Supplementary Materials for "A variational EM framework with adaptive edge selection for blind motion deblurring"

Liuge Yang and Hui Ji
Department of Mathematics, National University of Singapore, Singapore, 119076
yang_liuge@u.nus.edu and matjh@nus.edu.sg

1. Overview

This supplementary material is organized as follows. In Section 2, we will show the proofs of Proposition 1 and 2. Then, in Section 3, we will present more details on the experiments. This includes quantitative comparison of the results in terms of SSIM, and more examples on the dataset [5], and on real images.

2. Proofs of Proposition 1 and 2

2.1. Proof of Proposition 1

Proof. Since

$$p(\nabla z|\nabla g;\theta^{(t)}) \propto p(\nabla g|\nabla z;\theta^{(t)})p(\nabla z;\theta^{(t)}) = \mathcal{N}(\nabla g|k^{(t)}\otimes\nabla z,\tilde{\sigma}^2\mathrm{I})\mathcal{N}(\nabla z|0,\Sigma^{(t)}),$$

$p(\nabla z|\nabla g;\theta^{(t)})$ is a normal distribution with mean given by

$$\mathrm{argmin}_{\nabla z}\|\nabla g - k^{(t)}\otimes\nabla z\|_2^2 + \tilde{\sigma}^2\|(\Sigma^{(t)})^{-\frac{1}{2}}\nabla z\|_2^2. \tag{1}$$

Since $q(\nabla z)$ is restricted to be normal distribution with a constant covariance matrix, and the KL-divergence between two normal distributions is

$$KL(\mathcal{N}(\mu_1,\Sigma_1)||\mathcal{N}(\mu_2,\Sigma_2)) = \frac{1}{2}(tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1) + \log\frac{\det(\Sigma_2)}{\det(\Sigma_1)} - n),$$

the KL function is minimized when the mean of $q(\nabla z)$ equals to that of $p(\nabla z|\nabla g;\theta^{(t)})$ given by (1). □

2.2. Proof of Proposition 2

Proof. Denote $\nabla z^{(t+1)}$ for ∇z for simplicity of notation. By ignoring the irrelevant terms in

$$\min_{\theta\in\Theta}\frac{1}{2\tilde{\sigma}^2}\big[||\nabla g - k\otimes\nabla z^{(t+1)}||_2^2 + \tilde{\sigma}^2||\Sigma^{-\frac{1}{2}}\nabla z^{(t+1)}||_2^2\big] + \sum_i\log\sigma_i + \frac{\lambda N}{2\tilde{\sigma}^2}||k||_2^2 + \frac{\lambda}{2}\sum_i\frac{1}{\sigma_i^2}, \tag{2}$$

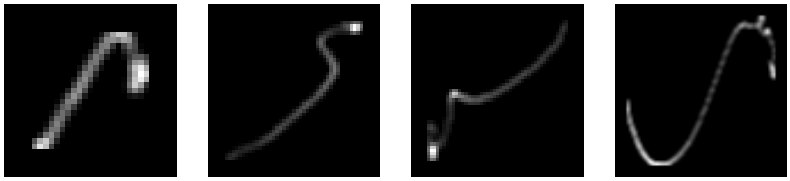
and let $\theta_Z^* = \{\sigma_1^*,...,\sigma_N^*\}$ denotes the optimal solution, we have

$$\theta_Z^* = \mathrm{argmin}_{\theta_Z\in\Theta}\sum_{i=1}^N(\log\sigma_i + \frac{|(\nabla z)_i|^2+\lambda}{2\sigma_i^2}).$$

Let $f_i(\sigma_i) = \log\sigma_i + \frac{a_i^2}{2\sigma_i^2}$, where $a_i = (|(\nabla z)_i|^2+\lambda)^{\frac{1}{2}}$. Let $\tilde{\sigma}_i^* = \mathrm{argmin}_{\sigma_i\geq\tau}f_i(\sigma_i)$, then by direct calculation, we have $\tilde{\sigma}_i^* = a_i$, if $a_i > \tau$ and τ otherwise. Since $\tilde{\sigma}_i^*$ is the unique optimal solution for each $f_i(\sigma_i)$, we have for each i, σ_i^* is either a_i or τ.

Let $\Lambda' = \{i : a_i > \tau, 1\leq i\leq N\}$. If $\#\{\Lambda'\}\leq M$, then $\sigma_i^* = \tilde{\sigma}_i^*$, which can be expressed as

$$\sigma_i^* = \begin{cases}(|(\nabla z^{(t+1)})_i|^2+\lambda)^{\frac{1}{2}} & \text{if } (|(\nabla z^{(t+1)})_i|^2+\lambda)^{\frac{1}{2}} > \tau \text{ and } i\in\Lambda,\\ \tau & \text{otherwise.}\end{cases} \tag{3}$$



(a) 51×51 (b) 75×75 (c) 75×75 (d) 101×101

Figure 1: Ground truth kernels and the corresponding sizes of the dataset [5].

	manmade	natural	people	saturated	text	average
Fergus-06 [2]	0.628	0.744	0.858	0.687	0.629	0.710
Cho-09 [1]	0.729	0.854	0.897	0.768	0.718	0.793
Xu-10 [13]	0.869	0.924	0.978	0.840	0.899	0.902
Krishnan-11 [3]	0.742	0.854	0.922	0.783	0.744	0.809
Levin-11 [6]	0.823	0.891	0.938	0.823	0.757	0.847
Sun-13 [11]	0.841	0.933	0.954	0.815	0.851	0.879
Xu-13 [14]	0.821	0.904	0.967	0.818	0.867	0.875
Zhang-13 [15]	0.761	0.885	0.961	0.808	0.775	0.838
Zhong-13 [16]	0.785	0.887	0.962	0.812	0.747	0.839
Michaeli-14 [7]	0.753	0.836	0.937	0.771	0.676	0.795
Pan-14 [9]	0.796	0.903	0.957	0.815	0.815	0.857
Perrone-14 [10]	0.820	0.917	0.957	0.794	0.815	0.860
DeepDeblur-17 [8]	0.659	0.793	0.902	0.769	0.618	0.748
DeblurGAN-17 [4]	0.608	0.728	0.852	0.727	0.609	0.705
Ours	**0.875**	**0.949**	**0.980**	**0.850**	**0.912**	**0.913**

Table 1: Quantitative comparison on the dataset in [5]. Performance is measured in average SSIM values on grayscale images. Different column denotes different category of images. The last column is the average SSIM value over the whole dataset.

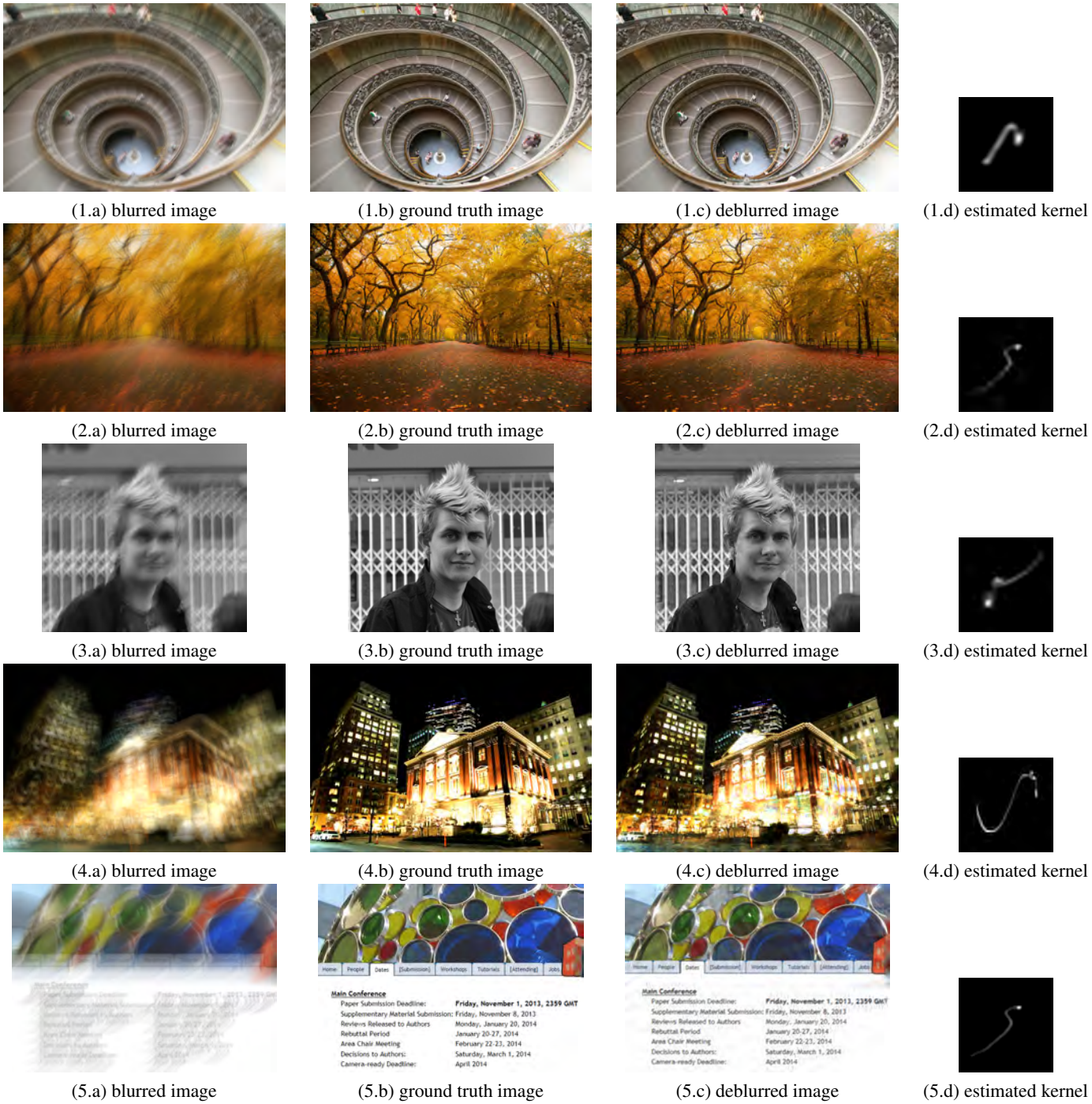
If $\#\{\Lambda'\} > M$, which breaks cardinality constraint, some of the $i \in \Lambda'$ has to be set to τ. For $a > \tau$, the cost of letting $\sigma^* = \tau$ instead of a is given by

$$h(a) = f(\tau) - f(a) = \frac{a^2}{2\tau^2} - \log a + \log \tau - \frac{1}{2},$$

where $f(\sigma) = \log \sigma + \frac{a^2}{2\sigma^2}$. Since $\forall a > \tau$, $h'(a) > 0$, the cost of letting $\sigma_i^* = \tau$ instead of a_i strictly increase as a_i increase. Since $a_i = (|(\nabla z)_i|^2 + \lambda)^{\frac{1}{2}}$, the optimal solution in this case will be achieved by letting $\sigma_i^* = \tilde{\sigma}_i^*$ when $i \in \Lambda$ and $\sigma_i^* = \tau$ otherwise, which can also be expressed by (3). The proof completes. □

3. Additional experiments and examples.

In the article, the quantitative comparison of different methods on the the synthetic dataset in Lai *et al*. [5] is listed in terms of the PSNR value. In this section, the quantitative comparison in terms of average SSIM is listed in Table 1 which used the same results as Table 1 in the article. See Figure 1 for the four ground truth kernels of different sizes used in [5] to generate the dataset, and see Figure 2 for visual inspection of the results on five images from the dataset [5] by the proposed method. These 5 images are taken from 5 categories respectively: "manmade", "natural", "people', "saturated" and "text". See Fig. 3 for visual comparison of different methods on more real images, including some real images summarized in [5].



(1.a) blurred image (1.b) ground truth image (1.c) deblurred image (1.d) estimated kernel

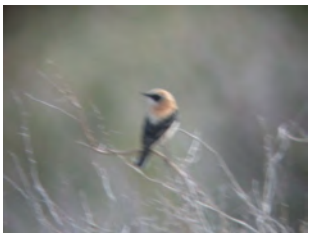
(2.a) blurred image (2.b) ground truth image (2.c) deblurred image (2.d) estimated kernel

(3.a) blurred image (3.b) ground truth image (3.c) deblurred image (3.d) estimated kernel

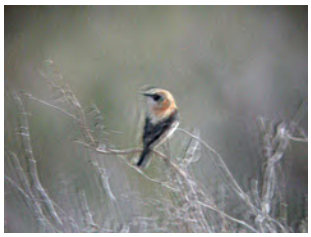
(4.a) blurred image (4.b) ground truth image (4.c) deblurred image (4.d) estimated kernel

(5.a) blurred image (5.b) ground truth image (5.c) deblurred image (5.d) estimated kernel

Figure 2: Demonstration of some results on the images from the dataset [5] by the proposed method. The first column shows blurry images; the second column shows ground truth images; the third column shows our deblurred results; the fourth column shows the kernels estimated by our algorithm. .

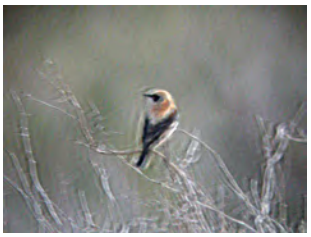


(1.a) input

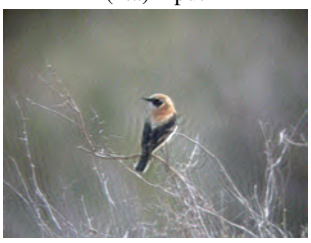


(1.b) Cho-09[1]

(1.c) Xu-10[13]

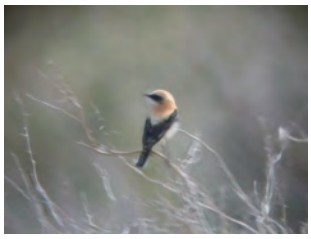


(1.d) Xu-13[14]

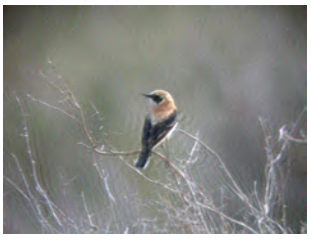


(1.e) Pan-14[9]

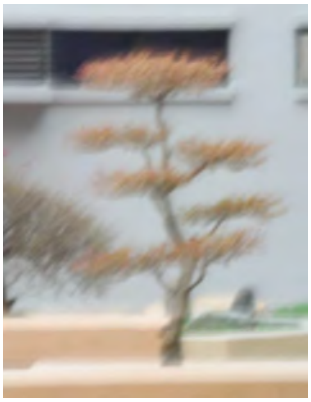
(1.f) Levin-11[6]



(1.g) DeepDeblur-17[8]

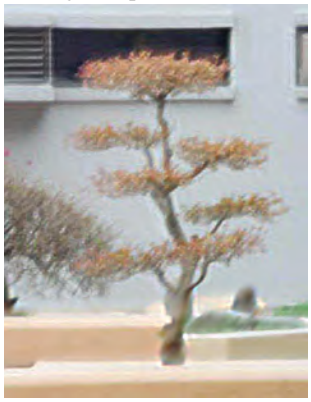


(1.h) Ours



(2.a) input

(2.b) Cho-09[1]



(2.c) Xu-10[13]

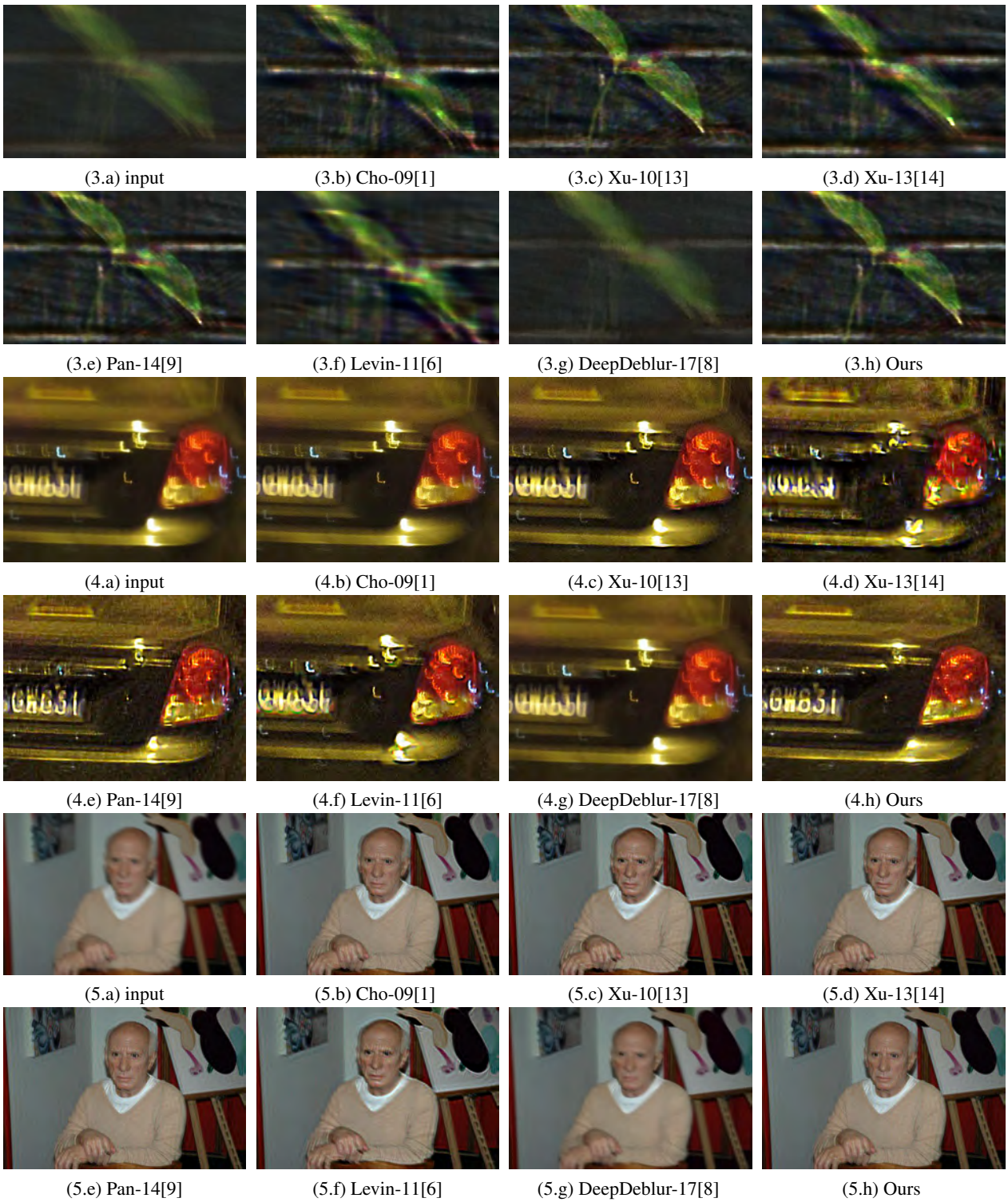
(2.d) Xu-13[14]

(2.e) Pan-14[9]

(2.f) Levin-11[6]

(2.g) DeepDeblur-17[8]

(2.h) Ours



(3.a) input (3.b) Cho-09[1] (3.c) Xu-10[13] (3.d) Xu-13[14]

(3.e) Pan-14[9] (3.f) Levin-11[6] (3.g) DeepDeblur-17[8] (3.h) Ours

(4.a) input (4.b) Cho-09[1] (4.c) Xu-10[13] (4.d) Xu-13[14]

(4.e) Pan-14[9] (4.f) Levin-11[6] (4.g) DeepDeblur-17[8] (4.h) Ours

(5.a) input (5.b) Cho-09[1] (5.c) Xu-10[13] (5.d) Xu-13[14]

(5.e) Pan-14[9] (5.f) Levin-11[6] (5.g) DeepDeblur-17[8] (5.h) Ours

Figure 3: Visual comparison of the results from different methods. They are better viewed using zoom-in.

References

[1] S. Cho and S. Lee. Fast motion deblurring. *ACM TOG (Proc. SIG- GRAPH Asia)*, 28(5):145:1–145:8, 2009.

[2] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *ACM TOG (Proc. SIGGRAPH)*, 25(3):787–794, 2006.

[3] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, 2011.

[4] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. *ArXiv e-prints*, 2017.

[5] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang. A comparative study for single image blind deblurring. In *CVPR*, 2016.

[6] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *CVPR*, 2011.

[7] T. Michaeli and M. Irani. Blind deblurring using internal patch recurrence. In *ECCV*, 2014.

[8] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, July 2017.

[9] J. Pan, Z. Hu, Z. Su, and M.-H. Yang. Deblurring text images via ℓ_0-regularized intensity and gradient prior. In *CVPR*, 2014.

[10] D. Perrone and P. Favaro. Total variation blind deconvolution: The devil is in the details. In *CVPR*, 2014.

[11] L. Sun, S. Cho, J. Wang, and J. Hays. Edge-based blur kernel estimation using patch priors. In *ICCP*, 2013.

[12] O. Whyte, J. Sivic, and A. Zisserman. Deblurring shaken and partially saturated images. *IJCV*, 110(2):185–201, 2014.

[13] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, 2010.

[14] L. Xu, S. Zheng, and J. Jia. Unnatural l_0 sparse representation for natural image deblurring. In *CVPR*, 2013.

[15] H. Zhang, D. Wipf, and Y. Zhang. Multi-image blind deblurring using a coupled adaptive sparse prior. In *CVPR*, 2013.

[16] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang. Handling noise in single image deblurring using directional filters. In *CVPR*, 2013.