Fast ℓ^0-Regularized Kernel Estimation for Robust Motion Deblurring

Jinshan Pan and Zhixun Su

***Abstract*—Blind image deblurring is a challenging problem in computer vision and image processing. In this paper, we propose a new ℓ^0-regularized approach to estimate a blur kernel from a single blurred image by regularizing the sparsity property of natural images. Furthermore, by introducing an adaptive structure map in the deblurring process, our method is able to restore useful salient edges for kernel estimation. Finally, we propose an efficient algorithm which can solve the proposed model efficiently. Extensive experiments compared with state-of-the-art blind deblurring methods demonstrate the effectiveness of the proposed method.**

***Index Terms*—ℓ^0-regularized method, blind image deblurring, image restoration, kernel estimation.**

I. Introduction

IMAGE deblurring is a widely existing problem in image formation process. Due to the imperfection of the imaging devices, it still remains an active research area in image processing communities [1]–[9]. The formation process of image blur is usually modeled as

$$B = I * k + \varepsilon, \tag{1}$$

where B, I, k, and ε represent the blurred image, latent image, blur kernel, and the additive noise, respectively. $*$ denotes the convolution operator. Because image deblurring is an ill-posed problem, most approaches introduce an image prior that favors natural images over degraded ones. By regularizing the problem in this fashion, a high quality result can be achieved.

A. Related Work

To make blind deblurring more tractable, a wide range of parametric image priors have been proposed. Early approaches usually adopted Gaussian prior smoothness penalties on the natural images or gradients. These priors cannot preserve the sharp edges of natural images. To overcome this limitation, Total Variation (TV) and its variations as popular choices of the regularization term have been proposed to solve deblurring problems [10], [11]. Since natural image gradients do not obey the Gaussian distribution, Fergus *et al.* [1] proposed a zero-mean Mixture of Gaussian to fit the distribution of natural image gradients. Shan *et al.* [2] used a certain parametric model to approximate the heavy-tailed natural image prior. Some other approaches [12], [13] proposed hyper-Laplacian prior to fit the distribution of natural image gradients. Roth and Black [14] developed the Fields of Experts model for learning the filters. Recently, sparse representation methods have also been employed in blind deblurring [15]–[17].

Summarizing above discussions, most priors imposed on image gradients can be regarded as some approximations of ℓ^0-metric which has a good natural interpretation of sparsity of image gradients, but minimizing ℓ^0-regularized model is an NP-hard problem.

B. Our Contributions

In this paper, we propose a new ℓ^0-regularized method for blur kernel estimation. By introducing this prior, our kernel estimation method is significantly different from previous works in the following aspects:

1) In our kernel estimation framework, we develop an efficient Alternating Direction Method (ADM) to solve the proposed model whose sub-problems have their own closed form solutions.
2) Due to adopting an adaptive structure map, the performance of our method is comparable or even better than salient edge selection methods [3], [5], [18]. In addition, our method is much simpler, because it does not take the additional salient edge selection to avoid the delta kernel solution.

To enable comparison and further testify to validity of our method, we will make our source code available online.

II. ℓ^0-Regularized Kernel Estimation

Most blind delurring methods can be attributed to recovering the sharp edges in kernel estimation. Therefore, instead of recovering latent images, our kernel estimation is performed on the high frequencies of images, i.e., given the blurred image B, we want to recover the edges ∇I. Our kernel estimation model is written as

$$\min_{x,k} \|x * k - y\|_2^2 + \gamma\|k\|_2^2 + \lambda\|x\|_0, \tag{2}$$

where x denotes $\nabla I = (\partial_x I, \partial_y I)^T$, y denotes $\nabla B = (\partial_x B, \partial_y B)^T$, and $\|\cdot\|_0$ is the ℓ^0-norm that counts the number of non-zero values of x.

Model (2) consists of 3 terms. The first term is the likelihood term, i.e., the restored data should be consistent with the observation with respect to the estimated degradation model. The second term is the Tikhonov regularization on kernel k which can stabilize the blur kernel estimation. The third term

Manuscript received January 21, 2013; revised April 27, 2013; accepted April 28, 2013. Date of publication May 07, 2013; date of current version June 25, 2013. This work was supported in part by the NSFC-Guangdong Joint Fund (Grant U0935004), the NSFC Fund (Grants 61173103 and 91230103), and by the National Science and Technology Major Project (Grant 2013ZX04005021). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kyoung Mu Lee.

The authors are with School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China (e-mail: jspan@mail.dlut.edu.cn; zxsu@dlut.edu.cn).

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

Digital Object Identifier 10.1109/LSP.2013.2261986

is ℓ^0-norm based regularization that can preserve the sparsity of natural image gradients.

A. Numerical Algorithm

Our kernel estimation method originates from traditional Maximum a-Posteriori (MAP) framework. It can be separated into the following two steps: (1) kernel estimation and (2) sharp edges restoration.

In the kernel estimation step, the blur kernel is obtained by

$$\min_k \|x * k - y\|_2^2 + \gamma \|k\|_2^2. \tag{3}$$

Obviously, model (3) is a least square problem whose solution can be fast obtained by

$$k = F^{-1}\left(\frac{\overline{F(\partial_x I)}F(\partial_x B) + \overline{F(\partial_y I)}F(\partial_y B)}{F(\partial_x I)^2 + F(\partial_y I)^2 + \gamma}\right), \tag{4}$$

where $F(\cdot)$ and $F^{-1}(\cdot)$ denote the Fast Fourier Transform (FFT) and inverse FFT, respectively, and $\overline{F(\cdot)}$ is the complex conjugate operator.

In the second step, i.e., sharp edges restoration, the problem can be reformulated as

$$\min_x \|x * k - y\|_2^2 + \lambda \|x\|_0. \tag{5}$$

Note that model (5) is a discrete optimization problem, which is difficult to solve by traditional gradient decent or other discrete optimization methods. Furthermore, the brute force search method is time-consuming. Inspired by the idea of ADM, we apply it to solve model (5).

We first convert model (5) to the following equivalent problem:

$$\begin{aligned} \min_{x,w} \quad & \lambda \|w\|_0, \\ \text{s.t.} \quad & x * k = y, \ w = x. \end{aligned} \tag{6}$$

Then, we have the following Lagrangian function:

$$\begin{aligned} L(x, w, J_1, J_2) = \lambda \|w\|_0 + J_1^T(x - w) + J_2^T(x * k - y) \\ + \frac{\mu}{2}(\|x * k - y\|_2^2 + \|x - w\|_2^2), \end{aligned} \tag{7}$$

where J_1 and J_2 are Lagrange multipliers and μ is a penalty parameter. Hence, the corresponding iterative scheme of ADM is

$$\begin{cases} w^{n+1} = \arg\min_w L(x^n, w, J_1^n, J_2^n), \\ x^{n+1} = \arg\min_x L(x, w^{n+1}, J_1^n, J_2^n), \\ J_1^{n+1} = J_1^n + \mu(x^{n+1} - w^{n+1}), \\ J_2^{n+1} = J_2^n + \mu(x^{n+1} * k - y). \end{cases} \tag{8}$$

It is easy to show that the minimization of $L(x, w^{n+1}, J_1^n, J_2^n)$ is equivalent to the following problem:

$$\min_x \|x * k + \frac{1}{\mu}J_2^n - y\|_2^2 + \|x - w^{n+1} + \frac{1}{\mu}J_1^n\|_2^2. \tag{9}$$

Obviously, model (9) is a least square problem and the corresponding solution can be obtained by using FFT, i.e.,

$$x = F^{-1}\left(\frac{\overline{F(k)}F(y - \frac{1}{\mu}J_2^n) + F(w^{n+1} - \frac{1}{\mu}J_1^n)}{\overline{F(k)}F(k) + 1}\right). \tag{10}$$

Similar to the x sub-problem, the minimization of $L(x^n, w, J_1^n, J_2^n)$ can be equivalently written as

$$\min_w \frac{\mu}{2}(\|x^n - w + \frac{1}{\mu}J_1^n\|_2^2) + \lambda \|w\|_0. \tag{11}$$

Although model (11)[1] involves a discrete ℓ^0-metric, we can still obtain its explicit solution by the following lemma:

Lemma II.1: Let z be a single variable, if the optimal solution of

$$\min_z (z - q)^2 + \beta \|z\|_0 \tag{12}$$

is z^*, then z^* is defined as

$$z^* = \begin{cases} q, & |q|^2 \geq \beta, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

where

$$\|z\|_0 = \begin{cases} 1, & |z| \neq 0, \\ 0, & \text{otherwise}. \end{cases} \tag{14}$$

According to Lemma II.1, the solution of (11) is

$$w = \begin{cases} x^n + \frac{1}{\mu}J_1^n, & (x^n + \frac{1}{\mu}J_1^n)^2 \geq \frac{2\lambda}{\mu}, \\ 0, & \text{otherwise}. \end{cases} \tag{15}$$

Hence, the ADM algorithm for solving model (5) is summarized in Algorithm 1.

Algorithm 1 ADM Algorithm for Solving **(5)**

Input: blurred image B and blur kernel k;
Initialize: $J_1^0 = J_2^0 = w^0 = 0, x^0 = y, \max_\mu = 10^6, \rho = 3, \mu = 0.01, \epsilon = 10^{-6}, n = 0$.
while not converged && $\mu \leq \max_\mu$ **do**
  Step 1: Compute w^{n+1} by **(15)**.
  Step 2: Compute x^{n+1} by **(10)**.
  Step 3: Update the multipliers and penalty parameters: $J_1^{n+1} = J_1^n + \mu(x^{n+1} - w^{n+1}), J_2^{n+1} = J_2^n + \mu(x^{n+1} * k - y), \mu = \rho\mu$.
  Step 4: Check the convergence conditions: $\|x^{n+1} * k - y\|_\infty < \epsilon$ and $\|w^{n+1} - x^{n+1}\|_\infty < \epsilon$.
$n = n + 1$.
end while
Output: sharp edges x^n.

Note that the ADM algorithm originates from Augmented Lagrangian Method (ALM), thus its convergence property is guaranteed in theory.

[1]The sub-problem can also be regarded as an ℓ^0-smoothing approach [19].

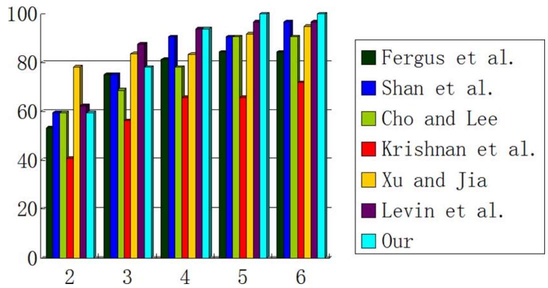

Fig. 1. Evaluation results: Cumulative histogram of the deconvolution error ratio across test examples.

Finally, we summarize our ℓ^0-regularized kernel estimation method in Algorithm 2.[2]

Algorithm 2 Algorithm for ℓ^0-Regularized Kernel Estimation

Input: blurred image B, parameters $\lambda = 0.04, \gamma = 0.001$, and kernel size: s.
for $i = 1 \rightarrow T$ **do**
  Step 1: Update kernel k by **(4)**.
  Step 2: Update sharp edges x by Algorithm 1.
$\lambda = \lambda * 0.55$.
end for
Output: blur kernel k.

B. Increasing the Robustness for Kernel Estimation

Not all edges in the interim latent image can help the kernel estimation. Like the strategy [18], we adopt the adaptive weighted method to enhance the robustness for kernel estimation. The modified model is defined as

$$\min_{x,k} \|x * k - y\|_2^2 + \gamma\|k\|_2^2 + \kappa\lambda\|x\|_0, \tag{16}$$

where $\kappa = \exp(-|r|^{0.8})$ and r is defined as

$$r = \frac{\|\sum_{\mathrm{q}\in N_h(\mathrm{p})} \nabla B(\mathrm{q})\|_2}{\sum_{\mathrm{q}\in N_h(\mathrm{p})} \|\nabla B(\mathrm{q})\|_2 + 0.5}, \tag{17}$$

in which B is the blurred image and $N_h(\mathrm{p})$ is an $h \times h$ window centered at pixel p. Equation (17) is first proposed by [5] to remove some narrow strips in the kernel estimation process.

Because r is a pixel-wise adaptive map, the optimization strategy for (16) is still the same as (2). We only need to modify (15) as

$$w = \begin{cases} x^n + \frac{1}{\mu}J_1^n, & (x^n + \frac{1}{\mu}J_1^n)^2 \geq \frac{2\kappa\lambda}{\mu}, \\ 0, & \text{otherwise}. \end{cases} \tag{18}$$

One can see that (18) has a similar form to the edge selection method discussed in [5]. This further verifies the effectiveness of model (16).

1) Some Notes About Adaptive Kernel Estimation Method: In our kernel estimation process, the adaptive ℓ^0-metric in (16) can be regarded as a selective method. This means that our kernel estimation is similar to the salient edge selection methods [3], [5], [18][3]. However, our method does not take the additional salient edge selection step (e.g., computing shock filtering, bilateral filtering, Gaussian filtering or adaptive TV denoising). This can greatly simplify the kernel estimation process and save the computing time. Furthermore, the proposed efficient solver ensures the fast convergence of our algorithm.

[2]The number of iterations T is set to be 30 in our experiment. To increase the accuracy of kernel estimates, we adopt coarse-to-fine strategy proposed by [3].

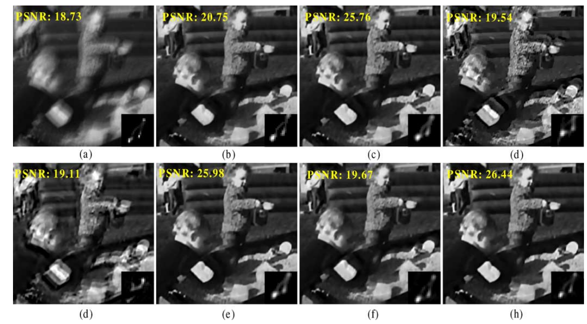

Fig. 2. One visualization example. (a) Blurred image and kernel. (b) – (g) represent the results of Fergus *et al.* [1], Shan *et al.* [2], Cho and Lee [3], Krishnan *et al.* [4], Xu and Jia [5], and Levin *et al.* [9], respectively. (h) Our results. Our recovered image has a higher PSNR value.

III. Latent Image Estimation

Once the kernel k has been estimated, we can use a variety of non-blind deconvolution methods to recover the latent image I from blurred image B. It is well known that TV regularized deblurring has been proven an effective method which can be efficiently solved by many fast algorithms. In this work, we adopt the following model to recover the latent image:

$$\min_{I} \|I * k - B\|_2^2 + \alpha\|\nabla I\|_1. \tag{19}$$

We set $\alpha = 0.001$ and employ the fast alternating minimization algorithm (e.g., [11], [12]) to solve (19).

IV. Experimental Results and Evaluations

A. Evaluation on Synthetic Data

We first perform quantitative evaluation of our method using the data set from [6]. The test data consists of 32 examples with 8 kernels. For evaluation with each example, we use the Sum of Squared Differences (SSD) ratio (see [6]) to measure the quality of restored images. We compare our method with state-of-the-art methods [1]–[5], [9]. The kernel estimates of these methods are all generated by their executable programs or source codes. The final deblurred results are obtained by using the sparse deconvolution method of [9] with the same parameter settings. In Fig. 1 we plot the cumulative histograms of the error ratios in the same way as [6].[4] As can be seen from Fig. 1, the performance of our method is better than the other methods on the whole, because all results' error ratio values are below 5. Furthermore, our method significantly outperforms those of Fergus *et al.* [1], Shan *et al.* [2], Krishnan *et al.* [4], and Cho and Lee [3]. In addition, compared with methods [3], [5], our

[3]Equation (18) provides an intuitive illustration.

[4]For example, in the histograms, a bin of 3 shows the percentage of test cases whose deconvolution error ratios are below 3 (see [6] for more details).

TABLE I
COMPARISON OF RUNNING TIME (/S)

Data	im01_ker01	im01_ker02	im01_ker03	im01_ker04	im01_ker05	im01_ker06	im01_ker07	im01_ker08
Fergus et al. [1]	417.94	318.59	370.11	278.89	403.96	385.65	380.75	459.51
Krishnan et al. [4]	139.28	128.14	119.45	218.79	85.38	225.44	215.36	212.73
Levin et al. [9]	109.18	86.19	70.30	271.32	56.81	132.48	168.50	166.21
Our	**43.07**	**35.13**	**44.74**	**39.98**	**34.24**	**53.35**	**44.98**	**44.93**

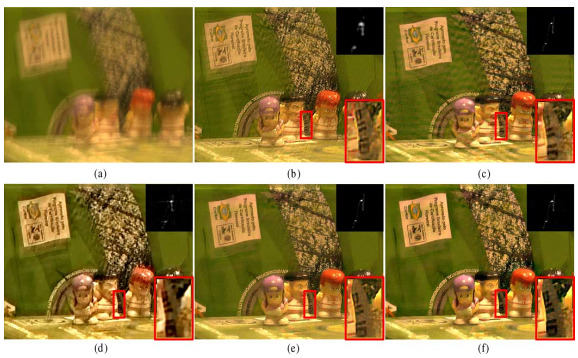

Fig. 3. A real challenging example with large blur. (a) Blurred image. (b) – (e) represent the results of Shan *et al.* [2], Cho and Lee [3], Krishnan *et al.* [4], and Xu and Jia [5], respectively. (f) Our results. The size of blur kernel is estimated as 95×95. (The images are best viewed on screen!).

method does not need to select salient edges from interim recovered images in every iterations. Thus, our kernel estimation process is simpler.

We choose an example from above experiment and show it in Fig. 2. In this example, the traditional Peak Signal to Noise Ratio (PSNR) is employed to measure the quality of the restored images. We can see that our result still outperforms the results of [1]–[5], [9] under this metric.

Our method is very efficient, because each sub-problem that is designed by our algorithm has its own explicit solutions. To demonstrate its efficiency, we compare with the methods whose source codes are available.

Table I shows the comparison of running time by using some test cases of the first experiment.[5] Compared with Fergus *et al.* [1], Krishnan *et al.* [4], and Levin *et al.* [9], our method needs much less computational time.

B. Evaluation on Real Challenging Examples With Large Blur

It is known that large blur kernel estimation is a challenging problem in blind deblurring. However, our method can estimate large blur kernels correctly. Fig. 3 is a challenging example with large blur[6]. From the estimated results of Fig. 3, we can see that the restored images of methods [2]–[4] still contain some noises and several visual artifacts. Compared with method [5], our method provides a correct blur kernel successfully, and the final restored image is visually comparable with [5].

V. CONCLUSION

We have presented an effective method for blind image deblurring. The proposed method develops an ℓ^0-regularized method to increase the robustness of kernel estimation. Although the ℓ^0-regularized problem is hard to be optimized, we develop an efficient ADM algorithm which splits this NP-hard problem into some simple sub-problems. We also show that each sub-problem has its own explicit solutions and can be easily solved. Furthermore, we find that our kernel estimation process is simpler than the state-of-the-art methods which exploit salient edges to estimate kernels. Extensive experiments testify to the superiority of our method over state-of-the-art methods, both qualitatively and quantitatively.

[5]The testing environment is a computer running MS Windows 7 64 bit version with an Intel Xeon CPU@2.53 GHz and 12 GB RAM.

[6]The image can be obtained from: http://www.cse.cuhk.edu.hk/~leojia/projects/robust_deblur/index.html.

REFERENCES

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

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

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

[4] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, 2011, pp. 2657–2664.

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

[6] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, 2009, pp. 1964–1971.

[7] Z. Hu and M.-H. Yang, "Good regions to deblur," in *Proc. Eur. Conf. Comput. Vision*, 2012, pp. 59–72.

[8] Y.-W. Tai and S. Lin, "Motion-aware noise filtering for deblurring of noisy and blurry images," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, 2012, pp. 17–24.

[9] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, 2011, pp. 2657–2664.

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

[11] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[12] D. Krishnan and R. Fergus, "Fast image deconvolution using H yper- L aplacian priors," *Adv. Neural Inf. Process. Syst.*, pp. 1033–1041, 2009.

[13] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.

[14] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, 2005, pp. 860–867.

[15] Z. Hu, J.-B. Huang, and M.-H. Yang, "Single image deblurring with adaptive dictionary learning," in *Proc. IEEE Conf. Image Process.*, 2010, pp. 1169–1172.

[16] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Sparse representation based blind image deblurring," in *Proc. IEEE Conf. Multimedia & Expo.*, 2011, pp. 1–6.

[17] 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, 2012.

[18] J. Pan, R. Liu, Z. Su, and X. Gu, Kernel Estimation from Salient Structure for Robust Motion Deblurring 2012 [Online]. Available: http://arxiv.org/abs/1212.1073

[19] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l_0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, p. 174, 2011.