IJCSI International Journal of Computer Science Issues, Volume 13, Issue 6, November 2016
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.20943/01201606.2533

The ℓ_0-norm-based Blind Image Deconvolution: Comparison and Inspiration

Hai-Song Deng[1], Wen-Ze Shao[2]

[1] School of Science, Nanjing Audit University
Nanjing, 211815, China

[2] College of Telecommunications and Information Engineering
Nanjing University of Posts and Telecommunications
Nanjing, 210003, China

Abstract—**Single image blind deblurring has been intensively studied since Fergus *et al.*'s variational Bayes method in 2006. It is now commonly believed that the blur-kernel estimation accuracy is highly dependent on the pursed salient edge information from the blurred image, which stimulates numerous ℓ_0-approximating blind deblurring methods via kinds of techniques and tricks. This paper, however, focuses on the four recent daring attempts which are all based on the simple and direct ℓ_0-norm. A systematic com- parative analysis is made towards those methods, clarifying their similarities and differences, and providing a benchmark evaluation on both the deblurring quality and computational efficiency. Results have demonstrated that the ℓ_0-norm alone is far enough to achieve top blind deblurring performance. Instead, details are to be paid with fairly more attention as working on the problem formulation as well as the algorithmic deduction. Inspired by the success of the bi-ℓ_0-ℓ_2-norm regularization, an attempt has been made to boost a recently proposed normalized sparsity-based blind deblurring method via simply borrowing core ideas behind the bi-ℓ_0-ℓ_2-norm regularization. Experimental results show that the boosting approach has leaded to a significant improvement in terms of both accuracy and efficiency. Finally, several possible extensions are discussed towards the bi-ℓ_0-ℓ_2-norm regularization.**

Keywords-blind deblurring; camera shake removal; variational Bayes; ℓ_0-norm minimization; split Bregman; half-quadratic

1. Introduction

Blind image deblurring is a highly ill-conditioned and cha- llengeable low-level vision task, the core problem of which is the accurate and robust estimation of blur-kernels. Since Fergus *et al.*'s influential variational Bayes method [1] in 2006, kinds of approaches have been proposed in the past decade [2-21, 32], and a detailed overview on the blind deblurring can be referred to Wang and Tao [22]. The focus in this paper is on the case of single image spatially-invariant blind deblurring, not only in that it could be used in a large number of potential applications due to the explosion of consumer digital photography, but also it serves as the basis for those more challengeable scenarios of multi-shot and spatially-variant blind deblurring.

Owing to the notorious ill-posedness of blur-kernel estimation, proper regularization terms or prior assumptions should be imposed so as to achieve reasonable estimates for the sharp image and the blur-kernel. Roughly speaking, most blur-kernel estimation approaches are rooted in the Bayesian framework, including two common inference principles: Maximum a Pos-teriori (MAP), e.g., [2-4, 9-13, 16-21, 32], and Variational Bayes (VB), e.g., [1, 5-8]. Note that in [14, 15], the MAP and VB principles are even combined in a sequential manner. It is also interesting to note that, most blind deblurring approaches concentrate intensively on the prior modeling of sharp images and the algorithmic deduction; as for the blur-kernels, they are either free of prior or imposed by a naive ℓ_q-norm-based prior $(0 \le q \le 2)$. In the subsequent, VB and MAP approaches are briefly reviewed, with an emphasis on the choice of priors for the images.

VB Methods. Fergus *et al.* [1] use a mixture-of-Gaussians (MoG) prior to model the image, with hyper-parameters in the prior learned in advance; and in the experiments, they empirically find that the MAP formulation with the same image and kernel priors completely fails. Inspired by Fergus *et al.*'s se-minal work, Levin *et al.* [5] provide a more profound analysis on blind deblurring, following which a simpler VB posteriori inference scheme is deduced [6]. Recently, a new blind motion deblurring method is proposed by Babacan *et al.* [7], imposing a general sparseness-inspired prior on the image and claiming that the non-informative Jeffreys prior is found more powerful than other options in practice, e.g., MoG [1, 6]. This finding is consistent with the theoretical analysis in [8], suggesting that the Jeffreys prior is optimal *to a certain degree*.

However, we should note that Babacan *et al.*'s empirical analysis does not build on the provided code by Levin *et al.* [6]. Actually, with a completely fair comparison[1] we have found that Levin *et al.* [6] outperforms Babacan *et al.* [7] in terms of the blur-kernel estimation accuracy.

MAP Methods. Compared with the VB principle, MAP is practiced more commonly. The reason lies in that, (*i*) it is more intuitive; (*ii*) it allows a simpler problem formulation; (*iii*) it is more flexible for imposing the regularization terms; and (*iv*) it typically leads to more efficient numerical implementation. Of course, many approaches of this type also exploit sparse image priors. Nevertheless, in order to achieve comparative or higher precision of blur-kernel estimation than the VB methods, MAP commonly approximates the ℓ_0-norm using sorts of techniques for dominant edge prediction as core clues to kernel estimation, either explicitly or implicitly. The explicit approximation includes the ℓ_p-norm-based image priors ($0<p<<1$) [16, 17], the reweighted ℓ_2-norm-based prior [18], the unnatural ℓ_0-norm-based sparse prior [2], the normalized sparsity-based sparse prior [19], and the log-operator-based sparse prior [20], i.e., logarithm of the Jeffreys prior in VB [7, 8]. As for those implicit ones [9-15], they usually include an additional step of texture/noise removal utilizing the smoothing and shock filters after performing an ℓ_2-based image estimation. As a matter of fact, [14, 15] have also incorporated the VB approach [6] to further improve the kernel estimation accuracy.

Following the discussion above, it is easily concluded that current leading blind deblurring approaches heavily depend on the ℓ_0-approximating image priors. The present paper, however, concentrates on the daring attempts by direct use of the ℓ_0-norm for blind image deblurring. Specifically, we make a systematic comparative analysis on the four recently reported papers [2, 3, 32, 4], clarify their similarities and differences in both problem formulation and algorithmic implementation, and give a benchmark evaluation on both the blind deblurring quality and computational efficiency. The experimental results show that the ℓ_0-norm-based image prior alone is far enough to achieve superior blind deblurring performance. And, details are to be paid with fairly more attention when working on the problem formulation as well as the algorithmic deduction. Inspired by the success of the bi-ℓ_0-ℓ_2-norm regularization, an attempt has been made to boost a recently proposed normalized sparsity-based blind deblurring method via simply borrowing core ideas behind the bi-ℓ_0-ℓ_2-norm regularization. Experimental results show that the boosting approach has leaded to a significant improvement in terms of both accuracy and efficiency.

[1] With the benchmark dataset in [5], the comparison in terms of the accuracy of blur-kernel estimation is made based on the same non-blind deconvolution algorithm in [23], using the Matlab codes provided respectively by Levin *et al.* [6] and Babacan *et al.* [7].

2. State-of-the-art ℓ_0-norm-based Single Image Blind Deblurring

Following the terminology of existing approaches [1-22, 32], the blurred image $\boldsymbol{y}$ is modeled by the spatially-invariant con- volution, mathematically formulated as

$$\boldsymbol{y} = \boldsymbol{k} * \boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{x}$ is the clear image, $\boldsymbol{k}$ is the blur-kernel, $*$ stands for a convolution operator, and $\boldsymbol{n}$ is assumed to be an additive white Gaussian noise. The task of blind image deblurring is generally divided into two independent stages, i.e., estimation of the blur-kernel $\boldsymbol{k}$ and then a non-blind deconvolution of the clear image $\boldsymbol{x}$ given the found $\boldsymbol{k}$. The inherent reason is that image priors exploited in the two stages are not necessarily the same.

To the best of our knowledge, state-of-the-art ℓ_0-norm-based natural image blind deblurring methods include [2, 3, 32, 4], all of which are based on the MAP. A basic formulation for the MAP estimates of $\boldsymbol{x}$ and $\boldsymbol{k}$ can be presented as

$$\min_{\boldsymbol{x},\boldsymbol{k}} \lambda \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \alpha_x \mathcal{Q}_x(\boldsymbol{x}) + \alpha_k \mathcal{Q}_k(\boldsymbol{k}), \tag{2}$$

where λ, α_x, α_k are positive tuning parameters, and $\mathcal{Q}_x(\boldsymbol{x})$, $\mathcal{Q}_k(\boldsymbol{k})$ are the regularization terms reflecting our priors on $\boldsymbol{x}$ and $\boldsymbol{k}$. We note that, in [2, 3, 32, 4] the positivity and unit sum constraints on the kernel $\boldsymbol{k}$ are also considered as minimizing (2) with respect to $\boldsymbol{k}$.

2.1. Unnatural ℓ_0-norm-based approach [2]

The specific objective in [2] for estimating the blur-kernel from the blurred image is

$$\min_{\boldsymbol{x},\boldsymbol{k}} \lambda \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \alpha_x \textstyle\sum_m \min(1, \frac{|(\nabla \mathbf{x})_m|^2}{\varepsilon^2}) + \alpha_k \| \mathbf{k} \|_2^2, \tag{3}$$

where ε is a positive small parameter, $\mathbf{x}$ and $\mathbf{k}$ are the vec-torized versions of $\boldsymbol{x}$ and $\boldsymbol{k}$, respectively, and $\nabla \triangleq (\nabla_h; \nabla_v)$ with ∇_h, ∇_v being the matrix versions of the first-order differ-ence operators D_h, D_v in the horizontal and vertical directions. With the half-quadratic regularization strategy[24], the authors [2] prove that (3) is equivalent to

$$\begin{aligned} \min_{\boldsymbol{x},z,\boldsymbol{k}} \; & \lambda \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \alpha_k \| \mathbf{k} \|_2^2 \\ & + \alpha_x \textstyle\sum_m \{ | \mathbf{z}_m |^0 + \frac{1}{\varepsilon^2} ((\nabla \mathbf{x})_m - \mathbf{z}_m)^2 \}, \end{aligned} \tag{4}$$

where $| \mathbf{z}_m |^0$ is a number to the zero power, and in implemen- tation ε is chosen as a decreasing sequence. Then, (4) can be solved by applying the block coordinate descent method.

Actually, using the penalty decomposition (PD) method for ℓ_0-norm minimization [25], we have alternative interpretation on Equation (4); that is, with PD the minimizers of (4) can be obtained by solving the following ℓ_0-norm-based objective

$$\min_{\boldsymbol{x},\boldsymbol{k}} \lambda \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \alpha_x \| \nabla \mathbf{x} \|_0 + \alpha_k \| \mathbf{k} \|_2^2, \tag{5}$$

where $\| \cdot \|_0$ denotes the ℓ_0-norm. Readers may refer to [25] for more details on the PD method. Notice that, as

estimating the blur-kernel in [2], the fidelity term of (5) is represented in the derivative domain. We should note that [2] has been extended to the blind deblurring of text images more recently in [34].

2.2. Reweighed ℓ_0-norm-based approach [3]

With the above alternative interpretation on the unnatural ℓ_0-norm-based blind deblurring approach [2], it becomes easier to formulate the reweighted ℓ_0-norm-based deblurring method [3] whose final objective is

$$\min_{\boldsymbol{x},\boldsymbol{k}} \lambda \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \alpha_{\boldsymbol{x}} \| \boldsymbol{\omega}^T \cdot \nabla \mathbf{x} \|_0 + \alpha_{\boldsymbol{k}} \| \mathbf{k} \|_2^2, \quad (6)$$

where $\boldsymbol{\omega} = \exp(-|\mathbf{r}|^{0.8})$ and $\mathbf{r}$ is the vectorized version of the matrix $\boldsymbol{r}$ defined as

$$\boldsymbol{r}(\mathbf{p}) = \frac{\left\| \sum_{\mathbf{q} \in N_l(\mathbf{p})} D_h \boldsymbol{y}(\mathbf{q}) \right\|_2 + \left\| \sum_{\mathbf{q} \in N_l(\mathbf{p})} D_v \boldsymbol{y}(\mathbf{q}) \right\|_2}{\sum_{\mathbf{q} \in N_l(\mathbf{p})} \left\| D_h \boldsymbol{y}(\mathbf{q}) \right\|_2 + \left\| D_v \boldsymbol{y}(\mathbf{q}) \right\|_2 + 0.5}, \quad (7)$$

where $N_l(\mathbf{p})$ is an $l \times l$ window centered at pixel $\mathbf{p}$. Note that, the motivation of incorporating the adaptive weight $\boldsymbol{\omega}$ into (5) is that, not all the edges in the sharp image $\boldsymbol{x}$ are helpful to the blur-kernel estimation [3]. Actually, Equation (7) is firstly proposed in [10] in order to remove some narrow strips which are to deteriorate the kernel estimation precision.

As in [2], it is easy to derive an analytical solution to $\boldsymbol{k}$ for (6). When it turns to minimizing (6) with respect to $\boldsymbol{x}$, the authors [3] use the idea of alternating direction method of multipliers (ADMM), originating from the operator splitting and augmented Lagrangian (OSAL) method. Specifically, a constrained minimization problem is solved to estimate $\boldsymbol{x}$ as this

$$\min_{\boldsymbol{x},\boldsymbol{k}} \alpha_{\boldsymbol{x}} \| \boldsymbol{\omega}^T \cdot \boldsymbol{\varpi} \|_0, \quad \text{s.t.} \quad \boldsymbol{\varpi} = \nabla \mathbf{x}, \ \boldsymbol{k} * \boldsymbol{x} = \boldsymbol{y}. \quad (8)$$

In addition, the continuation technique is applied to $\alpha_{\boldsymbol{x}}$ in implementation, i.e., it is decreasing as iterating. Also, the fidelity term of (6) is represented in the derivative domain as estimating both the blur-kernel and the sharp image.

2.3. Bi-ℓ_0-ℓ_2-norm regularization-based approach [4]

This approach follows a similar rationale to [2] and [3], but aims for pursing a better sharp image, which naturally leads to more accurate kernel estimation and therefore more successful blind deblurring. Specifically, a bi-ℓ_0-ℓ_2-norm regularization is imposed on both the sharp image and the blur-kernel, with the regularization terms in (2) respectively defined as

$$\mathcal{Q}_{\boldsymbol{x}}(\boldsymbol{x}) = \| \nabla \mathbf{x} \|_0 + \frac{\beta_{\boldsymbol{x}}}{\alpha_{\boldsymbol{x}}} \| \nabla \mathbf{x} \|_2^2, \quad (9)$$

$$\mathcal{Q}_{\boldsymbol{k}}(\boldsymbol{k}) = \| \mathbf{k} \|_0 + \frac{\beta_{\boldsymbol{k}}}{\alpha_{\boldsymbol{k}}} \| \mathbf{k} \|_2^2, \quad (10)$$

where $\beta_{\boldsymbol{x}}$ and $\beta_{\boldsymbol{k}}$ are positive tuning parameter [2]. Equation (9) corresponds to the ℓ_0-ℓ_2-norm-based image regularization, and Equation (10) corresponds to a similar regularization that serves the blur-kernel.

The rationale underlying (9) is the desire to get a recovered image with the salient edges from the original image, which govern the main blurring effect, as well as to force smoothness along prominent edges and inside homogenous regions. Such a sharp image is more reliable for recovering the true support of the desired motion blur-kernel than those alternative ones with unpleasant staircase artifacts. As for Equation (10), as the size of blur-kernel is large enough, it accounts for the sparseness property of typical blur-kernels, especially for the motion blur. The ℓ_0-part reduces those possibly moderate or strong isolated points in the blur-kernel, and just as practiced in [2, 3, 9, 10], the ℓ_2-part suppresses the possibly weak components.

As for the numerical implementation, the authors [4] apply the OSAL or ADMM approach to solve the ℓ_0-ℓ_2-minimization problem with respect to $\boldsymbol{x}$ and $\boldsymbol{k}$, respectively. The specific optimization problems to be solved are

$$\min_{\mathbf{w},\mathbf{x}} \lambda \| \mathbf{K}\mathbf{x} - \mathbf{y} \|_2^2 + \alpha_{\boldsymbol{x}} (\| \mathbf{w} \|_0 + \frac{\beta_{\boldsymbol{x}}}{\alpha_{\boldsymbol{x}}} \| \nabla \mathbf{x} \|_2^2), \ \text{s.t.} \ \mathbf{w} = \nabla \mathbf{x}. \quad (11)$$

$$\min_{\mathbf{g},\mathbf{k}} \lambda \| \nabla \mathbf{X} \mathbf{k} - \nabla \mathbf{y} \|_2^2 + \alpha_{\boldsymbol{k}} (\| \mathbf{g} \|_0 + \frac{\beta_{\boldsymbol{k}}}{\alpha_{\boldsymbol{k}}} \| \mathbf{k} \|_2^2), \ \text{s.t.} \ \mathbf{g} = \mathbf{k}. \quad (12)$$

where $\mathbf{K}, \nabla \mathbf{X}$ denote respectively the BCCB (block-circulant with circulant blocks) convolution matrix corresponding to $\mathbf{k}$ and $\nabla \mathbf{x}$, and $\mathbf{y}$ is the vectorized representation of $\boldsymbol{y}$. Notice from (12) that, as estimating the blur-kernel $\mathbf{k}$ the fidelity term is represented in the derivative image domain, the same as [2]. Additionally, the continuation strategy is applied to both the regularization parameters $\alpha_{\boldsymbol{x}}$ and $\alpha_{\boldsymbol{k}}$ while with different continuation factors.

We should make it clear here that, the applied continuation scheme in both [4] and [3] can be seen as a component of prior modeling for blind deblurring. But, there is no any explanation on the continuation in [3]. In fact, the essence of continuation on the image regularization parameter $\alpha_{\boldsymbol{x}}$ is to restore more and more accurate edge structures in a progressive manner, i.e., from the dominant to the faint. To make it clearer, if $\alpha_{\boldsymbol{x}}$ is set too small throughout all the iterations, the regularization effect would be so minor that very limited edges could be pursed as a core clue to blur-kernel estimation, therefore leading to poor quality estimated kernels; on the contrast, if $\alpha_{\boldsymbol{x}}$ is set too large, we will get a very cartooned image which generally has fairly less accurate edges while accompanied by unpleasant staircase artifacts in the homogeneous areas, hence reducing the kernel estimation precision. To alleviate this problem, a rational way is then to impose a decreasing sequence on $\alpha_{\boldsymbol{x}}$. A counterpart analysis can be made on

[2] The proposed ℓ_0-ℓ_2-norm regularization on $\mathbf{x}$ or $\mathbf{k}$ is somewhat akin to the elastic net regularization [26], which combines the ℓ_2- and the ℓ_1-norms in the ridge and LASSO regression [27]. However, the interest here is specifically in ℓ_0 and not ℓ_1, as it has been demonstrated both theoretically [5, 8, 28] and em-pirically [1] that a cost function (2) with an ℓ_1-norm-based image prior leads to a trivial and therefore a useless solution.

α_k, too. It is interesting to note that the decreasing ε in (4) appears to play the role of continuation, but actually it does not. On one hand, when we look at (3) the decreasing ε urges the image prior to approximate the ℓ_0-norm; on the other hand, it is just a component of the PD approach when we look at (5).

2.4. Incremental sparse approximation-based approach [32]

The core idea of [32] is to estimate the blur-kernel first from only the strongest edges in the image and then gradually refine this estimate by allowing for weaker and weaker edges. From the discussion above on the continuation scheme, we observe that the underlying ideas among [32, 3, 4] are similar to each other in some sense. And, it is now not hard to get why [32] is called the incremental sparse approximation-based approach.

TABLE 1. Comparative Summary of ℓ_0-norm-based Methods [2, 3, 32, 4]

Technical component	[2]	[3]	[32]	[4]
ℓ_0-image prior	√	√	√	√
ℓ_2-image prior	×	×	×	√
ℓ_0-kernel prior	×	×	×	√
ℓ_2-kernel prior	√	√	×	√
Continuation - α_x	×	√	√	√
Continuation - α_k	×	×	×	√
Additional trick for priors[3]	×	√	×	×
Numerical scheme	PD	ADMM	PG	ADMM
FFT	√	√	×	√
Derivative image fidelity	√	×	√	×
Derivative kernel fidelity	√	√	√	√
Multi-scale	√	√	√	√

In spite of the similarity in idea, the optimization problem in [32] does not build on (2) directly. Instead, a constrained one given as follows is proposed

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{k}} \ \tfrac{1}{2}\| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 \\ &\text{s.t.} \ \ \| \left|\nabla \mathbf{x}\right| \|_0 \le \alpha_x , \end{aligned} \tag{13}$$

where $\left|\nabla \mathbf{x}\right|$ is the vector formed by the gradient magnitude of $\mathbf{x}$ at each pixel. Except for the problem formulation, another major contribution in [32] is use of the projected gradient (PG) method for alternatively estimating $\mathbf{x}$ and $\mathbf{k}$. We just take $\mathbf{x}$ for example considering the limited paper space. The PG method seeks a solution of $\mathbf{x}$ by iterative updates of the form

$$\mathbf{x} \leftarrow \mathcal{P}_{\mathcal{X}}(\mathbf{x} - \tau_x \nabla_x \mathcal{F}(\boldsymbol{x},\boldsymbol{k})), \tag{14}$$

where $\mathcal{F}(\boldsymbol{x},\boldsymbol{k}) = \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 / 2$, $\mathcal{X} = \{\mathbf{x} \,|\, \| \left|\nabla \mathbf{x}\right| \|_0 \le \alpha_x\}$, τ_x is a step size, and $\mathcal{P}_{\mathcal{X}}(t)$ is the Euclidean projection of t onto $\mathcal{X}$ which can be computed in linear time via selection algorithms [33]. In order to perform the incremental sparse approximation, a strictly increasing sequence is applied to α_x as iterating, i.e., applying the continuation strategy. In addition, the same as [3] $\mathcal{F}(\boldsymbol{x},\boldsymbol{k})$ is represented in the derivative domain as estimating both the kernel and the image.

A note should be made that, one has to pay more attention to the choice of step size as operating the PG method. In [32], the respective step size for updating $\mathbf{x}$ and $\mathbf{k}$ is selected using the backtracking line search from an initial guess, however, which is not physically intuitive at all.

3. Comparative Studies

3.1. Algorithmic comparison

Table 1 presents a comparative summary of the four ℓ_0-norm-based blind deblurring approaches [2, 3, 32, 4]. It is clear that besides the ℓ_0-norm-based image prior, another two common implementation details in the four methods are the derivative representation of the fidelity term as estimating the blur-kernel and the multi-scale strategy[4] applied to the core algorithm of each method. We also see that except [2], other three methods all apply the continuation strategy to the image prior, which is empirically explained in [32] and [4] and demonstrated critical to the success of blind deblurring. A major distinction between [4] and [2, 3, 32] is that, the ℓ_2-norm-based image prior and the ℓ_0-norm-based kernel prior (also with continuation) are used in [4], too. The experiment results in Section 3 make us believe that this distinction is playing a fairly important role in achiev-ing superior blind deblurring performance. One more point is noted that it is very beneficial to make use of the fast Fourier transform (FFT) as deducing the numerical algorithm, e.g., [2, 3, 4].

Figure 1. The ground truth images and motion blur-kernels from the benchmark image dataset proposed by Levin *et al.* [5].

[3] E.g., Equation (7).

[4] The multi-scale scheme has been widely utilized in existing blind deblurring literature, e.g., [1-21, 32].

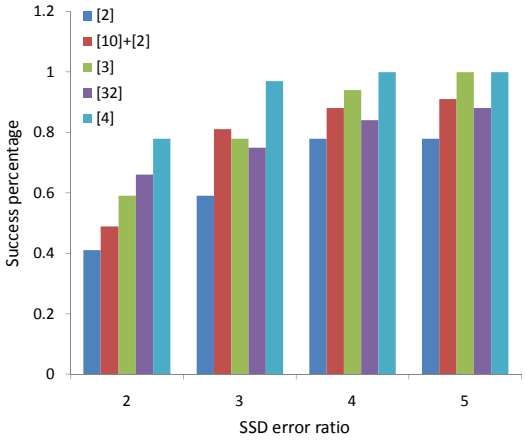

Figure 2. The cumulative histogram of SSD error ratios corresponding to blind deblurring approaches [2], [10]+[2], [3], [32], and [4]. The success percentage, i.e., SSD error ratio below 3, for each approach is: 59% [1], 81% [10]+[2], 78% [3], 75% [32], 97% [4].

3.2. Experimental comparison

This subsection makes an empirical comparative analysis on the four blind deblurring methods [2, 3, 32, 4]. Note that, [3] does not make their Matlab code online as claimed in the paper. But, the codes of methods [2, 32, 4] are available. In this case, it is better to compare different algorithms with the settings as specified in [3]. The deblurred images, the score of evaluation metric, and the computation time of [3] can be then cited directly in the present paper. For a fair comparison, we run the codes of [2, 32, 4] directly [5] to estimate blur-kernels, without altering any parameter settings; with estimated kernels, the same non-blind deconvolution method [23] is then used to generate final deblurred images for [2, 32, 4]. Note that, the non-blind deblurring method in [3] uses the same prior as in [23], i.e., the anisotropic discrete total variation.

Specifically, the evaluation is conducted on the benchmark image dataset proposed by Levin *et al.* in [5], downloaded from the author's homepage[6]. The dataset contains 32 real blurred images corresponding to 4 natural images of size 255×255 and 8 different motion blur-kernels with sizes ranging from 13×13 to 27×27, as provided in Figure 1. The SSD metric (Sum of Squared Difference) defined in [5] is used to quantify the error between the final non-blind deblurred image and the original image. Then, the same as [5, 3] the SSD error ratio between the images deblurred respectively utilizing the estimated kernel (its size is the same as the true one) and the true kernel is exploited as the final evaluation measure. The underlying idea [6] is to normalize for the fact that harder kernels achieve a larger reconstruction error even when estimated correctly.

[5] We should note that when we refer to [2] later in the experiments, we actually consider two versions of the executable C++ codes - the one reported in [2], and a combination of [2] and [10] which the authors released later on, due to its better performance. Both versions were taken from the author's webpage: http://www.cse.cuhk.edu.hk/leojia/deblurring.htm.

[6] www.wisdom.weizmann.ac.il/~levina/papers/LevinEtalCVPR2011Code.zip.

TABLE 2. Comparison of Running Time (in Seconds)

Blurred Images	[2]	[10]+[2]	[3][7]	[32]	[4][8]
Image01-Kernel01	0.912	1.158	43.07	44.36	3.765
Image01-Kernel02	0.928	1.196	35.13	33.96	3.692
Image01-Kernel03	0.956	1.150	44.74	25.98	3.802
Image01-Kernel04	0.974	1.182	39.98	70.49	3.835
Image01-Kernel05	0.955	1.165	34.24	21.07	3.798
Image01-Kernel06	0.924	1.168	53.35	36.01	3.847
Image01-Kernel07	0.945	1.178	44.98	46.91	3.753
Image01-Kernel08	0.917	1.157	44.93	47.03	3.741

Figure 2 provides the cumulative histogram of the SSD deconvolution error ratios across 32 test images for each algorithm. Following [5] and [3], the *r*'th bin in the figure counts the percentage of 32 blurred images achieving an error ratio below *r* [2]. For instance, the bar in Figure 2 corresponding to the bin 3 indicates the percentage of test images with SSD error ratios below 3. For each bin, the higher the bar, the better the deblurring performance. As pointed out by Levin *et al.* [6], deblurred images are visually plausible in general if their SSD error ratios are below 3, and in this case the blind deblurring is considered successful. The histogram provided in Figure 2 shows that the bi-ℓ_0-ℓ_2-norm-based blind deblurring method [4] achieves much higher success percentage, i.e., 97%. However, those of [32], [3], [2], and [10]+[2] are respectively 75%, 78%, 59% and 81%, demonstrating the superior performance of the more recent bi-ℓ_0-ℓ_2-norm regularization. We also see that [4] performs better than [32], [3], [2], and [10]+[2] throughout all the bins. It is interesting to note that [32], [3], and [10]+[2] have achieved similar success percentages, i.e., SSD error ratios below 3, while [3] is slightly better than [10]+[2] in other bins. Then, we can conjecture that the incorporation of [10] into [2] has played a similar role to the weight (7) in [3], which is used to select more important sharp edges as clues to blur-kernel estimation. And, we believe that [10]+[2] will work better if additionally exploiting the continuation scheme.

Table 2 also provides a comparison of running time for the 8 blurred images corresponding to Image01. The codes of [32], [3], and [4] are in Matlab, and those of [2] and [10]+[2] are in C++. It is seen that [2] and [10]+[2] are most efficient among several methods. It is also found that [4] is more efficient than [3, 32] which are of similar computational complexity. Hence, we now come to a conclusion that [4] has achieved a very good compromise between deblurring quality and running speed.

[7] The running time is directly cited from the paper [3], which is obtained on a computer with a Xeon CPU (2.53GHz) and 12GB memory, running Windows 7 (64 bit version).

[8] The running time of [4] and [32] is obtained on a portable computer with an Intel i7-4600M CPU (2.90GHz) and 8GB memory, also running Windows 7 (64 bit version).

4. Tikhonov Boosting Normalized Sparsity Regularization for Single Image Blind Deblurring

Inspired by the success of bi-ℓ_0-ℓ_2-norm regularization, it is natural to ask such a question that what we can benefit from it in the practice of blind deblurring, motivated by which, in this section we propose to boost a recent normalized sparsity (NS) [19]-based blind deblurring method via simply borrowing core ideas behind the bi-ℓ_0-ℓ_2-norm regularization. The main reason that we work on [19] is, the newly introduced NS-based image prior in [19] is claimed to favor sharp images over blurry ones, therefore better adapting blind deblurring than existing image regularizations [23, 24, 35, 30], while it is empirically proved much inferior to state-of-the-art blind deblurring methods [6, 4] in terms of deblurring quality despite its modeling advantage. For the clearness, we first briefly introduce the NS-based blind deblurring method [19], and then formulate our new Tikhonov boosting approach along with the numerical algorithm as well as empirical results on the benchmark data set [5].

4.1. NS-based blind deblurring via ISTA and IRLS [19]

With notations used in this paper, the minimizing objective function in [19] is posed as

$$\min_{\boldsymbol{x},\boldsymbol{k}} \lambda \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \frac{\|\nabla \mathbf{x}\|_1}{\|\nabla \mathbf{x}\|_2} + \alpha_k \| \mathbf{k} \|_1 . \quad (15)$$

It should be noted that, the fidelity term of (15) is represented in the derivative domain as estimating both the image and the kernel. The image regularizer is called the normalized sparsity (NS). As empirically shown in [19], for a sharp image this NS-based prior produces a lower cost than that for its blurry image, which is just what we desire in blind deblurring. To the best of our knowledge, NS is the one and only sparsity measure in the literature claimed satisfying the above property [19]. However, after numerous simulations it is found that the above property does not always hold for all the natural images.

As for solving (15), the authors of [19] respectively exploit the iterative soft-thresholding algorithm (ISTA) and iteratively reweighted least squares (IRLS) method to estimate the image and the kernel. In specific, provided the current estimates $\nabla \mathbf{x}_i$ and $\mathbf{k}_i$, the new gradient image $\nabla \mathbf{x}_{i+1}$ is estimated by

$$\nabla \mathbf{x}_{i+1} = \text{ISTA}\,(\mathbf{k}_i,\ \lambda \| \nabla \mathbf{x}_i \|_2, \nabla \mathbf{x}_i, t, N), \quad (16)$$

where t and N respectively denote the threshold parameter and the maximum iterations involved in ISTA. Note that, they are often specified by hand. As for the new blur-kernel $\mathbf{k}_{i+1}$, it is approximately updated by solving the following minimization function via conjugate gradient (CG), i.e.,

$$\min_{\boldsymbol{k}} \lambda \| \nabla \mathbf{X}_{i+1} \mathbf{k} - \nabla \mathbf{y} \|_2^2 + \alpha_k \mathbf{k}^{\mathrm{T}} \mathrm{diag}(\{1/\max(\mathbf{k}_i^j, \delta)\}) \mathbf{k}, \quad (17)$$

where $\mathbf{k}_i^j$ is the j-th item of $\mathbf{k}_i$, and δ is a very small constant. Then, $\mathbf{k}_{i+1}$ is projected onto the nonnegative and normalization constraint set. In practice, the authors perform 200 alternating updates of the gradient image and the blur-kernel. Besides, the multi-scale strategy is also applied for dealing with large-scale kernels.

4.2. Tikhonov boosting NS-based blind deblurring via ADMM

In this subsection, we are to demonstrate that the normalized sparsity alone is far enough for achieving top blind deblurring performance, just similar to the naive use of the ℓ_0-norm. And a fairly simple and straightforward remedy is the combination of the NS and the Tikhonov regularization as well as applying the continuation to the combined regularizer. The rationale is inherently the same as that of the bi-ℓ_0-ℓ_2-norm regularization. Thus, given current estimates of the image and the blur-kernel, i.e., $\mathbf{x}_i$ and $\mathbf{k}_i$, the next estimates can be obtained by solving

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{k}} \lambda \| \boldsymbol{k} * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + c_x^i \cdot \alpha_x \left(\frac{\|\nabla \mathbf{x}\|_1}{\|\nabla \mathbf{x}_i\|_2} + \frac{\beta_x}{\alpha_x} \| \nabla \mathbf{x} \|_2^2 \right) \\ + c_k^i \cdot \alpha_k \left(\| \mathbf{k} \|_0 + \frac{\beta_k}{\alpha_k} \| \mathbf{k} \|_2^2 \right), \end{aligned} \quad (18)$$

where c_x, c_k are fixed continuation factors, set respectively as $2/3$ and $4/5$ throughout the paper, and $0 \le i \le I-1$. We call (18) the Tikhonov boosting NS-based blind deblurring method.

Now, we come to the issue of algorithmic derivation for (18). Here we still use the ADMM scheme considering its simplicity and efficiency. For the sake of completeness, we are providing a detailed description of the overall algorithm in the following. We are to show that the new algorithm will lead to a significant improvement in terms of both accuracy and efficiency.

First of all, we solve the following minimization problem for $\mathbf{x}_{i+1}$

$$\min_{\boldsymbol{x}} \lambda \| \boldsymbol{k}_i * \boldsymbol{x} - \boldsymbol{y} \|_2^2 + c_x^i \cdot \alpha_x \left(\frac{\|\nabla \mathbf{x}\|_1}{\|\nabla \mathbf{x}_i\|_2} + \frac{\beta_x}{\alpha_x} \| \nabla \mathbf{x} \|_2^2 \right). \quad (19)$$

We let $\nabla \mathbf{x} = \mathbf{z}$, having an equivalent constrained optimization problem as

$$\begin{aligned} (\mathbf{z}_{i+1}, \mathbf{x}_{i+1}) = \arg\min_{\boldsymbol{x},\mathbf{z}} \frac{\lambda}{c_x^i} \| \boldsymbol{k}_i * \boldsymbol{x} - \boldsymbol{y} \|_2^2 \\ + \alpha_x \left(\frac{\|\mathbf{z}\|_1}{\|\nabla \mathbf{x}_i\|_2} + \frac{\beta_x}{\alpha_x} \| \nabla \mathbf{x} \|_2^2 \right) \ s.t.\ \ \nabla \mathbf{x} = \mathbf{z}. \end{aligned} \quad (20)$$

Applying the augmented Lagrangian scheme, we obtain

$$\begin{aligned} (\mathbf{z}_i^{l+1}, \mathbf{x}_i^{l+1}) = \\ \arg\min_{\mathbf{x},\mathbf{z}} \frac{\lambda}{c_x^i} \| \mathbf{K}_i \mathbf{x} - \mathbf{y} \|_2^2 + \alpha_x \left(\frac{\|\mathbf{z}\|_1}{\|\nabla \mathbf{x}_i^l\|_2} + \frac{\beta_x}{\alpha_x} \| \nabla \mathbf{x} \|_2^2 \right) \\ + \vartheta_x^{l*} (\nabla \mathbf{x} - \mathbf{z}) + \frac{\xi_x}{2} \| \nabla \mathbf{x} - \mathbf{z} \|_2^2, \end{aligned} \quad (21)$$

where $\mathbf{x}_0^0 = \mathbf{x}_0, \mathbf{z}_0^0 = \mathbf{0}$, $0 \le l \le L-1$, ϑ_x^l is a Lagrange multiplier for the constraint $\mathbf{z} = \nabla \mathbf{x}$ and updated by

$$\vartheta_x^{l+1} = \vartheta_x^l + \xi_x (\nabla \mathbf{x}_i^{l+1} - \mathbf{z}_i^{l+1}), \quad (22)$$

and ξ_x is an augmented Lagrangian penalty parameter, which is specified universally as 100 throughout the paper. With some directly algebraic operations, $\mathbf{z}_i^{l+1}, \mathbf{x}_i^{l+1}$ can be easily calculated by solving (21), i.e.,

$$\mathbf{z}_i^{l+1} = \mathrm{sgn}(\nabla \mathbf{x}_i^l + \frac{\vartheta_x^l}{\xi_x}) \cdot \max(abs(\nabla \mathbf{x}_i^l + \frac{\vartheta_x^l}{\xi_x} - \frac{\alpha_x}{\xi_x \|\nabla \mathbf{x}_i^l\|_2})), \quad (23)$$

$$\mathbf{x}_i^{l+1} = \left(\frac{\lambda}{c_k^i}\mathbf{K}_i^*\mathbf{K}_i + (\beta_x + \frac{\xi_x}{2})\nabla^*\nabla\right)^{-1} \cdot \left(\frac{\lambda}{c_k^i}\mathbf{K}_i^*\mathbf{y} + \frac{\xi_x}{2}\nabla^*(\mathbf{z}_i^l - \frac{1}{\xi_x}\vartheta_x^l)\right), \quad (24)$$

where $\mathbf{K}_i^*$ is the conjugate transpose of $\mathbf{K}_i$ and ϑ_x^0 is set as $\mathbf{0}$. The final $\mathbf{z}_{i+1}$, $\mathbf{x}_{i+1}$ are then estimated as $\mathbf{z}_{i+1} = \mathbf{z}_i^L$, $\mathbf{x}_{i+1} = \mathbf{x}_i^L$.

Subsequently, similar to [4] and [19] we solve the following optimization problem for $\mathbf{k}_{i+1}$ in the derivative image domain

$$\min_{\mathbf{k}} \lambda \|\nabla \mathbf{X}_{i+1}\mathbf{k} - \nabla \mathbf{y}\|_2^2 + c_k^i \cdot \alpha_k(\|\mathbf{k}\|_0 + \frac{\beta_k}{\alpha_k}\|\mathbf{k}\|_2^2). \quad (25)$$

We let $\mathbf{k} = \mathbf{s}$, and have an equivalent constrained optimization problem as

$$(\mathbf{s}_{i+1}, \mathbf{k}_{i+1}) = \arg\min_{\mathbf{k},\mathbf{s}} \frac{\lambda}{c_k^i}\|\nabla \mathbf{X}_{i+1}\mathbf{k} - \nabla \mathbf{y}\|_2^2 + \alpha_k(\|\mathbf{s}\|_0 + \frac{\beta_k}{\alpha_k}\|\mathbf{k}\|_2^2) \ \ s.t. \ \ \mathbf{k} = \mathbf{s}. \quad (26)$$

In a similar way, we apply the augmented Lagrangian scheme and have

$$(\mathbf{s}_i^{j+1}, \mathbf{k}_i^{j+1}) = \arg\min_{\mathbf{k},\mathbf{s}} \frac{\lambda}{c_k^i}\|\nabla \mathbf{X}_{i+1}\mathbf{k} - \nabla \mathbf{y}\|_2^2 + \alpha_k(\|\mathbf{s}\|_0 + \frac{\beta_k}{\alpha_k}\|\mathbf{k}\|_2^2) + \vartheta_k^{j*}(\mathbf{k} - \mathbf{s}) + \frac{\xi_k}{2}\|\mathbf{k} - \mathbf{s}\|_2^2, \quad (27)$$

where ξ_k is the augmented Lagrangian penalty parameter fixed universally as 1×10^6 throughout the paper, ϑ_x^j is the Lagrange multiplier for the constraint $\mathbf{k} = \mathbf{s}$ and updated by

$$\vartheta_k^{j+1} = \vartheta_k^j + \xi_k(\mathbf{k}_i^{j+1} - \mathbf{s}_i^{j+1}), \quad (28)$$

and $\mathbf{k}_0^0 = \mathbf{k}_0$, $\mathbf{s}_0^0 = \mathbf{0}$, $0 \le j \le J-1$. Then, $\mathbf{s}_i^{j+1}, \mathbf{k}_i^{j+1}$ can be easily calculated by solving (27) as follows

$$\mathbf{s}_i^{j+1} = \Theta_{\mathrm{Hard}}\left(\mathbf{k}_i^j + \frac{1}{\xi_k}\vartheta_k^j, \ (\frac{2\alpha_k}{\xi_k})^{\frac{1}{2}}\right), \quad (29)$$

$$\mathbf{k}_i^{j+1} = (\frac{\lambda}{c_k^i}\sum_{d\in\mathbf{\Lambda}}(\mathbf{X}_{i+1})_d^*(\mathbf{X}_{i+1})_d + (\beta_k + \frac{\xi_k}{2})\mathbf{I})^{-1} \cdot (\frac{\lambda}{c_k^i}\sum_{d\in\mathbf{\Lambda}}(\mathbf{X}_{i+1})_d^*\mathbf{y}_d + \frac{\xi_k}{2}(\mathbf{s}_i^j - \frac{1}{\xi_k}\vartheta_k^j)), \quad (30)$$

where $d \in \mathbf{\Lambda} \triangleq \{h, v\}$, $\mathbf{y}_d = \nabla_d \mathbf{y}$, and $(\mathbf{X}_{i+1})_d$ represents the convolution matrix corresponding to the image gradient $(\mathbf{x}_{i+1})_d$ i.e., $(\mathbf{x}_{i+1})_d = \nabla_d \mathbf{x}_{i+1}$. With the calculated $\mathbf{k}_i^{j+1}$, it should be further projected onto the constraint set $\mathcal{C} = \{\mathbf{k} \ge 0, \ \|\mathbf{k}\|_1 = 1\}$ considering the properties of nonnegativity and normalization for a typical blur-kernel. The final $\mathbf{s}_{i+1}$, $\mathbf{k}_{i+1}$ can be estimated as $\mathbf{s}_{i+1} = \mathbf{s}_i^J$, $\mathbf{k}_{i+1} = \mathbf{k}_i^J$.

We should note that, it is conceptually quite easy to compute (23) and (29) because of their pixel-wise operations. Besides, (24) and (30) can be also calculated very efficiently by use of the fast Fourier transform (FFT) as we have assumed that both $\mathbf{K}_i$ and $(\mathbf{X}_{i+1})_d$ are BCCB matrices. In order to account for large-scale blur-kernels, we also exploit a multi-scale strategy and uniformly set 4 scales for all the blurry images throughout the paper. Other parameters involved in (18) are also provided here: $\alpha_x = 50$, $\beta_x = 1.5$, $\alpha_k = 0.15$, $\beta_k = 1.5$, $\lambda = 100$, just for the clearness. As for I, J, L, they are set as 10. Moreover, the same as previous section, we use the non-blind deconvolution algorithm [3] for the final image deblurring.

TABLE 3. Ssd Error Rations of [19] and the Proposed Method Assuming True Sizes of Blur-kernels

Proposed	Image01	Image02	Image03	Image04
Kernel01	1.30	2.82	1.14	2.55
Kernel02	1.20	1.83	1.11	2.57
Kernel03	1.26	2.30	0.99	1.35
Kernel04	1.04	1.50	1.58	2.64
Kernel05	1.14	2.93	1.36	1.66
Kernel06	1.48	1.87	2.05	4.39
Kernel07	1.66	2.11	2.49	2.44
Kernel08	0.97	1.68	1.09	0.93
[19]	Image01	Image02	Image03	Image04
Kernel01	1.53	5.48	1.12	2.98
Kernel02	1.13	1.43	1.06	2.60
Kernel03	1.65	1.74	1.26	2.70
Kernel04	3.72	2.60	0.97	1.41
Kernel05	1.45	1.51	1.48	2.92
Kernel06	2.62	10.5	2.57	7.36
Kernel07	1.47	2.04	1.10	2.33
Kernel08	4.50	8.82	0.99	3.70

4.3. Performance analysis and validation

In this subsection, we make a comparative analysis towards [19] and the proposed Tikhonov-boosting method. The first set of experiments are conducted assuming that we have original sizes of blur-kernels in the benchmark dataset. Table 3 shows SSD error ratios of 32 test images for the two methods, and we also plot the cumulative histograms of SSD error ratios for [4] [19], and the proposed method in Figure 3. Note that, we make use of the provided MATLAB codes by the authors of [19] for estimating blur-kernels (The involved parameters are adjusted for the best performance). It is clear that our boosting method has achieved much more accurate estimation than [19]. In the meanwhile, seen from Figure 3 that our boosting approach has reached comparable performance to the previous state-of-the-art method [4] in some sense. To further test the robustness of [19] and the proposed approach, we conduct a second group of experiments where we assume that all the blur-kernels are of size 31×31. The SSD error ratios of the two methods are listed in Table 4, from which we conclude that the boosting method is more robust than [19] whose success percentage is just 56% while ours is 91%. Of course, in this case our boosting method also achieves better performance than [19] as for other bins. In the meanwhile, it is worth noting that our boosting approach is of almost the same computational efficiency as [4]. However, [19] is of much higher cost and dependent on the input sizes of blur-kernels. As for Image01, Table 5

shows the running time of [19] and our method as the sizes of eight blur-kernels are assumed known. **Note that,** the proposed method can be made real-time by integral use of parallel implementation and GPU (Graphics Processing Unit) acceleration in the future [36]-[42].

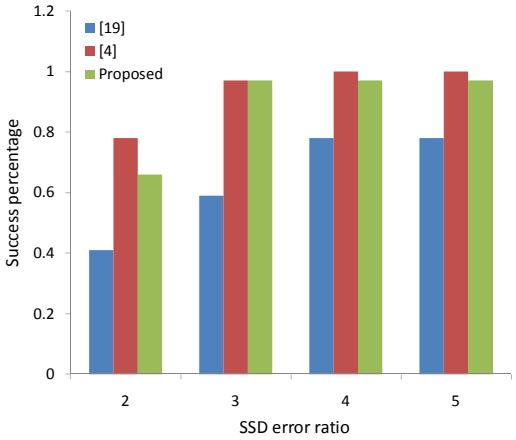

Figure 3. The cumulative histograms of SSD error ratios corresponding to [19], [4], and the proposed method in the case of knowing true blur-kernel sizes. The success percentage, i.e., SSD error ratio below 3, for each method is: 78% [19], 97% [4], 97% (Proposed).

TABLE 4. Ssd Error Ratios of [19] and the Proposed Method Assuming Sizes of All the Blur-kernels are 31×31

Proposed	Image01	Image02	Image03	Image04
Kernel01	1.22	3.35	1.20	2.44
Kernel02	1.15	1.74	1.13	2.74
Kernel03	1.10	2.46	1.00	1.37
Kernel04	1.04	1.50	1.58	2.64
Kernel05	1.16	2.97	1.39	1.78
Kernel06	1.58	2.44	1.63	2.73
Kernel07	1.52	5.72	2.29	2.41
Kernel08	0.99	2.53	1.05	1.05
[19]	Image01	Image02	Image03	Image04
Kernel01	2.09	6.98	1.24	4.08
Kernel02	1.66	2.31	1.27	5.11
Kernel03	2.99	3.20	2.08	63.5
Kernel04	1.00	1.36	0.97	1.35
Kernel05	2.78	3.18	2.03	6.80
Kernel06	3.89	46.8	3.41	6.38
Kernel07	1.43	21.7	1.69	5.29
Kernel08	0.87	1.44	0.95	0.93

TABLE 5. Running Time (in Seconds) of [19] (First Row) and Proposed Method (Second Row) as for Image01 Blurred by Different Kernels

01	02	03	04	05	06	07	08
51.65	50.59	52.60	110.5	36.62	88.65	93.47	101.4
3.786	3.655	3.848	3.896	3.778	3.889	3.788	3.789

5. Discussions and Conclusions

Single Fergus *et al.*'s influential variational Bayes approach [1] on camera shake removal, blind deblurring has been intensively studied in the past decade. It is interesting to note that the image priors in most current methods essentially attempt to approximate the ℓ_0-norm by various techniques and tricks, implying that unnatural image priors are more preferred than the common natural ones. In a distinct perspective, this paper con-centrates on four daring attempts which use the naive ℓ_0-norm for blind image deblurring [2, 3, 32, 4]. A detailed comparative analysis is made towards the four approaches, clarifying their similarities as well as differences, and providing a benchmark evaluation using Levin *et al.*'s blurred image dataset [5]. The experimental results show that the ℓ_0-norm alone is far enough to achieve top blind deblurring performance. Actually, details have to be paid with fairly more attention as working on the problem formulation and the algorithmic deduction for the ℓ_0-norm-based blind deblurring. Inspired by the success of the bi-ℓ_0-ℓ_2-norm regularization, we have also tried to boost a recent normalized sparsity-based blind deblurring method via simply borrowing core ideas behind the bi-ℓ_0-ℓ_2-norm regularization. The experimental results show that the boosting approach has leaded to a significant improvement in terms of both accuracy and efficiency.

Considering the superior performance of the bi-ℓ_0-ℓ_2-norm regularization in blind deblurring [4], we would like to discuss its possible extensions to several other imaging problems. A natural extension is to apply the regularization to non-uniform blind deblurring, just in a similar spirit to [2]. The second work is to apply it for blind video deblurring. We should note that essentially natural image priors are the key components for successful non-blind video deblurring, e.g., [29]. However, we doubt whether it is still the case as for the blind scenario. The last research to be carried out is to apply the bi-ℓ_0-ℓ_2-norm regularization for blind super-resolution (SR). Surprisingly, despite the similarity between blind deblurring and blind SR, it seems there exists a big gap between the two highly related problems. Particularly, the attention given to the nonparame- tric blind SR is very faint, while the counterpart blind deblur-ring problem is very popular and extensively treated. To the best of our knowledge, [11, 31] are the only two methods for nonparametric blind SR. And it is interesting that the work in [11] has presented a nonparametric kernel estimation approach for both blind SR and blind deblurring in a unified framework. However, it is restricting its treatment to the single-mode blur-kernels. In addition, [11] does not originate from a rigorous optimization principle, but rather builds on the detection and prediction of step edges as an important clue to the blur-kernel estimation, just similar to [10]. Because of the more rigorous problem formulation in [4], we strongly expect that a more advanced non-parametric blind SR approach can be developed based on [4].

Acknowledgment

The research is supported in part by the National Natural Science Foundation (NSF) of China (61402239, 61302178), the NSF of Jiangsu Province (BK20130868, BK20130883) and Guangxi Province (2014GXNSFAA118360), and the NSF of Jiangsu Advanced Institutions (13KJB510022), the Jiangsu Key

Laboratory of Image and Video Understanding for Social Safety (Nanjing University of Science and Technology, 30920140122007), and the NSF of NUPT (NY213007).

References

[1] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, "Removing camera shake from a single photograph," ACM Trans. Graph., 25(3), pp. 787-794, 2006.

[2] L. Xu, S. Zheng, J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," IEEE CVPR, pp.1107-1114, 2013.

[3] J. Pan, Z. Su, "Fast L_0-regularized kernel estimation for robust motion deblurring," IEEE Signal Processing Letters, 20(9), pp. 1107-1114, 2013.

[4] W.-Z. Shao, H.-B. Li, M. Elad, "Bi-ℓ_0-ℓ_2-norm regularization for blind motion deblurring," Journal of Visual Communication and Image Representation, 33, pp. 42-59, 2015.

[5] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, "Understanding blind deconvolution algorithms," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2354-2367, Oct. 2011.

[6] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," IEEE CVPR, pp. 2657-2664, 2011.

[7] S.D. Babacan, R. Molina, M.N. Do, A.K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," A. Fitzgibbon et al. (Eds.): ECCV, Part VI, LNCS 7577, pp. 341-355, 2012.

[8] D. Wipf, H. Zhang, "Analysis of Bayesian blind deconvolution," EMMCVPR, pp. 40-53, 2013.

[9] S. Cho, S. Lee, "Fast motion deblurring," ACM Trans. Graph., vol. 28, no. 5, article no. 145, Dec. 2009.

[10] L. Xu, J. Jia, "Two-phase kernel estimation for robust motion deblurring," ECCV, Part I, LNCS 6311, pp. 157-170, 2010.

[11] N. Joshi, R. Szeliski, D.J. Kriegman, "PSF estimation using sharp edge prediction," IEEE CVPR, pp. 1-8, 2008.

[12] J.H. Money, S.H. Kang, "Total variation semi-blind deconvolution using shock filters," Image and Vision Computing, vol. 26, no. 2, pp. 302-314, 2008.

[13] E. Faramarzi, D. Rajan, M.P. Christensen, "Unified blind method for multi-image super-resolution and single/multi-image blur deconvolu- tion," IEEE Trans. Image Processing, vol. 22, no. 6, pp. 2101-2114, 2013.

[14] C. Wang, Y. Yue, F. Dong, Y. Tao, X. Ma, G. Clapworthy, X. Ye, "Enhancing Bayesian estimators for removing camera shake," Computer Graphics Forum, vol. 32, no. 6, pp. 113-125, 2013.

[15] C. Wang, L. Sun, P. Cui, J. Zhang, S. Yang, "Analyzing image deblurring through three paradigms," IEEE Trans. Image Processing, vol. 21, no. 1, pp. 115-129, 2012.

[16] M. Almeida, L. Almeida, "Blind and semi-blind deblurring of natural images," IEEE Trans. Image Processing, vol. 19, no. 1, pp.36-52, Jan. 2010.

[17] J. Kotera, F. Sroubek, P. Milanfar, "Blind deconvolution using alternating maximum a posteriori estimation with heavy-tailed priors," R. Wilson et al. (Eds.): CAIP, Part II, LNCS 8048, pp. 59-66, 2013.

[18] D. Krishnan, J. Bruna, R. Fergus, "Blind deconvolution with non-local sparsity reweighting," arXiv:1311.4029v2, 2014.

[19] D. Krishnan, T. Tay, R. Fergus, "Blind deconvolution using a normalized sparsity measure," IEEE CVPR, pp. 233-240, 2011.

[20] D. Perrone, R. Diethelm, P. Favaro, "Blind deconvolution via lower-bounded logarithmic image priors," EMMCVPR, LNCS, vol. 8932, pp. 112-125, 2015.

[21] Q. Shan, J. Jia, A. Agarwala, "High-quality motion deblurring from a single image," ACM SIGGRAPH, New York, NY, USA, pp. 1-10, 2008.

[22] R. Wang, D. Tao, "Recent progress in image deblurring," arXiv: 1409.6838v1, 2014.

[23] D. Krishnan, R. Fergus, "Fast image deconvolution using hyper-laplacian priors," NIPS, vol. 22, pp. 1033-1041, 2009.

[24] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," IEEE Trans. Image Processing, vol. 6, no. 2, pp. 298-311, 1997.

[25] Y. Zhang, B. Dong, Z. Lu, "L_0 Minimization for wavelet frame based image restoration", Math. Comput., vol. 82, no. 282, pp. 995-1015, 2013.

[26] H. Zou, T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society (Series B), vol. 67, no. 2, pp. 301-320, April 2005.

[27] R. Tibshirani,"Regression shrinkage and selection via the lasso," J. Royal. Statist. Soc B., vol. 58, no. 1, pp. 267-288, 1996.

[28] A. Benichoux, E. Vincent, and R. Gribonval, "A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors," ICASSP, pp.6108-6112, 2013.

[29] S.H. Chan, R. Khoshabeh, K.B. Gibson, P.E. Gill, T.Q. Nguyen, "An augmented Lagrangian method for total variation video restoration," IEEE Trans. Image Processing, vol. 20, no. 11, pp. 3097-3111, 2011.

[30] Y. Weiss and W. T. Freeman, "What makes a good model of natural images?," IEEE CVPR, pp. 1-8, 2007.

[31] T. Michaeli, M. Irani, "Nonparametric blind super-resolution," IEEE ICCV, pp. 945-952, 2013.

[32] P. Shearer, A.C. Gilbert, A.O. Hero III, "Correcting camera shake by incremental sparse approximation," IEEE ICIP, pp. 572-576, 2013.

[33] E.G. Birgin, J.M. Martinez, M. Raydan, "Nonmonotone spectral projected gradient methods on convex sets," SIAM Journal on Optimization, vol. 10, no. 4, pp. 1196-1211, 2000.

[34] J. Pan, Z. Hu, Z. Su, M.-H. Yang, "Deblurring text images via L0 regu-larized intensity and gradient prior," IEEE CVPR, pp. 2901-2908, 2014.

[35] S. Roth and M.J. Black, "Fields of experts," International Journal of Computer Vision, 82(2), pp. 205-229, 2009.

[36] K. Li, W. Ai, Z. Tang, F. Zhang, L. Jiang, K. Li, K. Hwang, "Hadoop recognition of biomedical named entity using conditional random fields," IEEE Transactions on Parallel and Distributed Systems, 26(11), pp. 3040-3051, 2015.

[37] K. Li, C. Liu, K. Li, "An approximation algorithm based on game theory for scheduling simple linear deteriorating jobs," Theoretical Computer Science, 543(10), pp. 46-51, 2014.

[38] K. Li ; W. Yang, K. Li, "Performance analysis and optimization for SpMV on GPU using probabilistic modeling," IEEE Transactions on Parallel and Distributed System, 26(1), pp. 196-205, 2015.

[39] G. Xiao, K. Li, K. Li, X. Zhou, "Efficient top-(k,l) range query processing for uncertain data based on multicore architectures," Distributed and Parallel Databases, 33(3), pp. 381–413, 2015.

[40] K. Li, J. Liu, L. Wan, S. Yin, K. Li, "A cost optimal parallel algorithm for the 0-1 knapsack problem and its performance on multicore CPU and GPU implementations," Parallel Computing, 43, pp. 27-42, 2015.

[41] T.K. Truong, K. Li, Y. Xu, "Chemical reaction optimization with greedy strategy for the 0-1 knapsack problem, " Applied Soft Computing, 13(4), pp. 1774-1780, 2013.

[42] J. Mei, K. Li, A. Ouyang, K. Li, "A profit maximization scheme with guaranteed quality of service in cloud computing, " IEEE Transaction on Computers, 64(11), 3064-3078, 2015.

[43] H. Zheng, K. Li, "A fast algorithm with less operations for length-N= q2mDFTs, " IEEE Transactions on Signal Processing, 63(3), pp. 673-683, 2015.