Understanding and Eliminating the Large-kernel Effect in Blind Deconvolution

Li Si-Yao¹, Qian Yin*¹, Ping Guo¹,²
¹Beijing Normal University, ²Beijing Institute of Technology
lisiyao@mail.bnu.edu.cn, yinqian@bnu.edu.cn, pguo@ieee.org

Abstract

Blind deconvolution consists of recovering a clear version of an observed blurry image without specific knowledge of the degradation kernel. The kernel size, however, is a required hyper-parameter that defines the range of the support domain. In this study, we experimentally and theoretically show how large kernel sizes yield inferior results by introducing noises to expected zeros in the kernel. We explain this effect by demonstrating that sizeable kernels lower the square cost in optimization. We also prove that this effect persists with a probability of one for noisy images. To eliminate this effect, we propose a low-rank based penalty that reflects structural information of the kernel. Compared to the generic $\ell_1$, our penalty can respond to even a small amount of random noise in the kernel. Our regularization process reduces the noise and efficiently enhances the success rate of large kernel sizes. We also compare our method to state-of-art approaches and test it using real-world images.

1. Introduction

Blind deconvolution algorithms attempt to recover the latent clear version of an observed blurry image without previously specified knowledge of the degradation kernel. The model of blurred images is

$$y = x \ast k + n \quad (1)$$

where $y$ is the observed blurry image, $x$ is the latent clear image, $k$ is the kernel (a.k.a. the Point Spread Function (PSF)), and $n$ represents random Gaussian noise. Maximize a posteriori (MAP) approaches redefine this ill-posed problem as regularized least squares optimization [2] as follows:

$$\min_{x, k} \left( \|x \ast k - y\|^2 + \lambda g(x) + \sigma h(k) \right). \quad (2)$$

Because the least squares part contains two variables, this optimization is generally realized in an alternate way:

$$x^{(i+1)} = \min_x \left( \|x \ast k^{(i)} - y\|^2 + \lambda g(x) \right) \quad (3)$$

$$k^{(i+1)} = \min_k \left( \|x^{(i+1)} \ast k - y\|^2 + \sigma h(k) \right). \quad (4)$$

Prior function $g(x)$ reflects how likely it is that $x$ will be a clear image. A series of works have attempted to define $g$ by fitting the shape of derivative histograms of clear images, including the mixture of Gaussian functions [6], the hyper-Laplacian functions [11, 14] and the combination of Laplacian and Gaussian functions [20]. These $\ell_\alpha$-like regularizations are prone to failure because they lower the cost associated with blurry images [15]. The blind deconvolution task has been redefined to address the sparse structure of sharp images rather than to fit the shapes of sparsity distribution. Krishnan et al. [12] applied an unnatural $\ell_1/\ell_2$ sparsity measurement, which favors axes in high dimensional spaces, that reflects flat areas and salient edges in clear images; Xu et al. [23] proposed a generalized $\ell_0$ regularization.

Although $k$ is unknown, most implementations use the kernel size as a hyper-parameter that limits the largest potential support domain. Previous studies have shown that improper kernel sizes yield failures [6, 16]. If the input size is smaller than ground truth, errors in the estimated kernel are inevitable due to the lack of expansion capacity. In this study, we experimentally and theoretically show that in the absence of regularization, kernel sizes greater than ground truth introduce increased noises into estimated kernels and result in higher errors.

The current literature lacks details on the constraints of $k$, and $k$ has been considered sparsity without being precisely defined. In some implementations, the generic sparse regularizations $\ell_2$ [23, 7, 3, 24], $\ell_1$ [12, 20, 17] and $\ell_\alpha$ ($0 < \alpha < 1$) [25] are used. In addition, some works implemented refinements after the k-step. Cho and Lee [3] removed any elements smaller than 1/20 of the maximum of $k$. Xu and Jia [23] proposed an iterative support domain detector based on the differences of elements of $k$. Only two characteristics of $k$ are recognized: (i) the sum of $k$ equals 1 and (ii) the elements of $k$ are no smaller than 0. With respect to (i), Kundur et al. [13] indicates that the mean of a clear image is preserved in the blurry. Levin et al. [15] emphasizes that constraints on the sum prevent scaling conflicts. For (ii), the PSF is naturally greater than zero because the...
Figure 1. Large kernels produce inferior results. (a) The blurry image. (b) The numerical errors on kernel size; (c-f) Deblurred results: in the first row are restored images; in the second row are corresponding estimated kernels; in the third row are the support domains ($k > 0$); in the last row are the domains of $k > \max(k)/30$. Adjacent positive pixels are colored identically and zeros are white.

contribution of light is positive. These constraints are realized as projections after the $k$-step [12, 25]. Using 1D simulations, Perrone et al. [19] showed that projections after alternate Gradient Descent prevent the trivial (delta) solution. In this study, we revisit the question of what constitutes a good kernel and propose a novel regularization method to eliminate the large-kernel effect.

2. Large-kernel effect

If an input kernel size is greater than ground truth, the estimated kernel is expected to be equivalent to a zero-padded smaller kernel. However, impurities emerge in areas that should be zero. This effect yields image deblurring failures. In this section, we describe and analyze the failures in detail.

2.1. Phenomena

We tested our implementation described in Section 4 (but set $\mu$ and $\sigma$ zero) using one-half to three times the size of the truth kernel iteratively. Figure 1(d-f) shows that the larger the kernel size than ground truth the poorer the deblurred image. At the same time, kernels become more noisy with higher numerical errors. Figure 1(b) shows the relationship curves of the error ratio of restored $x$ and the Sum of Squared Difference (SSD) of estimated $k$ on kernel size, respectively. Although the parameters that performed well on truth size were kept identical for the larger sizes during the experiment, the error increases continuously with kernel size.

The errors in kernels were considered noises [20]. The noises, however, are not random but structural. To illustrate this point, we clustered the neighboring positive elements of the estimated kernels by using different colors. Figure 1(c-f) shows the support domains of large kernels separated into patches, with larger sizes yielding more clusters. Gaps (zeros) between clusters is caused by the non-negativity projection. Random noise cannot yield such large, connected and regular pieces of positivity and negativity.
It should be noted here that in experiments of Figure 1, we ignored regularization $h$, then the k-step corresponds to a least squares problem

$$k^{(i+1)} = \min_k \|x^{(i+1)} * k - y\|^2. \tag{5}$$

We also avoid using the popular multi-scaled scheme in these experiments.

2.2. Analysis

We attribute the errors in Figure 1 to the Plumping Theorem; that is, the padding of linear independent columns to a thin matrix (with more rows than columns) lowers the least squares cost.

**Theorem 1. (Plumping Theorem)** Let $A = [v_1 \ldots v_n]$, where $v_i \in \mathbb{R}^m$ ($m \geq n + 1$). Let $B = [w_1 A w_2]$ where $w_1, w_2 \in \mathbb{R}^m$ and rank($B$) $> \text{rank}(A)$. Given an $m$-$D$ random vector $b$ whose elements are i.i.d. and the continuous probability density function $p$, for $u \in \mathbb{R}^m$

$$\Pr \{ \inf \{ \|Bu - b\|^2 \} < \inf \{ \|Au - b\|^2 \} \} = 1.$$

**Proof.**

$$\Pr \{ \inf \{ \|Bu - b\|^2 \} \geq \inf \{ \|Au - b\|^2 \} \} = \Pr \{ b \in \mathbb{R}^m \setminus \text{span}(w_1, w_2) \} = \int_{\Omega} dp(b)$$

where $\Omega = \mathbb{R}^m \setminus \text{span}(w_1, w_2)$. For $w_1, w_2$ are linear independent to $\{v_1, \ldots, v_n\}$, we have $w_1 \neq 0$ or $w_2 \neq 0$. Hence, the Lebesgue measure of $\Omega$ is zero, and the probability is zero.

The Plumping Theorem also suggests different optimal solutions to least squares for the plumper matrix and the original matrix.

In (1), $x, y \in \mathbb{R}^{M \times N}$, $k \in \mathbb{R}^{L \times K}$. The degradation (convolution) component can be rewritten into linear transforms. Note that $f(x, k) = x * k$; thus,

$$f(x, k) = T_k x = T_x k \tag{6}$$

where $T_k$ and $T_x$ are block banded Toeplitz matrices $[1, 8]$; $T_k \in \mathbb{R}^{MN \times MN}$ and $T_x \in \mathbb{R}^{MN \times LK}$; $T$ represents the column-wise expanded vector of matrix $T$.

In 1D cases, where $N = K = 1$, assuming $L = 2l + 1$, $T_x(L)$

$$
\begin{bmatrix}
x_{l+1} & \cdots & x_2 & x_1 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & x_2 & x_1 & \ddots & \vdots \\
x_{M-1} & \cdots & \vdots & \vdots & x_2 & \ddots & 0 \\
0 & \cdots & x_{M-1} & \vdots & \vdots & \ddots & x_1 \\
\vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & x_M & x_{M-1} & \cdots & x_{M-l}
\end{bmatrix}
= 
\begin{bmatrix}
J^{-l}x & \ldots & J^{-1}x & x & J^{(1)}x & \ldots & J^{(l)}x
\end{bmatrix}.
\tag{7}
$$

During blind deconvolution iterations, for identical values of $x^{(i)}$, a large $L$ introduces a greater number of columns onto both sides of $T_x(i)$ and results in different solutions. To illustrate this point, we tested a 1D version of blind deconvolution without kernel regularization and took different values of $L$ (true and double and four times the truth size) for the 50th k-step optimization after 49 truth-size iterations (see Figure 2). Figure 2(a-c) show that the optimal solutions in different sizes differ slightly on the main body that lies within the ground truth size (colored in red) but greatly outside this range (colored in green), where zeros are expected. Figure 2(d-f) compare ground truth to the estimated kernels in (a-c) after nonnegativity and sum-to-one projections. Larger sizes yield a greater number of positive noises; hence, they lower the weight of the main body after the projection and change the outlook.

Even if $x^{(i)}$ successfully iterates to truth $x$, Theorem 1 implicates the large-kernel effect will remain for random Gaussian noise $n$. Kaltofen and Lobo [10] proved that for an $M$-$by$-$M$ Toeplitz matrix composed of finite filed of $q$ elements,

$$\Pr \{ \text{rank}(T_{M \times M}) = M \} = 1 - 1/q. \tag{8}$$

We assert that if $L \leq M$ and $x$, the elements of which are i.i.d., as

$$p(x_j) = \begin{cases} 
\beta \exp(-\gamma |x_j|^\alpha) & x_j \in [-1, 1] \\
0 & \text{otherwise}
\end{cases} \tag{9}$$

then

$$\Pr \{ \text{rank} \left( T_x(L + 2) \right) > \text{rank} \left( T_x(L) \right) \} = 1. \tag{10}$$

We show a stronger assertion than (10).

**Theorem 2.**

$$\Pr \{ \text{rank}(T_x(M) = M) = 1. \tag{11}$$

**Proof.** See [21].
Figure 2. The estimated kernels in 1D Blind deconvolution simulations. (a-c) are optimized kernels of different sizes after the 50th iteration. (d-f) are corresponding normalized kernels to left after nonegativity and sum-to-one projections. (g) shows the residue squares cost of (5) of the optimized kernel on its size. In this experiment, $x$ is a 255×1 vector extracted from a real image and the truth $k$ is generated by column-wisely adding a 23×23 truth kernel from Levin’s dataset [15]. The signal prior is $\ell_1\ell_2$. This figure is recommended to view in color.

3. Regularization

Although previous works consider kernels to be sparsity [23, 17], they lack elaborate descriptions. For clear images, the same concept was clarified as sharp edges and flat planes. In this section, we indicate the concrete meaning of sparsity on $k$. Then, we propose a low-rank based regularization. Finally, we compare this regularization to previous $\ell_\alpha$ on response to noise.

3.1. Ideal kernel

Image degradation is generally caused by motion or inaccurate camera focusing [9, 18]. Assuming the exposure of cameras is uninterrupted, an ideal PSF is continuous in a single support domain. The origin of image degradation also suggests that most elements of large-size $k$ are zero, which are distributed surrounding the support domain. The kernels composed of noises (with many pieces of domains) in Figure 1(d-f) should be recognized as impure.

We can see in Figure 1(d-f) that the main body of $k$ can be determined even without any kernel regularization, but noise in large kernels yields failures. Thus the regularization on $k$ is required to eliminate the noise.

As motions are diverse, it is impossible to measure them in a fixed pattern, e.g. by limiting elements greater than the threshold of a fixed number. It is also unreliable to use statistical $\ell_\alpha$. An extreme example consists of disrupting a truth kernel and randomly reorganizing its elements, with the $\ell_\alpha$ cost unchanged. In contrast, we use the structural information of whole matrix $k$ to reduce noise.

3.2. Low-rank regularization

In the literature, low-rank regularization is widely used. In Compressive Sensing Recovery problems, low rank is applied to preserve the similarity of neighbouring images [4]. Tai et al. [22] used low-rank to remove the redundancy in convolution kernels of CNNs.

Because direct rank optimization is an NP-hard problem, continuous proximal functions are required. Fazel et al. [5] propose

$$\log \det \left((X + \delta I)\right)$$

(11)

as a heuristic proxy for $X \in S^N_+$ where $I$ is the N-by-N
identity matrix and \( \delta \) is a small positive number. A 1D case is shown in Figure 3.

To allow this approximation to play a role in general matrices, the low-rank object is substituted to \( (XX^T)^{1/2} \) \cite{4}. The regularization function then becomes

\[
h(X) = \log \det((XX^T)^{1/2} + \delta I) = \sum_j \log(s_j + \delta) \quad (12)
\]

where \( s_j \) is the i-th singular value of \( X \).

The intention to take low-rank regularization on \( k \) is based on a common phenomenon of noisy matrices. Figure 4(a-b) shows a nonnegativity projected Gaussian noise matrix and its singular values in decreasing order. For a noise matrix, where light and dark alternate irregularly, the distribution of singular values decays sharply at lower indices, then it breaks and drag a relatively long and flat tail to the last. This phenomenon is generic \cite{1}. Based on this fact, noise matrices are distinguished by high log det cost from real kernels (see Figure 4(c)). Figure 4(d) shows that the singular values of low-rank regularized kernel are distributed similarly as ground truth, compared with the impure one.

Compared to previously used \( \ell_p \) norms, low-rank regularization responds more efficiently to noise. To illustrate this point, we generated a noisy kernel by adding a small percentage (\( \varepsilon \)) of Gaussian noise and \( 1 - \varepsilon \) of the real kernel. Figure 5 shows that the low-rank cost rapidly adjust favorably to the noise but \( \ell_p \) norms fail.

![Figure 4](image1.png)

**Figure 4.** The characteristics of clean kernels and noisy matrices on singular values. (a) shows the support domain (black) of a 47 \times 47 random Gaussian noise matrix after nonnegativity projection. (b) is the distribution of singular values of (a). In (c) we compare the log det cost of random Gaussian noise matrix (black), the truth kernel from \cite{15} after zero-padding (red), and the Gaussian PSF with size/6 standard deviation (blue) on kernel size. In (d) we compare the distribution of scaled singular values (the maximum to 1) of clean, impure and regularized kernels. This figure is recommended to view in color.

![Figure 5](image2.png)

**Figure 5.** Comparison on respond to noise. The cost ratio is calculated as \( 1 + \frac{\text{cost}(\varepsilon)-\text{cost}(0)}{\text{cost}(0)} \). This figure is recommended to view in color.

### 4. Implementation

#### 4.1. Updating \( k \)

It should be firstly noted that the log det function is non-convex (it is actually concave on \( S_+ \)). To solve the low-rank regularized least squares problem

\[
\min_k \left( \|x^{(i+1)} * k - y\|^2 + \sigma h(k) \right) \quad (13)
\]

where \( h \) is defined in (12), we introduce the auxiliary variable \( \Psi = k \) and reformulate (13) into

\[
\min_{k, \Psi} \left( \|x^{(i+1)} * \Psi - y\|^2 + \sigma \log \det \left( (kk)^{1/2} + \delta I \right) \right)
\]

s.t. \( \Psi = k \) \quad (14)

Using the Lagrange method, (14) is solved by two alternate sub-optimizations

\[
\begin{align*}
\Psi^{(j+1)} &= \min_{\Psi} \left( \|x^{(i+1)} * \Psi - y\|^2 + \mu \|\Psi - k^{(j)}\|^2 \right) \\
k^{(j+1)} &= \min_k \left( \frac{1}{2\tau} \|k - \Psi^{(j+1)}\|^2 + \sigma h(k) \right)
\end{align*}
\]

where \( j \) is the iteration number while \( \mu \) and \( \tau \) are trade-off parameters.
The \( \Psi \)-substep is convex and accomplished using the Conjugate Gradient method. For \( k \)-substep, the requirement for low rank is limited; otherwise, the regularization may change the main body of kernel. An extreme result is a zero matrix. Thus, our strategy is to lower the rank at \( \Psi \) locally. Using the first-order Taylor expansion of \( h \) at fixed matrix \( Z \):

\[
h_Z(X) = h(Z) + \text{tr} \left( (ZZ^T)^{1/2} + \delta I \right)^{-1} \left( (XX^T)^{1/2} - (ZZ^T)^{1/2} \right)
\]

where \( \delta_i \) is the \( i \)-th eigenvalue of \( Z \), the \( k \)-substep is transformed into an iterative optimization

\[
k^{(t+1)} = \min_k \left( \frac{1}{2\tau} \| k - \Psi^{(j+1)} \|^2 + \sigma h_{k^{(t)}}(k) \right)
\]

where \( t \) is the inner iteration number. For convenience we set \( \sigma \) as a flag (if \( \sigma = 0 \), the \( k \)-substep will be skipped) and only tuned \( \tau \) as the trade-off parameter.

Define the proximal mapping of function \( \phi \) as follows:

\[
\text{prox}_{\phi}(v) = \min_u \left( \frac{1}{2}\| u - v \|^2 + \phi(u) \right).
\]

Dong et al. [4] proved that one solution to the proximal mapping of \( \tau h_Z \) is

\[
\text{prox}_{\tau h_Z}(X) = U (\Sigma - \tau \text{diag} (w))_+ V^T
\]

where \( U \Sigma V^T \) is SVD of \( X \), \( w_i = 1 / (\delta_i + \delta) \) and \( (A)_+ = \max \{ A, 0 \} \). Local low-rank optimization is iteratively implemented via the given parameter \( \tau \) (see Algorithm 1).

It should be noted here that \( \mu \) is designed to exponentially grow with \( j \) to allow more freedom of \( \Psi \) for early iterations.

### 4.2. Updating \( x \)

As \( \ell_1 \) norms prefer the blurry images, we take \( g = \ell_1/\ell_2 \) in this work, as proposed by Krishnan et al. [12]. The non-convex \( \ell_1/\ell_2 \) regularized least squares problem is solved by fixing \( \ell_2 \), as known on the previous iteration \( x \). Hence, the problem is transformed into a linear optimization problem

\[
\min_x \left( \| x * k - y \|^2 + \lambda \| x \|_1 \right)
\]

and is solved by iteratively minimizing a convex envelope (for small \( \eta \) small enough) equivalent to

\[
x^{(t+1)} = \text{prox}_{\eta \lambda \| x \|_1} (u^{(t)}) = \text{sgn}(u^{(t)})(|u^{(t)}| - \eta \lambda)_+
\]

where \( u^{(t)} = x^{(t)} - \eta \left[ \nabla_x \| x * k - y \|^2 \right]_{x=x^{(t)}} \).

### 5. Experiments and Results

In response to the high error ratios of large kernels mentioned in Figure 1 and to show the efficiency of our regularization, we used the same parameters except \( \mu \) and \( \sigma \).
in Figure 1. Figure 7 shows how the low-rank regularized kernels are much more robust to kernel size. Our regularization method performs robust on larger kernels upon dataset from [15]. Figure 8 shows the success rates of state-of-the-art methods versus our implementations with and without (set $\mu$ and $\sigma$ zero) low-rank regularization. Parameters are fixed during the whole experiment, where $\sigma = 1$, $\mu = 1$, $\tau = 5 \times 10^{-5}$, $\text{OuterIterMax} = 25$, $\text{CGIterMax} = 3$ and $\text{innerIterMax} = 10$. Low-rank regularization works effectively by diminishing noises (see Figure 9). In addition, we test several real-world images to reveal the robustness of our regularization on kernel sizes (see Figure 6). We avoid multi-scaling scheme except in the experiment of Figure 8.

Our implementation processed images of $255 \times 255$ pixels with truth blur sizes in 103 sec on average on a Lenovo ThinkCentre computer with Core i7 processor. Our MATLAB codes and deblurring results will be distributed after the present paper has been officially accepted for publication.

6. Discussion

In this paper, we demonstrate that kernel sizes larger than ground truth produce increased noises. We attribute the large-kernel effect to the Plumping Theorem. We also discuss the the concept of sparsity on $k$. The proposed low-rank regularization reflects the structural information of the kernel and reduces the large-kernel effect efficiently.
Figure 9. Deblur examples from Levin’s set. (a) are blurry images. In (b-e) we display restored images and estimated kernels (upside) and domains of \( k > \max(k)/30 \) (downside). (b): ground truth kernels; (c): truth sizes using low-rank regularization; (d): double size of truth without regularization; (e): double size of truth using low-rank regularization.

In practical implementations, even for noise-free \( y \), the intermediate \( x^{(i)} \) is unlikely to iterate to truth, hence some parts of \( y \) will be treated as implicit noises [20], which may intensify the effect even more than expected.

7. Acknowledgments

Li Si-Yao thanks Dongwei Ren for his generous suggestions and constructive comments. We thank Junfeng Li, Yongjian Hu and Wenyi Zeng for their suggestions and inspirations on the proof to Theorem 2.

This work is funded by National Natural Science Foundation of China (NSFC) (61472043, 61375045) and Joint Research Fund in Astronomy under cooperative agreement between the National Natural Science Foundation of China (NSFC) and Chinese Academy of Sciences (CAS) (U1531242). Qian Yin is the corresponding author.
References

[1] H. C. Andrews and B. R. Hunt. Digital image restauration. 1977. 3, 5

[2] T. F. Chan and C.-K. Wong. Total variation blind deconvolution. IEEE transactions on Image Processing, 7(3):370–375, 1998. 1

[3] S. Cho and S. Lee. Fast motion deblurring. In ACM Transactions on Graphics (TOG), volume 28, page 145. ACM, 2009. 1

[4] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang. Compressive sensing via nonlocal low-rank regularization. IEEE Transactions on Image Processing, 23(8):3618–3632, 2014. 4, 5, 6

[5] M. Fazel, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In American Control Conference, 2003. Proceedings of the 2003, volume 3, pages 2156–2162. IEEE, 2003. 4

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In ACM Transactions on Graphics (TOG), volume 25, pages 787–794. ACM, 2006. 1

[7] D. Gong, M. Tan, Y. Zhang, A. Van den Hengel, and Q. Shi. Blind image deconvolution by automatic gradient activation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1827–1836, 2016. 1

[8] R. M. Gray et al. Toeplitz and circulant matrices: A review. Foundations and Trends® in Communications and Information Theory, 2(3):155–239, 2006. 3

[9] A. K. Jain. Fundamentals of Digital Image Processing. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989. 4

[10] E. Kaltofen and A. Lobo. On rank properties of toeplitz matrices over finite fields. In Proceedings of the 1996 international symposium on Symbolic and algebraic computation, pages 241–249. ACM, 1996. 3

[11] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In Advances in Neural Information Processing Systems, pages 1033–1041, 2009. 1, 6

[12] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 233–240. IEEE, 2011. 1, 2, 6

[13] D. Kundur and D. Hatzinakos. Blind image deconvolution. IEEE signal processing magazine, 13(3):43–64, 1996. 1

[14] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. ACM transactions on graphics (TOG), 26(3):70, 2007. 1

[15] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 1964–1971. IEEE, 2009. 1, 4, 5, 7

[16] S. Liu, H. Wang, J. Wang, S. Cho, and C. Pan. Automatic blur-kernel-size estimation for motion deblurring. The visual computer, 31(8):733–746, 2015. 1

[17] J. Pan, Z. Lin, Z. Su, and M.-H. Yang. Robust kernel estimation with outliers handling for image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2800–2808, 2016. 1, 4

[18] A. P. Pentland. A new sense for depth of field. IEEE transactions on pattern analysis and machine intelligence, (4):523–531, 1987. 4

[19] D. Perrone and P. Favaro. Total variation blind deconvolution: The devil is in the details. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2909–2916, 2014. 2

[20] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In Acm transactions on graphics (tog), volume 27, page 73. ACM, 2008. 1, 2, 8

[21] L. Si-Yao, Q. Yin, and P. Guo. Supplementary material, 2017. https://igpr.bnu.edu.cn/~lisyao/pdf/large_kernel_supplement.pdf. 3

[22] C. Tai, T. Xiao, Y. Zhang, X. Wang, et al. Convolutional neural networks with low-rank regularization. arXiv preprint arXiv:1511.06067, 2015. 4

[23] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In European Conference on Computer Vision, pages 157–170. Springer, 2010. 1, 4

[24] L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1107–1114, 2013. 1

[25] W. Zuo, D. Ren, S. Gu, L. Lin, and L. Zhang. Discriminative learning of iteration-wise priors for blind deconvolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3232–3240, 2015. 1, 2