results show the superiority of the proposed method. The source code and results are available at: <https://github.com/shenjianbing/text-image-deblur>.

Index Terms— L_0 -norm, motion deblurring, text image.

I. INTRODUCTION

TEXT documents such as advertisements, receipts, and information signboards captured by hand-held cameras are very common in our daily lives. They provide great conveniences for information acquisition and memorization. However, acquired documents are often motion blurred due to the camera shaking during photographing. How to eliminate such a degradation and obtain sharp texts are thus very important. In this paper, we propose a text image motion deblurring method by considering the sparsity of both image and kernel.

There has been abundant research on natural image deblurring [3]. A few methods [5] implicitly include the image sparsity with the dictionary, which has been adopted for other ill-posed inverse problems [1], [2], [6] without solving them

directly. However, its performance depends on the quality of the generated dictionary. Therefore, most studies explicitly incorporate the image sparsity as prior constraints, which generally utilize the heavy-tail or other types of sparse statistics, e.g., sparse dark channel prior, as a strong sparsity prior to regularize the solution. However, generalizing these techniques directly to text image deblurring could be problematic because they emphasize the importance of color variations and continuity in natural images while text images are usually sparsely toned with less contrast variation and more color-constant stroke regions. Fig. 1(b) shows such an example where the motion blurred text image [Fig. 1(a)] cannot be successfully deblurred with the state-of-the-art deblurring method [7].

The merits of the sparse toning of text images were not discovered in the text image deblurring community before their recent introduction by Pan *et al.* [8]. They discovered the sparse distributions of the text image intensity and its gradients, and utilized them as L_0 -norm-based constraints. Impressive results were obtained, outperforming previous studies [9], [10] by the traditional text-specific properties, e.g., sharp contrast between texts and background. Unfortunately, existing methods including the sparsity one [8] still cannot recover the clear latent texts, as shown in Fig. 1(c) and (d). Recently, a convolutional neural network (CNN)-based text deblurring method [11] was proposed. Fig. 1(e) shows the deblurred result with this method. Such an end-to-end learning approach without kernel estimation highly depends on the variety and quantity of the training samples. Therefore, there still lacks a robust text image motion deblurring method.

A common characteristic of the motion blur kernel is that it tends to be sparse, where most values close to zero. This is due to the fact that the motion kernel identifies the smooth motion path of the camera. But, to the best of our knowledge, none of the existing text deblurring methods explicitly considers the sparsity of the blur kernel as a L_p norm constraint. We believe the text motion deblurring method can be more robust by exploiting the kernel sparsity. In this paper, the sparsity of the blur kernel is explicitly modeled via a L_0 -regularizer to obtain a high quality estimation of the blur kernel.

Actually, the kernel sparsity has been widely adopted for natural image deblurring [12], [13]. Among them, the heavy-tailed style norm [14] is not sparse enough to depict the sparsity of motion blur kernels. The sparsity oriented norm mixtures are inefficient due to the high number of parameters to tune. Other common sparsity forms include L_1 norm [15], [16] or L_2 continuity norm mixtures [13]. However, the former has been embedded in the optimization process and thus can be omitted, while the latter complicates the deblurring

Manuscript received July 29, 2017; revised January 26, 2018; accepted October 8, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61502005, in part by the Anhui Science Foundation under Grant 1608085QF129, in part by the Key Programs for Science and Technology Development of Anhui Province under Grant 1604d0802004, and in part by the Beijing Natural Science Foundation under Grant 4182056. This paper was recommended by Associate Editor W. Hu. (Corresponding author: Jianbing Shen.)

X. Fang and Q. Zhou are with the School of Computer Science and Technology, Anhui University, Hefei 230601, China (e-mail: fangxianyong@ahu.edu.cn).

J. Shen is with the Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: shenjianbing@bit.edu.cn).

C. Jacquemin is with the Department of Computer Science, Université Paris-Sud, Université Paris-Saclay, 91403 Orsay, France (e-mail: christian.jacquemin@lmsi.fr).

L. Shao is with the Inception Institute of Artificial Intelligence, Abu Dhabi, UAE, and also with the School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, U.K. (e-mail: ling.shao@uea.ac.uk).

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/TCYB.2018.2876511

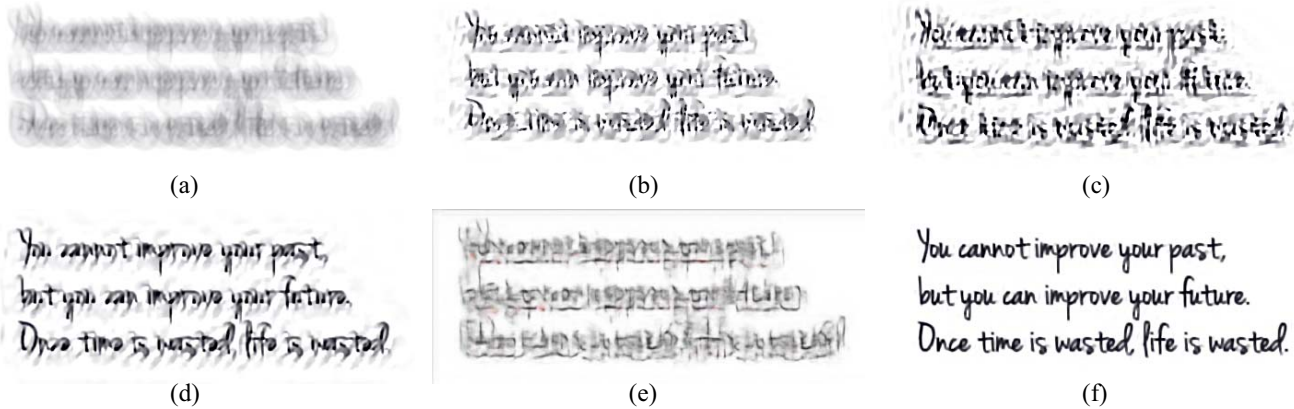


Fig. 1. Deblurring of a motion blurred text image by four state-of-the-art methods and our method, separately. Among the four methods, one (Pan *et al.* [7]) is for natural image deblurring with the other three, Pan *et al.* [8], Cao *et al.* [9], and Hradiš *et al.* [11], for text image deblurring.

process without significant performance improvement by our experiments. A simple yet efficient kernel sparsity solution is expected. Consequently, the proposed text image motion deblurring model combines the sparsity constraints for the image, its gradient, as well as the blur kernel. It is formulated as an optimization energy function where the sparsity priors are all represented through the L_0 norm. A new optimization method which integrates half-quadratic splitting and the fast conjugate gradient is also developed.

It is also observed that the estimated kernel image is often contaminated by noise surrounding the motion trajectory. Obtaining a clear and noise-free kernel is difficult due to the approximate estimation property of the optimization process. Existing methods often implicitly incorporate denoising into the kernel optimization process, however, noise is still difficult to be eliminated significantly. The noisy kernel will degrade the quality of the deblurred image.

To obtain a noise-free kernel, we propose a structure-preserving kernel denoising method, which explicitly denoises the kernel after kernel optimization and is thus independent of the kernel optimization technique. It can keep the shape and intensities of the kernel unchanged while adaptively filtering out the noisy pixels. This is more efficient than existing image denoising methods [17], [18] because they do not consider the special motion structure in the kernel, contrary to our method. The final deblurring result in Fig. 1(f) reveals that our method is much better than the other four state-of-the-art methods.

The main contributions of our method are summarized as follows.

- 1) A new text image motion deblurring model containing the sparsity constraints for both image and kernel, which extraordinarily matches the sparse text image and smooth camera motion path during the unstable photographing.
- 2) An efficient energy function representing the sparsity constraint only by the L_0 norm, which is optimized by a half-quadratic splitting and fast conjugate descent integrated algorithm.
- 3) A structure-preserving kernel denoising method to denoise the estimated blur kernel, which can keep the

shape and intensities of the kernel unchanged while eliminating the noises.

II. RELATED WORK

Text image motion deblurring is a class-specific study of the broader image deblurring area [19], [20]. Similar to other intensively studied inverse problems, e.g., image segmentation [21], object tracking [29], [33], and visual saliency [27], [39], various solutions in different aspects have been proposed, e.g., motion model, outlier, prior, and hardware reconfiguration. Our interest is to explore the merits of sparsity prior for the specific text image motion deblurring. Therefore, the sparsity in image deblurring methods is first reviewed. Then the status of text image deblurring will be reviewed.

A. Natural Image Deblurring Using Sparsity

The sparsity of natural images is implicitly included by some methods, e.g., Hu *et al.* [5] encoded the image patch with sparse coefficients using an over-complete dictionary. Such methods can recover the latent image without solving the ill-posed deconvolution, and thus have also been adopted for other inverse problems [1], [2], [4], [6]. However, their optimal performance is affected by the quality of dictionary, considering the vast differences among natural images.

Therefore, most studies explicitly include the image sparsity as a prior constraint, where the heavy-tailed gradient distribution of image statistics is widely formulated by different L_p ($0 < p < 1$) norms [12], [22]. Some approximation formats are also proposed, such as Gaussian mixture model and piecewise function approximation [16]. However, the heavy-tailed distribution or the norm mixtures is generally not suitable for text images which are usually colored uniformly.

The gradient sparsity is also introduced in other formats, e.g., the approximated L_0 norm [23] and the normalized L_1 norm [15]. Still there can be limits in describing the sparsity of text images [8]. Some studies consider extending the pixel-wise sparsity constraint to other forms. For example, Cai *et al.* [24] represented the sparsity of framelet with the L_1 norm, while Pan *et al.* [7] and Yan *et al.* [25] included the sparsity priors of the dark and bright channels with the

L_0 norm. However, these studies can fail when applied to text images, e.g., the dark channel may not exist for most text images which contain uniformly colored strokes with nonzero-valued colors.

The sparsity of the motion kernel is also studied. Some researchers [14] simulated it to be the same heavy-tailed distribution as the natural image gradient distribution. Others such as [15] and [16] modeled it by the L_1 norm. Other similarly complicated prior constraints make use of several kernel priors together. The prior constraint associated with the sparse L_0 and continuous L_2 norms [13]. The merits of our kernel sparsity constraint when comparing with these methods are as follows.

- 1) It is closer to the true motion kernel intensity distribution than the heavy tailed sparsity constraints.
- 2) It does not use the L_1 norm due to the embedded L_1 property of the blur kernel.
- 3) It is easier to deploy the optimization process than the complicated ones because of only one coefficient.

B. Text Image Motion Deblurring

Some methods are specifically designed for text images. Chen *et al.* [26] computed the intensity distribution prior, observing that the intensities between the foreground texts and the background are significantly different. Their method relies on text segmentation to separate the background and the two-tone or bi-level property of text, similar to some two-tone image deblurring studies [28], [30]. However, these methods may not be effective for real images which are usually not two-tone. Cho *et al.* [10] used the stroke width transform to separate the text image into text and nontext regions, which gives poor results when the characters are small and connected. Lou *et al.* [31] exploited the sparse characteristics of natural text images and estimated the latent image directly with an over-complete dictionary. Cao *et al.* [9] also adopted a sparsity-based technique and created two types of dictionaries, one for the natural scene and the other for the characters at multiple scales. Pan *et al.* [8] further formulated such a sparsity as L_0 constraints for the intensity and gradient distributions of the latent image, without a training process. We also use these sparsity constraints in our deblurring model.

However, existing text image deblurring methods may fail due to neglecting the sparsity of the motion blur kernel, even though such a sparsity has already been considered in natural image deblurring methods. As far as we know, it has never been applied to text image motion deblurring. Therefore, we propose to make use of kernel sparsity through the L_0 -norm-based kernel constraint. Our model drops the L_2 kernel prior widely utilized in most deblurring methods [8], [13], [23] due to the following reasons.

- 1) The data term of the deblurring model has partially included the kernel continuity.
- 2) The L_2 norm of the motion blur kernel is very small in comparison with its L_0 kernel norm and thus its contribution to the convergence can be negligible.
- 3) The additional parameter tuning and optimization burden will be introduced with the additional L_2 norm to make the optimization more complex.

Along with the booming research of deep neural networks [36], [39], the CNN was introduced for text image deblurring [11]. Taking the raw blurry patch as input, the method can recover the latent image via the trained 15-layer CNN. This idea is interesting, but its performance highly relies on the quantity and variety of training documents or text patches. In addition, the training process is time-consuming even with high quality GPU implementation due to a vast number of parameters to tune.

Existing text deblurring methods often take the blur kernel as a single objective to optimize. However, in natural image deblurring, recently, much better results are obtained through separated optimizations [32]. The kernel is separated into three parts (trajectory, intensity, and point spread function) and each part is optimized independently with different priors. However, the kernel width along the skeleton is kept unchanged, which may be incorrect for real photographs due to the projective transform and random camera motion. In addition, processing three separate optimizations instead of a single combined one increases the computational load and may introduce noise into the kernel. Our proposed denoising method relaxes the requirement of a fixed width and instead filters out the trajectory noises along its skeleton without modifying the kernel structure. It also does not require additional optimization.

Image denoising is a traditional research topic with a lot of methods available, including the well-known ones, i.e., the nonlocal means, K-SVD, and BM3D. These methods also have a lot of variants, which have almost reached the optimally possible denoising performance [34]. They usually aim for natural images, where optimization or learning is often incorporated. However, our denoising method simply uses the trajectory as the guidance without optimization or learning and is specifically designed for the motion-based kernel.

III. SPARSITY-BASED MOTION DEBLURRING MODEL AND ITS ENERGY FUNCTION

We now present the novel sparsity-based model where only the sparsities for the image intensity, image gradient, and kernel are taken as prior constraints.

A. New Model

Assume that \mathbf{B} , \mathbf{K} , and \mathbf{L} are the blurred image, the blur kernel, and the latent image, respectively. Image deblurring can be formulated as an energy minimization problem whose objective function consists of three terms

$$\min_{\mathbf{K}, \mathbf{L}} D(\mathbf{B}, \mathbf{K}, \mathbf{L}) + \gamma_K P(\mathbf{K}) + \gamma_L P(\mathbf{L}) \quad (1)$$

where $D(\mathbf{B}, \mathbf{K}, \mathbf{L})$ represents the data term. It measures the difference between the \mathbf{B} and the convolution between \mathbf{K} and \mathbf{L}

$$D(\mathbf{B}, \mathbf{K}, \mathbf{L}) = \|\mathbf{B} - \mathbf{K} \otimes \mathbf{L}\|_2^2 \quad (2)$$

where \otimes means the convolution operation. $P(\mathbf{K})$ and $P(\mathbf{L})$ denote, respectively, the prior regularizers of \mathbf{K} and \mathbf{L} , with γ_K and γ_L being their coefficients.

The prior knowledge of \mathbf{K} and \mathbf{L} is crucial to regularize such an ill-posed inverse problem, and to facilitate estimating the

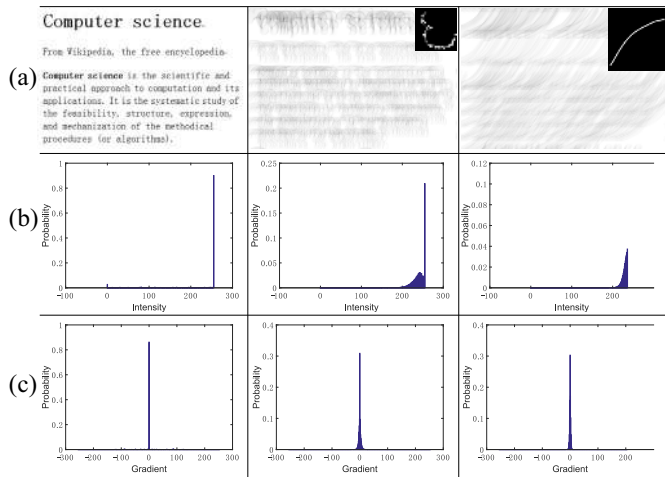


Fig. 2. Sparse properties of the image intensity and gradients of motion blurred text images. An example clear image and its four differently blurred images are shown. (a) Image and its blurred ones. (b) Intensity distributions. (c) Gradient distributions.

blur kernel and the latent image, as a result. Different types of prior constraint represented by different matrix norms have been brought forward so far [13].

The text sparsity priors are adopted from the sparsity properties discussed by Pan *et al.* [8]. Observing the text image, we can see that the characters and background usually have uniform intensities in the clear image and nonuniform ones in the blurred one. In the intensity histogram of the image, only few intensity positions will be significantly higher than other places for a clear image, while there is no such a zero peak for a blurred image. The same observation can also be made in the gradient distribution. Fig. 2 shows the intensity and gradient distributions of an example image blurred by different motion blur kernels. It can be seen that: 1) the intensities of the clear image contain a lot of zeros while these of the blurred image are close to zero but not null and 2) the clear image contains more zero gradients than the blurred image.

A motion blur kernel identifies the smooth path of the camera, so it tends to be sparse, with most values outside the sparse motion path being null. Consequently, the kernel can also be constrained as a sparse kernel. Fig. 3 shows two example kernels obtained from the naturally blurred images by the method in [8]. Most of kernel values are zero for each kernel and thus these kernels are all very sparse as their intensity histograms show. Therefore, we take the sparsities of both image and its kernel as the priors. Now, let us discuss the energy function including these constraints for optimizing the proposed model.

B. Energy Function

The values of the image intensity and its gradient, and the blur kernel are mostly zeros as we have discussed. Therefore, their prior distributions can be regularized by the L_0 norm which counts the number of nonzero values of a vector. Taking these priors together, we can obtain the energy function for the proposed model with only the L_0 sparsity norm as the constraints for the image and its gradient, and the blur kernel.

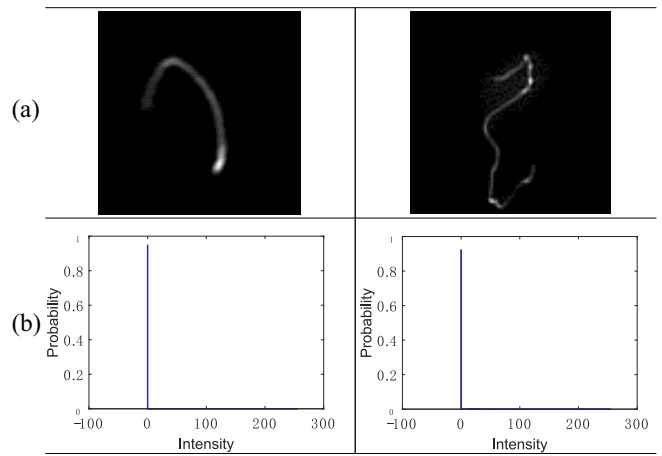


Fig. 3. Intensity histograms of two example kernels estimated from five naturally shaking text images by the method of Pan *et al.* [8]. (a) Two kernels. (b) Intensity distributions.

The sparse properties of the image and its gradients can be formulated as

$$P_s(\mathbf{L}) = \|\mathbf{L}\|_0 \quad (3)$$

$$P_s(\nabla\mathbf{L}) = \|\nabla\mathbf{L}\|_0. \quad (4)$$

The image prior is then obtained by taking (3) and (4) together, with σ being the weight

$$P(\mathbf{L}) = \sigma P_s(\mathbf{L}) + P_s(\nabla\mathbf{L}). \quad (5)$$

Similarly, the sparse constraint for the kernel can be

$$P(\mathbf{K}) = \|\mathbf{K}\|_0. \quad (6)$$

The interesting part of the new model is that it is perhaps the first one to include the kernel prior only by the L_0 -norm represented sparsity constraint. The advantage of this configuration will be discussed in the following section.

IV. ANALYSIS ON THE KERNEL PRIOR CONSTRAINT

This section mainly analyzes the two L_0 -norm related forms of kernel sparsity, the L_1 -norm, and the L_2 -norm incorporated mixture of norms, and shows that either the L_1 norm or the L_2 included norm mixture is not necessary for a proper regularization. About the other two types of published norms: 1) the heavy-tail style norm cannot correctly depict the kernel sparsity distribution (Fig. 3) and thus may not be a proper choice and 2) the mixture of norms without the L_2 norm can be too complicated for an efficient optimization with more parameters to tune in comparison with our single L_0 -norm. Therefore, these norms are not discussed here.

A. L_1 Norm-Based Sparsity Prior

Assume that the motion blur kernel \mathbf{K} with the size of $m \times n$. Generally, it has two properties as follows.

1) *Non-Negativity*: The element of \mathbf{K} , $K_{i,j}$, is always non-negative, i.e.,

$$K_{i,j} \geq 0. \quad (7)$$

2) *Normalization*: The sum of the elements of \mathbf{K} is equal to 1

$$\sum_{i=1}^m \sum_{j=1}^n K_{i,j} = 1. \quad (8)$$

These two conditions amount to

$$\|\mathbf{K}\|_1 = 1. \quad (9)$$

In practice, these two conditions are usually implicitly taken in charge by existing proposed deblurring models including ours (Algorithm 1 in Section VII) through a kernel normalization process. The negative elements are first set to zeros and then the kernel is normalized to one. Some studies [35], [37] even explicitly show this process. Therefore, it is not necessary to include an L_1 kernel norm explicitly.

B. Insignificant Kernel Continuity Regularizer

The constraint of kernel continuity is often formulated by L_2 norm, however, such a regularizer could be insignificant. First, the kernel continuity is already partially included in the data term (2). Our following analysis will also show that the L_2 -norm is very small for a blur kernel in comparison with the L_0 -norm, and thus it can be negligible for the optimization process. Second, the additional continuity regularizer introduces one additional parameter to tune, where an additional optimization budget may make the deblurring system more difficult to converge.

1) *Negligible Continuity Regularizer*: Let us first present a proposition based on the two properties of a blur kernel.

Proposition 1: We assume that the size of motion blur kernel \mathbf{K} is $m \times n$. Its L_2 -norm is smaller than its L_0 -norm

$$\|\mathbf{K}\|_2 < \|\mathbf{K}\|_0. \quad (10)$$

Proof: Only one nonzero element in the blur kernel cannot blur an image and, therefore, the number of the nonzero elements of \mathbf{K} is always larger than 1, i.e.,

$$\sum_{i=1}^m \sum_{j=1}^n \text{sgn}(K_{i,j}) \geq 2 \quad (11)$$

where $\text{sgn}(\cdot)$ is the sign function.

According to the non-negativity of \mathbf{K} by (7), (11) suggests that $K_{i,j} < 1$ if further considering the normalization property of \mathbf{K} by (8) and, consequently,

$$\|\mathbf{K}\|_2 = \sqrt{\|\mathbf{K}\|_2^2} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n K_{i,j}^2} < \sqrt{\left(\sum_{i=1}^m \sum_{j=1}^n K_{i,j}\right)^2} = 1. \quad (12)$$

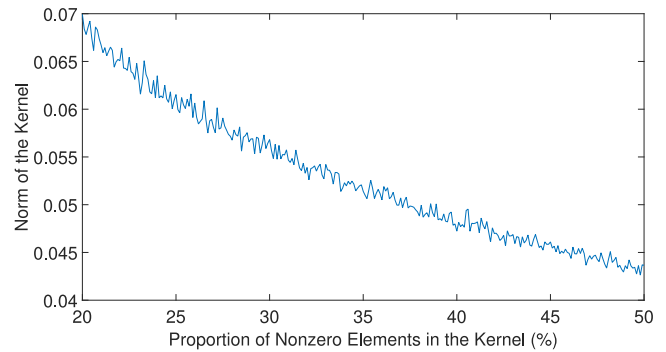
The L_0 -norm of \mathbf{K} based on (7) can be formulated as

$$\|\mathbf{K}\|_0 = \sum_{i=1}^m \sum_{j=1}^n \text{sgn}(K_{i,j}). \quad (13)$$

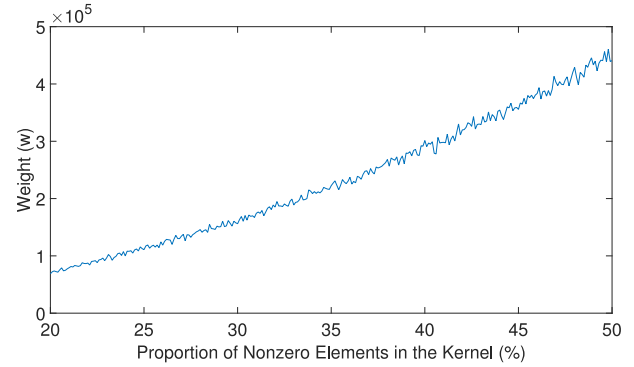
Considering the nonzero property in (11), we obtain

$$\|\mathbf{K}\|_0 \geq 2. \quad (14)$$

Finally, combining (12) and (14) leads to (10). ■



(a)



(b)

Fig. 4. L_2 norms of motion blur kernels and the associated weights (17) balancing the L_0 norm and the L_2 norm. (a) L_2 norms. (b) Weights.

In practice, the motion trajectory is often very sparse with mostly zero values in the blur kernel (Fig. 3), whose space complexity can be assumed as almost linear, i.e., it approximates $O(\max(m, n))$. In other words, $\|\mathbf{K}\|_0$ is in the order of $\alpha(\max(m, n))$ with α being a scalar and thus significantly larger than $\|\mathbf{K}\|_2$ which is less than 1 (12). Fig. 4(a) shows the L_2 norms of an example 41×41 kernel which contains 20%–50% nonzero elements randomly generated to simulate the sparsity of the motion kernel, considering that generally more than half kernel elements are zeros. The L_0 norm for this sparse range is between 336 and 840, while the L_2 norm is only between 0.0426 and 0.0699 in Fig. 3. The difference between the two norms is apparently large: the L_0 norm can be about $1.9718e4$ times the value of the L_2 norm.

The above observation on the kernel sparsity can be summarized as the following assumption based on Proposition 1.

Assumption 1: For a motion blur kernel \mathbf{K} of an image captured by a shaking camera, its L_2 -norm is generally much smaller than its L_0 -norm

$$\|\mathbf{K}\|_2 \ll \|\mathbf{K}\|_0. \quad (15)$$

This assumption and the previous discussions on both L_2 and L_0 norms [e.g., (12) and (14)] show that the kernel continuity prior can be neglected in the deblurring model.

2) *Additional Computation Budget if Additionally Including the Continuity Regularizer*: L_2 norm for the kernel continuity prior can be scaled to the same importance as L_0 norm if both of them are adopted. However, an additional parameter should

be chosen to balance the contributions of them for the kernel regularizer in comparison with (6), i.e.,

$$P(\mathbf{K}) = \gamma_{K_0} \|\mathbf{K}\|_0 + \gamma_{K_2} \|\mathbf{K}\|_2^2 \quad (16)$$

where γ_{K_0} and γ_{K_2} are the coefficients. Such a balance can be reformulated as follows by considering their equal contributions to the total energy:

$$\|\mathbf{K}\|_0 = w \|\mathbf{K}\|_2^2 \quad (17)$$

with the weight $w = (\gamma_{K_2}/\gamma_{K_0})$ according to (16).

However, finding a proper w is nontrivial because the L_2 norms of different images can be quite different. For example, Fig. 4(b) shows w oscillates drastically from $6.8700e4$ to $4.6064e5$ because of the varying L_2 norms [Fig. 4(a)]. It is thus difficult to tune the contribution of the L_2 norm with such a changing weight if both L_0 and L_2 norms of the unknown kernel are considered for blind deblurring.

Because of norm mixture, additional optimization related problems will emerge when incorporating the extra L_2 norm. On one hand, this optimization can be unstable because of the difficulty in selecting a creditable weight. On the other hand, additional costs (e.g., time and storage) will be introduced even though the optimized solution of (2) may be obtained with this dual norm-based kernel constraint. Therefore, only the L_0 sparsity norm is adopted in our method for the kernel without considering the explicit continuity constraint incorporated by other models [8], [13].

V. OPTIMIZATION VIA HALF-QUADRATIC SPLITTING

We now present the optimization method for the proposed model. \mathbf{K} and \mathbf{L} in the deblurring model (1) are alternatively optimized via the following two functions, separately:

$$\min_{\mathbf{K}} D(\mathbf{B}, \mathbf{K}, \mathbf{L}) + \gamma_K P(\mathbf{K}) \quad (18)$$

$$\min_{\mathbf{L}} D(\mathbf{B}, \mathbf{K}, \mathbf{L}) + \gamma_L P(\mathbf{L}). \quad (19)$$

The properties of \mathbf{K} given by (7)–(9) are mandatorily held in the optimization process (see step 17 in Algorithm 1), which can also eliminate the scale ambiguity of \mathbf{K} and thus improve the estimation accuracy. Both (18) and (19) contain a computationally intractable L_0 regularization term. To work around this shortcoming, the half-quadratic splitting-based L_0 minimization method [38] can be used, which can approximately solve the L_0 optimization iteratively with auxiliary variables. This scheme makes both (18) and (19) convex and thus ensures their convergences.

However, \mathbf{K} cannot be obtained robustly if computed directly from the intensity [40]. To resolve this problem, we further adopt the fast conjugate gradient method [41] to compute it in the gradient space after a half-quadratic splitting. In comparison with the traditional conjugate gradient method, the fast method can converge faster with a much fewer number of Fourier transforms, thanks to the derivatives adopted.

A. Kernel Optimization With Fast Conjugate Gradient

To compute \mathbf{K} while fixing \mathbf{L} , (18) is reformulated as follows in the gradient space:

$$\min_{\mathbf{K}} D(\nabla \mathbf{B}, \nabla \mathbf{L}, \mathbf{K}) + \gamma_K P(\mathbf{K}). \quad (20)$$

The corresponding alternative model, after introducing the auxiliary variable \mathbf{N} , is

$$\min_{\nabla \mathbf{K}, \mathbf{N}} D(\nabla \mathbf{B}, \nabla \mathbf{L}, \mathbf{K}) + \theta_1 \|\mathbf{N} - \mathbf{K}\|_2^2 + \gamma_K \|\mathbf{N}\|_0 \quad (21)$$

with θ_1 being the coefficient. It is often set to a relatively large number (e.g., 2000 in our experiment) so that the above new form approximates (20). This new form can be solved by estimating \mathbf{K} and \mathbf{N} alternatively.

1) *Estimating \mathbf{K}* : Denote x and y being two derivative directions. \mathbf{K} is computed by the fast conjugate gradient method through the following energy function:

$$E(\mathbf{K}) = \|\mathbf{B}_* - \mathbf{K} \otimes \mathbf{L}_*\|_2^2 + \theta_1 \|\mathbf{N} - \mathbf{K}\|_2^2 \quad (22)$$

$$(\mathbf{L}_*, \mathbf{B}_*) \in \{(\nabla_x \mathbf{L}, \nabla_x \mathbf{B}), (\nabla_y \mathbf{L}, \nabla_y \mathbf{B}), (\nabla_{xx} \mathbf{L}, \nabla_{xx} \mathbf{B}), (\nabla_{yy} \mathbf{L}, \nabla_{yy} \mathbf{B}), ((\nabla_{xy} \mathbf{L} + \nabla_{yx} \mathbf{L}), \nabla_{xy} \mathbf{B})\}. \quad (23)$$

Consequently, \mathbf{K} can be obtained from (22) by

$$\frac{\partial E(\mathbf{K})}{\partial \mathbf{K}} = \mathbf{X}^T \mathbf{X} \mathbf{K} + \theta_3 (\mathbf{N} - \mathbf{K}) - \mathbf{X}^T \mathbf{Y} = 0 \quad (24)$$

where \mathbf{X} and \mathbf{Y} are composed of the partial derivatives of \mathbf{L} and \mathbf{B} , respectively.

However, $\mathbf{X} \mathbf{K}$ is very large and therefore this computation is time and space consuming. To accelerate it, we further use the fast Fourier transform (FFT) [denoted as $F(\cdot)$]. In particular, $F(\mathbf{X}_*)$ and $\mathbf{X}^T \mathbf{Y}$ can be precomputed before optimization. $\mathbf{X}^T \mathbf{X} \mathbf{K}$ can be computed through FFT as

$$\mathbf{X}^T \mathbf{X} \mathbf{K} = F^{-1} \left(\sum_{\mathbf{L}_*} w_* \overline{F(\mathbf{L}_*)} \odot F(\mathbf{L}_*) \odot F(\mathbf{K}) \right) \quad (25)$$

where $F^{-1}(\cdot)$ represents the inverse FFT; \odot means the element-wise multiplication; and w_* is the weight for each element of \mathbf{L}_* . $\sum_{\mathbf{L}_*} w_* \overline{F(\mathbf{L}_*)} \odot F(\mathbf{L}_*)$ can also be precomputed before the optimization starts.

2) *Estimating \mathbf{N}* : Optimization of \mathbf{N} is similar to \mathbf{K}

$$E(\mathbf{N}) = \theta_1 \|\mathbf{N} - \mathbf{K}\|_2^2 + \gamma_K \|\mathbf{N}\|_0. \quad (26)$$

The solution for \mathbf{N} is expressed as [38]

$$\mathbf{N} = \begin{cases} \mathbf{K} & |\mathbf{K}|^2 \geq \frac{\gamma_K}{\theta_1} \\ 0 & \text{else.} \end{cases} \quad (27)$$

B. Image Optimization

To estimate \mathbf{L} , (19) is reformulated by fixing \mathbf{K} and introducing two auxiliary variables, \mathbf{P} and $\mathbf{M} = (\mathbf{M}_x, \mathbf{M}_y)^T$

$$\min_{\mathbf{L}, \mathbf{P}, \mathbf{M}} D(\mathbf{B}, \mathbf{L}, \mathbf{K}) + \theta_2 \|\mathbf{L} - \mathbf{P}\|_2^2 + \theta_3 \|\nabla \mathbf{L} - \mathbf{M}\|_2^2 + \gamma_L (\sigma \|\mathbf{P}\|_0 + \|\mathbf{M}\|_0) \quad (28)$$

where coefficients θ_2 and θ_3 are set to be relatively large numbers as θ_1 for approximating (19). \mathbf{P} and \mathbf{M} are initialized to

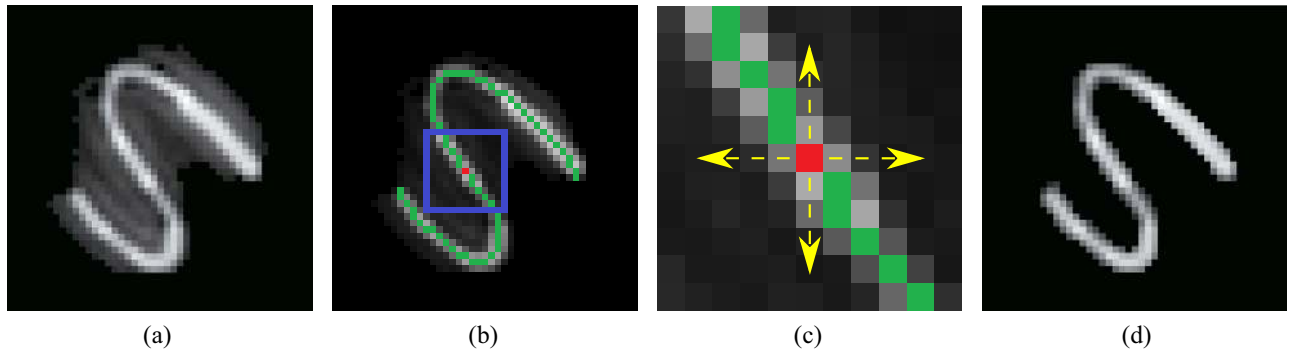


Fig. 5. Principle of the structure-preserving kernel denoising. (a) Noisy kernel. (b) Skeleton in green for (a) and the corresponding sliding window in blue on the current pixel in red. (c) Enlarged patch inside the sliding window of (b). (d) Final denoised kernel. The yellow direction lines in (c) give the directions used to filter out the noisy pixels surrounding the current pixel.

be 0. \mathbf{L} , \mathbf{P} , and \mathbf{M} can be obtained alternatively by minimizing each one independently while fixing the others.

For \mathbf{P} and \mathbf{M} , their optimization equations are similar to that of N when computing K

$$E(\mathbf{P}) = \theta_2 \|\mathbf{L} - \mathbf{P}\|_2^2 + \gamma_L \sigma \|\mathbf{P}\|_0 \quad (29)$$

$$E(\mathbf{M}) = \theta_3 \|\nabla \mathbf{L} - \mathbf{M}\|_2^2 + \gamma_L \|\mathbf{M}\|_0. \quad (30)$$

For \mathbf{L} , its optimization equation is

$$E(\mathbf{L}) = D(\mathbf{B}, \mathbf{L}, \mathbf{K}) + \theta_2 \|\mathbf{L} - \mathbf{P}\|_2^2 + \theta_3 \|\nabla \mathbf{L} - \mathbf{M}\|_2^2. \quad (31)$$

Equation (31) can be solved as a least squares problem [8]. Accordingly, the closed-form solution is

$$\mathbf{L} = F^{-1} \left(\frac{\overline{F(\mathbf{K})} F(\mathbf{B}) + \theta_2 F(\mathbf{P}) + \theta_3 F_{\mathbf{M}}}{\overline{F(\mathbf{K})} F(\mathbf{K}) + \theta_2 + \theta_3 \overline{F(\nabla)} F(\nabla)} \right) \quad (32)$$

where $\overline{F(\cdot)}$ is the complex conjugate of $F(\cdot)$, with $F_{\mathbf{M}} = \overline{F(\nabla_x)} F(\mathbf{M}) + \overline{F(\nabla_y)} F(\mathbf{M})$.

VI. STRUCTURE-PRESERVING KERNEL DENOISING

The optimization process may result in so noisy kernels that the motion trajectory can be barely identified as the example kernel shown in Fig. 5(a). Deblurring the image with such a kernel would degrade the quality of the recovered image. The real kernel associated with the continuous motion process is continuous and without noise. Therefore, a denoising step can be incorporated inside the deblurring process to obtain a clean kernel made of a continuously varying width along its trajectory. The structure-preserving kernel denoising method is proposed based on this idea.

A. Principle of the Denoising Method

Considering that the blur kernel is generated by the motion of the camera during the exposure time τ , we can define the clean kernel and its noisy version being $\mathbf{K}(t)$ and $\mathbf{K}'(t)$, respectively, according to the time t . Their relationship is

$$\mathbf{K}'(t) = \mathbf{K}(t) + \mathbf{n}(t) \quad (33)$$

with $\mathbf{n}(t)$ being the noise function. The denoising process turns into finding $\mathbf{K}'(t)$ at t that can minimize the difference between

$\mathbf{K}(t)$ and $\mathbf{K}'(t)$ during τ

$$\min_{\mathbf{K}'(t)} \int_{t=0}^{\tau} |\mathbf{K}(t) - \mathbf{K}'(t)|. \quad (34)$$

However, the kernel at t is not a simple pixel but an irregular part of the motion trajectory with different intensities and span widths, corresponding to different duration times and projective deformations due to the camera speed and motion path, separately. Therefore, (34) can be reformulated to minimize the differences of the shape and pixel intensity composed kernel structures between \mathbf{K} and \mathbf{K}'

$$\min_{\mathbf{K}'(t)} \int_{t=0}^{\tau} |S(\mathbf{K}(t)) - S(\mathbf{K}'(t))| \quad (35)$$

with $S(\cdot)$ representing the structure function.

Equation (35) seeks the best structure of \mathbf{K}' approximating that of \mathbf{K} without noisy pixels. However, the solution of (35) is nontrivial considering the discrete kernel intensity and irregular kernel shape. Checking (35), we can see that the best \mathbf{K}' is the one closest to \mathbf{K} at each time t during τ , i.e., the best \mathbf{K}' is the one closest to \mathbf{K} at each position on the kernel structure. Consequently, (35) can be reformulated as the following discrete form at each discrete time \bar{t} :

$$\min_{\mathbf{K}'(\bar{t})} |S(\mathbf{K}(\bar{t})) - S(\mathbf{K}'(\bar{t}))|. \quad (36)$$

For a kernel image, its trajectory represents the kernel structure and, therefore, (36) means the optimized kernel can be obtained by filtering its trajectory from the beginning to the end. Consequently, we can propose the sliding window-based kernel denoising strategy. It uses a sliding window to adaptively filter the trajectory for noise removal so that a clean blur kernel with a width-varying trajectory curve can be obtained. The structure of the resulting kernel \mathbf{K}' is as close as possible to that of the original kernel \mathbf{K} .

B. Structure-Preserving Denoising Using Sliding Window

The denoising process consists of two main steps (Fig. 5): 1) skeleton detection and 2) window sliding. The first step locates the skeleton of the trajectory in the noisy kernel image. The skeleton of the trajectory is a pixel-thin curve structure centered inside the shape and is used as the sliding path of the

center of the sliding window. The second step slides a window along the skeleton and filters out the noisy pixels.

In the first step, the skeleton detection method adopted in [42] is taken. It uses the ordered region-growing algorithm to extract the connectivity among all nodes in the kernel image represented as an acyclic graph. The skeleton can then be obtained by specifying endpoints or through a pruning process. Fig. 5(b) shows the detected skeleton for the kernel shown in Fig. 5(a), which depicts the camera motion path during capture. In the second step, a sliding window is moved along the skeleton pixel by pixel so that the noisy pixels surrounding the current skeleton pixel can be eliminated. The denoised kernel can be obtained finally, whose trajectory will continually vary in intensity and width along the skeleton, reflecting the true photographing process of different durations and paths. Then the structure, i.e., the shape and intensities, of the true kernel is kept after the denoising step.

The noisy pixels normally lie around the motion path inconsistently or show rather lower intensities than the pixels that belong to the real kernel [Fig. 5(a)]. Therefore, the noisy pixels are searched along four directions from the skeleton pixel, i.e., two horizontal and vertical directions [the yellow arrows in Fig. 5(c)]. The pixels with significantly low values are set to be the noisy pixels and then eliminated.

However, the intensities of a true kernel can vary anywhere along the skeleton. Then a unified threshold is inadequate for judging whether the intensity of a pixel surrounding the current skeleton pixel is significantly low or not. In the proposed method, we combine a global and a local approaches. First, a global threshold is used before the sliding starts to eliminate the apparent noisy pixels in the whole kernel. Second, a local threshold is applied during the sliding to eliminate the missed noisy pixels surrounding the current skeleton pixel. In practice, the global threshold is usually set to one tenth of the max intensity values in the whole patch with the local one set to three tenths of the intensity of the current skeleton pixel. Fig. 5(d) shows the denoised kernel for the original noisy one, where the kernel structure keeps unchanged.

VII. PROPOSED ALGORITHM

The image is deblurred in a multiscale manner [8], [20] which iteratively estimates the kernel and the latent image in the coarse-to-fine style. For each level of the pyramid, the residual deconvolution of Pan *et al.* [8] is adopted. The method first estimates L_l using the Laplacian prior [43] and then computes L_0 without taking the intensity prior into account. The difference image between L_0 and L_l is filtered by a bilateral filter so that ringing effects can be suppressed. The final clear latent image L is then obtained by subtracting the filtered difference image from L_l .

Algorithm 1 shows the main steps for the deblurring process. In the algorithm: 1) $maxLevel$ is determined by the size of the minimal pyramid image of K , which is set in the same way as Pan *et al.* [8]; 2) $maxIter$ is set to be 5; and 3) K_1 is initialized to be $\mathbf{0}$ except its two central elements are set to 0.5. The coefficients of constraints γ_L , γ_K , and σ are usually set to 0.004, 0.004, and 1, respectively.

Algorithm 1 L_0 -Norm Text Image Deblurring Algorithm

Input: B , max iterations $maxIter$, max pyramid level $maxLevel$
Output: K , L

- 1: Build the pyramid images of B , $B_i (1 \leq i \leq maxLevel)$, from the coarsest level $i = 1$ to the finest level $i = maxLevel$;
- 2: $L_1 = B_1$
- 3: Initialize K_1 ;
- 4: $iLevel = 1$;
- 5: **while** $iLevel \leq maxLevel$ **do**
- 6: $iTer = 1$;
- 7: **while** $iTer \leq maxIter$ **do**
- 8: Estimate L_{iLevel} by the half-quadratic splitting method with B_{iLevel} and K_{iLevel} (Section V-B);
- 9: Estimate K_{iLevel} in gradient space by the half-quadratic splitting and fast conjugate gradient methods with B_{iLevel} and L_{iLevel} (Section V-A);
- 10: Normalize K_{iLevel} (Section IV-A);
- 11: $iTer = iTer + 1$;
- 12: **end while**
- 13: Upsample K_{iLevel} to be $K_{iLevel+1}$;
- 14: $iLevel = iLevel + 1$;
- 15: **end while**
- 16: Denoise $K_{maxLevel}$ by the structure-preserving denoising (Section VI-A);
- 17: Normalize $K_{maxLevel}$ (Section IV-A);
- 18: $K = K_{maxLevel}$;
- 19: Estimate L by the residual deconvolution method with B and K .

VIII. EXPERIMENTAL RESULTS

Three state-of-the-art text image deblurring methods are adopted primarily for comparing the experimental results qualitatively and quantitatively: 1) the CNN-based method by Hradiš *et al.* [11]; 2) the L_0 -regularized image intensity and gradient priors by Pan *et al.* [8]; and 3) the dictionary-based method by Cao *et al.* [9]. Hradiš *et al.* [11] is a direct end-to-end learning-based method, while Pan *et al.* [8], Cao *et al.* [9], and ours take the traditional optimization. Neither of Pan *et al.* [8] and Cao *et al.* [9] considers the kernel sparsity and in most experiments, to show the performance of our kernel denoising strategy, our method is presented in two versions, *our method without denoising* and *our method with denoising*, meaning the deblurring by our method without or with the proposed kernel denoising strategy, respectively. The artificial dataset of Pan *et al.* [8] is taken for the statistical performance comparison, which contains 15 images and eight kernels.

A. Qualitative Results

We first show the deblurring performances with different sparsity-based methods, including the natural image oriented methods: Krishnan *et al.* [15], Xu *et al.* [23], and Zuo *et al.* [14]. Krishnan *et al.* [15] adopted the ratio of L_1 to L_2 norm and the L_1 norm, respectively, for the image and the

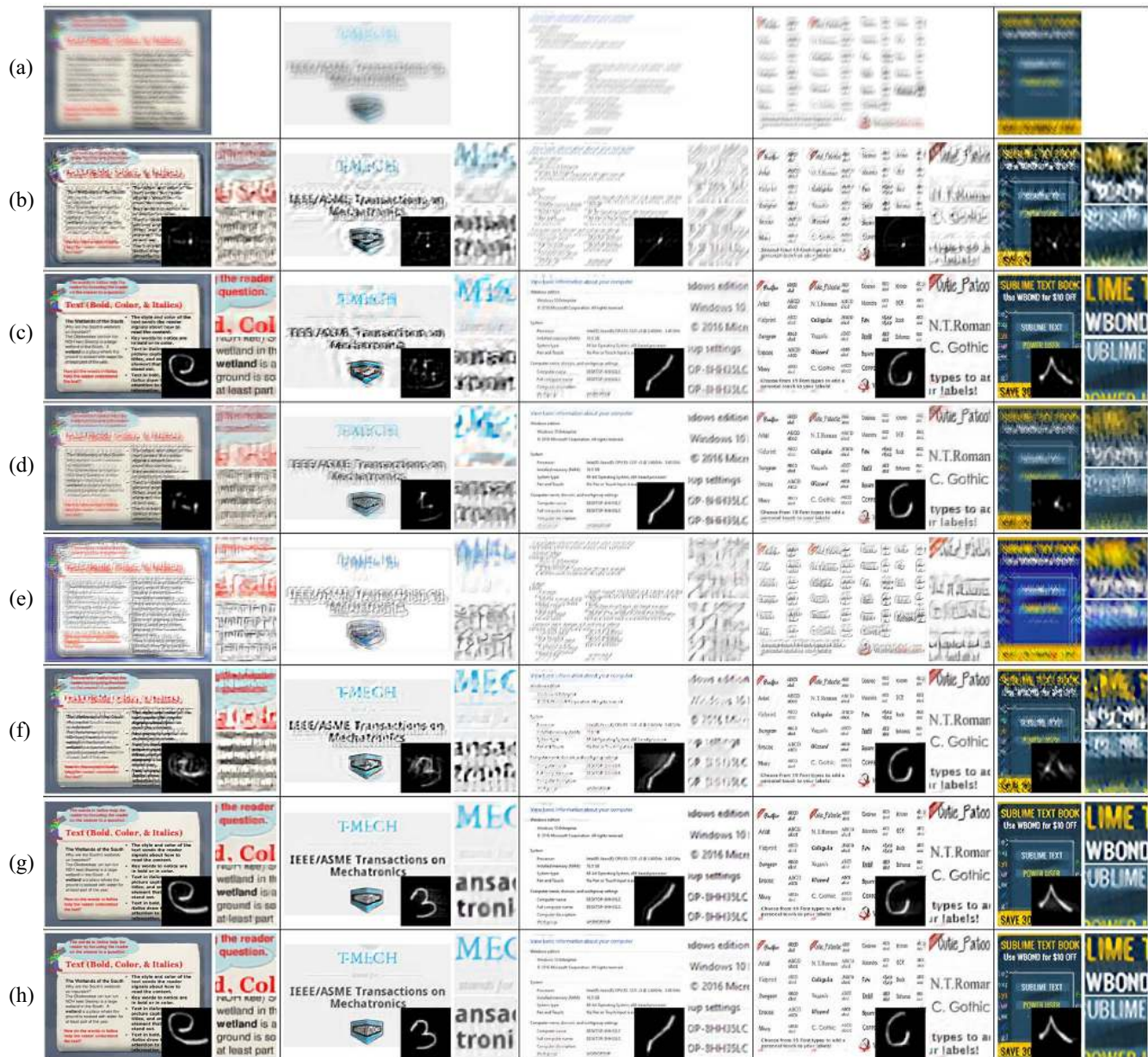


Fig. 6. Comparison of the deblurring results of ten sparsity-based methods. (a) Blurred image. (b) Krishnan *et al.* [15]. (c) Xu *et al.* [23]. (d) Zuo *et al.* [14]. (e) Hradiš *et al.* [11]. (f) Pan *et al.* [8]. Our method (g) without denoising and (h) with denoising. The estimated kernel is displayed at the bottom of each deblurring result with two scaled clips on its right. Note that there is no kernel estimated by Hradiš *et al.* [11].

kernel prior constraints; Xu *et al.* [23] used a L_0 norm approximation for the image constraint; and Zuo *et al.* [14] imposed the hyper-Laplacian constraint for both image and kernel. The deblurring results of these three methods and other five text image oriented methods including ours are shown in Fig. 6.

Highly relying on the types of the training blurred images, Hradiš *et al.* [11] obtains the worst performance among the ten even though it aims for text images. There are apparent ringing artifacts and blurring among all existing sparsity-based methods, no matter whether they are natural image or text image. Especially, the existing text image oriented methods are apparently not better than the natural image oriented ones, e.g., Pan *et al.* [8] versus Xu *et al.* [23].

However, the proposed L_0 -norm-based deblurring method (the last two rows in Fig. 6) achieves better performances than the other compared methods. Our method performs the best with denoising.

A clearer kernel always associates with a higher quality latent estimation, e.g., Pan *et al.* [8] versus Cao *et al.* [9] and the comparison between ours. It justifies the importance of our simultaneous constraints for the image, its gradient and the kernel, as well as the importance of kernel denoising. We now give the comparison with the text image deblurring methods, Hradiš *et al.* [11], Pan *et al.* [8], and Cao *et al.* [9]. Fig. 7 shows the deblurring results of two blurred images taken from the dataset of Pan *et al.* [8]. Another experiment with the significantly blurred real text images is also taken into

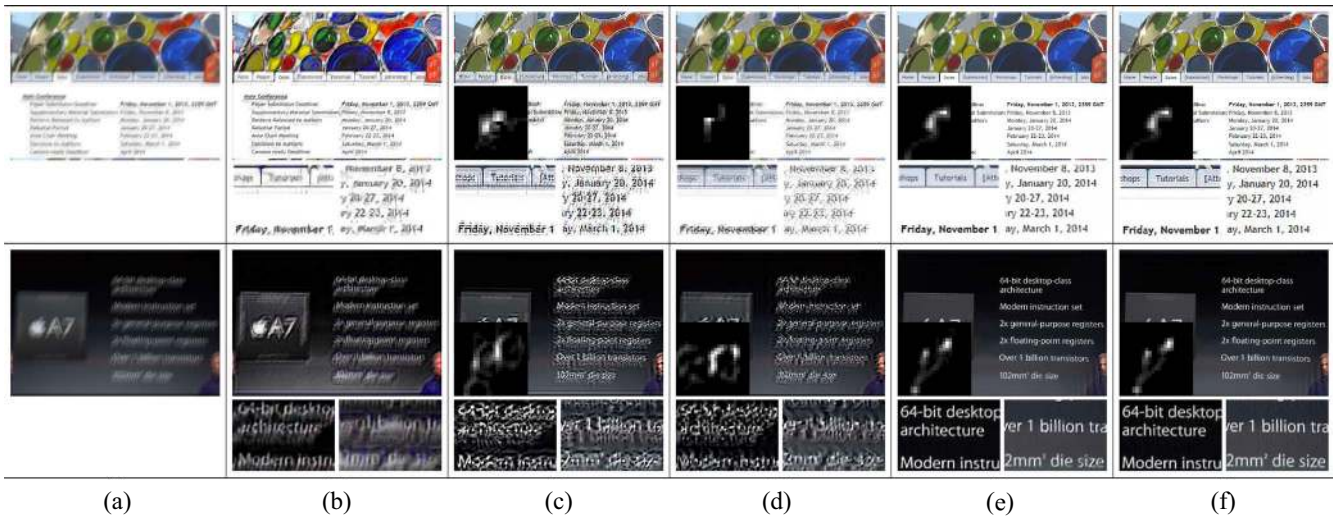


Fig. 7. Deblurring results with the dataset in [8]. The estimated kernel is displayed in the similar way as Fig. 6 with two scaled clips shown below. (a) Blurred image. (b) Hradiš *et al.* [11]. (c) Pan *et al.* [8]. (d) Cao *et al.* [9]. Our method (e) without denoising and (f) with denoising.

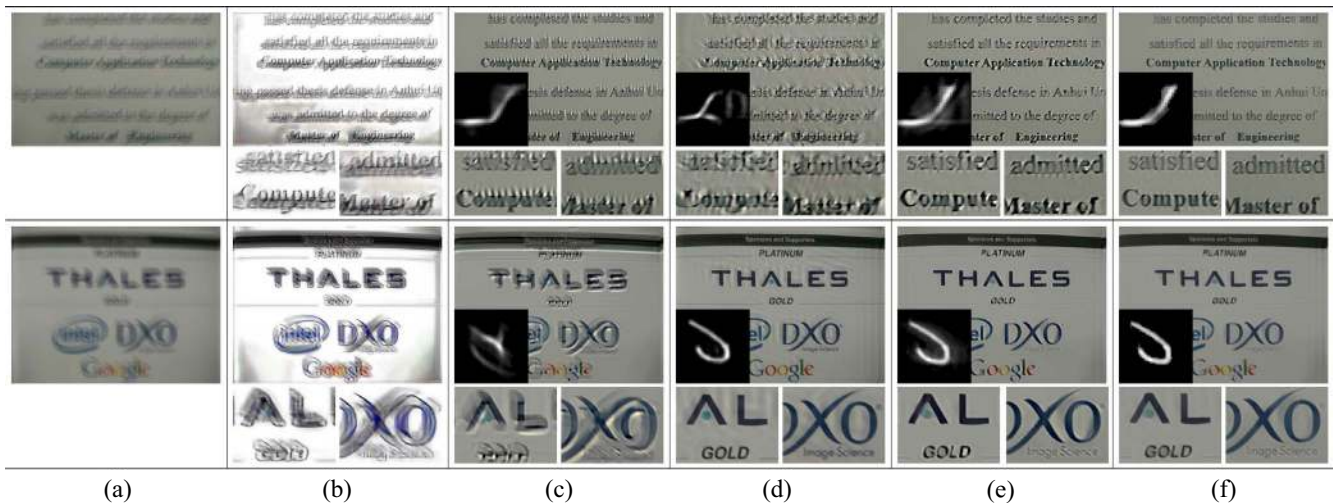


Fig. 8. Deblurring results on real images. The results are shown in the same way as Fig. 7. (a) Blurred image. (b) Hradiš *et al.* [11]. (c) Pan *et al.* [8]. (d) Cao *et al.* [9]. Our method (e) without denoising and (f) with denoising.

consideration (Fig. 8). Hradiš *et al.* [11] still returns the worst results among all methods while while neither Pan *et al.* [8] nor Cao *et al.* [9] can effectively remove the blur. Our method obtains the best results, with much clearer denoised kernels.

B. Quantitative Results

In the following, we will present the statistical comparison of the convergence speed and accuracy of the kernel estimation. We have discussed that the L_2 -norm-based kernel constraint is not necessarily useful (Section IV-B). More theoretical analysis of the performance of L_0 norm or L_1 norm is a nontrivial task. Therefore, the convergence performance of the kernel estimation between the proposed L_0 kernel prior-based method and the L_2 kernel prior-based methods of both Pan *et al.* [8] and Cao *et al.* [9] are compared. For fair comparison, the kernel denoising step in our method is omitted. The cross-correlation-based kernel similarity metric used by

Pan *et al.* [8] is adopted, which is computed by the maximum response of normalized cross-correlation between two kernels.

Fig. 9(a) gives the example convergence processes of both Pan *et al.* [8] and ours during the iterations with the eighth kernel. The process of Cao *et al.* [9] is not shown because their code cannot be debugged iteration after iteration, therefore, the intermediate kernels are not available for comparison. Both Pan *et al.* [8] and ours take a multiscale way where each scale is made of five iterations. kernels are converged in every sequence of five iterations and then restart at a coarser status with a lower similarity. lower similarity. However, the average similarity of Pan *et al.* [8] oscillates approximately between 0.7 and 0.75 while ours always increases at each scale and reaches a similarity of more than 0.85 after 25 iterations. Therefore, the convergence speed of our method can be faster and more accurate than that of Pan *et al.* [8]. Fig. 9(b) shows the final averaged similarities for each kernel, where our method is uniformly and significantly better than the other two methods.

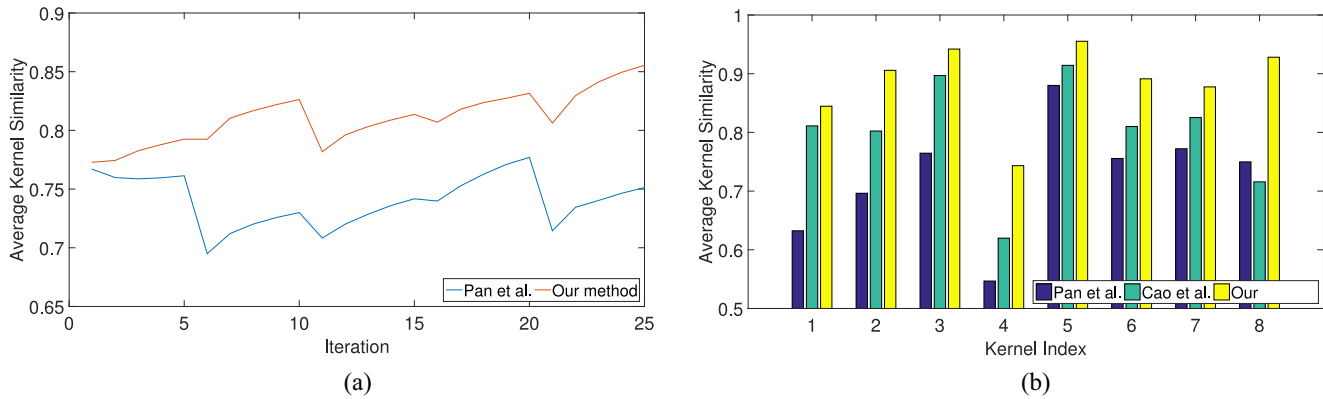


Fig. 9. Kernel estimation performance with the dataset of Pan *et al.* [8]. (a) Example convergence for the eighth kernel during the iteration. (b) Average convergence for each kernel. Note that Cao *et al.* [9] is not included in (a) as explained in the texts.

IX. CONCLUSION

We introduced a new motion deblurring method for text images. While previous methods omit the sparsity of the motion blur kernel, the proposed method incorporates the kernel sparsity as well as the sparsities of both image and its gradients. Formulated all the sparsities by the L_0 norm, this new model is optimized by the half-quadratic splitting coupling with fast gradient descent. A structure-preserving denoising method aiming at keeping the kernel structure unchanged was also presented. It can yield a noise-free kernel with a curvilinear trajectory while keeping the kernel intensities and shape unchanged. Experimental results demonstrated the advantages of the method both qualitatively and quantitatively.

REFERENCES

- [1] X. Lu, Y. Yuan, and X. Zheng, "Joint dictionary learning for multispectral change detection," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.
- [2] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1451–1462, Apr. 2014.
- [3] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang, "A comparative study for single image blind deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 1701–1709.
- [4] J. Shen, X. Yang, X. Li, and Y. Jia, "Intrinsic image decomposition using optimization and user scribbles," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 425–436, Apr. 2013.
- [5] Z. Hu, J.-B. Huang, and M.-H. Yang, "Single image deblurring with adaptive dictionary learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2010, pp. 1169–1172.
- [6] X. Lu, Y. Yuan, and P. Yan, "Alternatively constrained dictionary learning for image superresolution," *IEEE Trans. Cybern.*, vol. 44, no. 3, pp. 366–377, Mar. 2014.
- [7] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 1628–1636.
- [8] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring text images via L_0 -regularized intensity and gradient prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 2901–2908.
- [9] X. Cao, W. Ren, W. Zuo, X. Guo, and H. Foroosh, "Scene text deblurring using text-specific multiscale dictionaries," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1302–1314, Apr. 2015.
- [10] H. Cho, J. Wang, and S. Lee, "Text image deblurring using text-specific properties," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 7576, 2012, pp. 524–537.
- [11] M. Hradiš, J. Kotera, P. Zemčík, and F. Šroubek, "Convolutional neural networks for direct text deblurring," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2015, pp. 6.1–6.13.
- [12] J. Pan, Z. Lin, Z. Su, and M.-H. Yang, "Robust kernel estimation with outliers handling for image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2800–2808.
- [13] W.-Z. Shao, H.-B. Li, and M. Elad, "Bi-L0-L2-norm regularization for blind motion deblurring," *J. Vis. Commun. Image Represent.*, vol. 33, pp. 42–59, Nov. 2015.
- [14] W. Zuo, D. Ren, D. Zhang, S. Gu, and L. Zhang, "Learning iteration-wise generalized shrinkage-thresholding operators for blind deconvolution," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1751–1764, Apr. 2016.
- [15] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 233–240.
- [16] 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.
- [17] J. Pang and G. Cheung, "Graph Laplacian regularization for image denoising: Analysis in the continuous domain," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1770–1785, Apr. 2017.
- [18] E. Luo, S. H. Chan, and T. Q. Nguyen, "Adaptive image denoising by mixture adaptation," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4489–4503, Oct. 2016.
- [19] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2007, pp. 257–265.
- [20] A. Rajagopalan and R. Chellappa, Eds., "Motion deblurring: Algorithms and systems," in *Image Processing and Machine Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
- [21] J. Shen *et al.*, "Real-time superpixel segmentation by DBSCAN clustering algorithm," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5933–5942, Dec. 2016.
- [22] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 1964–1971.
- [23] L. Xu, S. Zheng, and J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2013, pp. 1107–1114.
- [24] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 562–572, Feb. 2012.
- [25] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, "Image deblurring via extreme channels prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 6978–6986.
- [26] X. Chen, X. He, J. Yang, and Q. Wu, "An effective document image deblurring algorithm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 369–376.
- [27] W. Wang, J. Shen, L. Shao, and F. Porikli, "Correspondence driven saliency transfer," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5025–5034, Nov. 2016.
- [28] T.-H. Li and K.-S. Lii, "A joint estimation approach for two-tone image deblurring by blind deconvolution," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 847–858, Aug. 2002.
- [29] J. Shen, D. Yu, L. Deng, and X. Dong, "Fast online tracking with detection refinement," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 162–173, Jan. 2018.

- [30] K. Donaldson and G. K. Myers, "Bayesian super-resolution of text in video with a text-specific bimodal prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 1188–1195.
- [31] Y. Lou, A. L. Bertozzi, and S. Soatto, "Direct sparse deblurring," *J. Math. Imag. Vis.*, vol. 39, no. 1, pp. 1–12, 2011.
- [32] L. Fang, H. Liu, F. Wu, X. Sun, and H. Li, "Separable kernel for image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 2885–2892.
- [33] X. Dong *et al.*, "Occlusion-aware real-time object tracking," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 763–771, Apr. 2017.
- [34] A. Levin, B. Nadler, F. Durand, and W. T. Freeman, "Patch complexity, finite pixel correlations and optimal denoising," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 73–86.
- [35] D. Gong, M. Tan, Y. Zhang, A. Van den Hengel, and Q. Shi, "Blind image deconvolution by automatic gradient activation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 1827–1836.
- [36] W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 38–49, Jan. 2018.
- [37] L. Xiao, J. Wang, W. Heidrich, and M. Hirsch, "Learning high-order filters for efficient blind deconvolution of document photographs," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 734–749.
- [38] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, p. 174, 2011.
- [39] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2368–2378, May 2018.
- [40] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 2657–2664.
- [41] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, p. 145, 2009.
- [42] P. J. Yim, P. L. Choyke, and R. M. Summers, "Gray-scale skeletonization of small vessels in magnetic resonance angiography," *IEEE Trans. Med. Imag.*, vol. 19, no. 6, pp. 568–576, Jun. 2000.
- [43] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2009, pp. 1033–1041.



Xianyong Fang (M'16) is currently a Full Professor with the School of Computer Science and Technology and also the Director of the Institute of Media Computing, Anhui University, Hefei, China. He was a Post-Doctoral Researcher with the Centre National de la Recherche Scientifique, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingenieur, Orsay, France. His current research interests include computer vision and computer graphics.

Qiang Zhou is currently pursuing the master's degree with the School of Computer Science and Technology, Anhui University, Hefei, China.

His current research interest includes image and video deblurring.

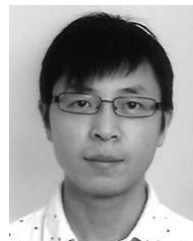


Jianbing Shen (M'11–SM'12) is a Full Professor with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. He has published about 100 journal and conference papers, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE CVPR, and the IEEE ICCV. His current research interests include computer vision and deep learning.

Mr. Shen was a recipient of several flagship honors, including the Fok Ying Tung Education Foundation from the Ministry of Education (MOE), the Program for Beijing Excellent Youth Talents from the Beijing Municipal Education Commission, and the Program for New Century Excellent Talents from MOE.

Christian Jacquemin received the Ph.D. degree from the Department of Computer Science, University of Paris 7, Paris, France, in 1991.

He is a Full Professor with the University of Paris-Sud, Université Paris-Saclay, Paris, and a Researcher of Computer Science with the Centre National de la Recherche Scientifique, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingenieur, Orsay, France. His current research interests include information visualization, interactive graphics and vision, and advanced graphics rendering.



Ling Shao (M'09–SM'10) is the Chief Scientist of the Inception Institute of Artificial Intelligence, Abu Dhabi, UAE, and also a Full Professor with the School of Computing Sciences, University of East Anglia, Norwich, U.K. His current research interests include computer vision, image processing, pattern recognition, and machine learning.

Mr. Shao is an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and other journals. He is a fellow of IAPR, IET, and BCS.