A MAP-Estimation Framework for Blind Deblurring Using High-Level Edge Priors

Yipin Zhou, Nikos Komodakis

A MAP-estimation Framework for Blind Deblurring Using High-level Edge Priors[*]

Yipin Zhou[1] and Nikos Komodakis[2]

[1] Brown University
yipin_zhou@brown.edu
[2] Universite Paris-Est, Ecole des Ponts ParisTech
nikos.komodakis@enpc.fr

Abstract. In this paper we propose a general MAP-estimation framework for blind image deconvolution that allows the incorporation of powerful priors regarding predicting the edges of the latent image, which is known to be a crucial factor for the success of blind deblurring. This is achieved in a principled, robust and unified manner through the use of a global energy function that can take into account multiple constraints. Based on this framework, we show how to successfully make use of a particular prior of this type that is quite strong and also applicable to a wide variety of cases. It relates to the strong structural regularity that is exhibited by many scenes, and which affects the location and distribution of the corresponding image edges. We validate the excellent performance of our approach through an extensive set of experimental results and comparisons to the state-of-the-art.

1 Introduction

The problem of blind image deconvolution has regained lately a lot of research interest in the computer vision community [24, 40, 14, 4, 25, 12, 38, 26, 1, 11, 37, 27, 13, 39, 28, 10, 29, 34, 31, 22, 33, 21, 16]. By examining more closely the various state of the art algorithms that have been proposed recently, it becomes clear that there exist at least two elements that can play a crucial role for the success of blind image deconvolution: edge prediction and the use of proper priors.

The first element relates to the ability of one to correctly predict part of the true edges of the unknown deblurred image. The more of these edges can be detected during the deconvolution process, the better for the quality of the estimated results. Of course, the challenge is that this is often very difficult to achieve due to the inherent blurriness associated with the provided input image.

The second element is that the good performance of the recent blind deconvolution algorithms relies heavily on the successful use of various types of image priors, which naturally serve the purpose of reducing the severe ill-posedness of the above problem. To mention just a few characteristic examples, there has been recently made use of priors related to the distributions of image gradients

---

[*] Part of this work was done while the first author was an intern at Ecole des Ponts ParisTech.

obeyed by natural images [7], color priors [15], normalized sparsity priors [20], compactness priors over the so-called motion density function of the camera [9], discrete MRF image priors [19], patch priors [32] as well as smoothness priors that aim to reduce ringing artifacts [30].

Fig. 1: Scenes very often exhibit important structural regularities.

In this paper we want to capitalize on the above two findings. In other words, we wish to be able to utilize priors that are even more powerful, and which will help us to estimate edges much more robustly during blind deconvolution. To do that, in this paper we rely on an additional observation, which directly relates to the fact that a large part of the images nowadays depict scenes that exhibit strong geometric regularities with respect to the location and distribution of the corresponding image edges. One of the many reasons that this happens, for instance, is because a lot of these images display man-made objects or are captured inside man-made environments (both indoor and outdoor). Actually, one certain aspect of this phenomenon has been first noticed in an earlier work by Coughlan and Yuille [6], where it has been experimentally shown to hold for a large variety of scenes (not only urban but also rural ones).

The presence of such regularities suggests an opportunity for utilizing priors that can significantly constrain (and thus hopefully improve) the estimation of edges in this case. If we take a look at the images of Fig. 1, for instance,we can immediately see that edge pixels do not appear isolated or in arbitrary curves, but instead typically form line segments. Moreover, these segments are not arbitrary either. Instead, many of them are collinear (*i.e.*, can be grouped into lines) and, furthermore, many of the resulting lines converge into vanishing points.

One of the goals of this work is exactly to allow successfully taking advantage of all such amount of scene-specific prior knowledge during the deblurring process. More generally, our aim here is to propose a sound MAP-estimation framework for blind deconvolution, based on which one would be able to incorporate in a principled, unified and robust manner multiple types of constraints or priors, importantly including any available prior information regarding prediction of image edges.

In blind deconvolution, the idea of exploiting domain-specific properties had been previously used for the deblurring of text images [3]. More generally, using scene-specific prior knowledge for improving the performance of computer vision algorithms is an idea that has already been applied with great success in other contexts in the past. For instance, one characteristic example was in the context of camera orientation estimation [5], while two more recent examples

are the works of Furukawa *et al.* [8] on multi-view stereo reconstruction, where significant improvements are shown through the use of a "Manhattan world" assumption, and Barinova *et al.* [35] on horizon estimation with also excellent results.

We conclude this section by briefly mentioning the main contributions of this work, which are as follows:

- It introduces the idea of utilizing *scene-specific* edge priors for tackling the blind image deblurring problem.
- It successfully makes use of one such prior (related to the strong regularities that exist in many of the existing images with respect to the location and distribution of their edges) for improving the performance of blind deconvolution. This prior is generic enough, goes beyond the Manhattan world assumption, and is applicable to a wide variety of cases.
- More generally, by building upon recent work [19] that shows the importance of utilizing sparse discrete MRF image priors in this context, it proposes a solid MAP-estimation framework that manages to formulate blind deconvolution as optimization over a *single unified* energy function that can take into account various types of constraints.
- Importantly, given the significance of correct edge prediction to blind deconvolution, such a framework enables one to incorporate available prior edge information (both low-level and high-level) into the deconvolution process in a principled, sound, and robust manner.

2   Our blind deconvolution approach

In this section we describe the MAP-estimation framework that we propose for blind image deconvolution. As usual, we are considering a model where the blurry input image $\mathbf{I}$ is assumed to be the result of a convolution of a latent image $\mathbf{x}$ with a blur kernel $\mathbf{k}$ plus some additive noise $\mathbf{n}$, *i.e.*,

$$\mathbf{I} = \mathbf{x} \otimes \mathbf{k} + \mathbf{n}, \tag{1}$$

where the symbol $\otimes$ denotes the convolution operator.

In blind deconvolution, we need to recover both $\mathbf{x}$ and $\mathbf{k}$ using as only input the image $\mathbf{I}$. To that end, here we propose minimizing an energy function of the following form:

$$E(\mathbf{k}, \mathbf{x}, \mathbf{e}|\mathbf{I}) = E_{\text{data}}(\mathbf{k}, \mathbf{x}|\mathbf{I}) + E_{\text{kernel}}(\mathbf{k}) + E_{\text{img}}(\mathbf{x}|\mathbf{e}) + E_{\text{edge}}(\mathbf{e}|\mathbf{x}). \tag{2}$$

This energy consists of 4 main terms, corresponding to a data term $E_{\text{data}}(\mathbf{k}, \mathbf{x}|\mathbf{I})$, a prior-related term $E_{\text{kernel}}(\mathbf{k})$ concerning the kernel $\mathbf{k}$, a prior-related term $E_{\text{img}}(\mathbf{x}|\mathbf{e})$ for the image $\mathbf{x}$, and a prior-related term $E_{\text{edge}}(\mathbf{e}|\mathbf{x})$ concerning the image edges of the latent image $\mathbf{x}$ (where $\mathbf{e}$ is an appropriate set of variables used for specifying image edges).

The role of these terms is to properly constraint (in a soft manner) all the different elements involved in blind deconvolution with the goal of ensuring that,

in the end, convergence towards the correct kernel will take place. The last two terms are of particular importance in this regard, since their role is to help together for correctly predicting some of the main structural elements of $\mathbf{x}$ that play a crucial role for obtaining high quality blur kernels. We next define each of the above terms, and also explain their role in more detail.

2.1 Data term $E_{\mathbf{data}}(\mathbf{k}, \mathbf{x}|\mathbf{I})$

This is a standard data term used in blind deconvolution, defined as

$$E_{\mathrm{data}}(\mathbf{k}, \mathbf{x}|\mathbf{I}) = \|\mathbf{k} \otimes \mathbf{x} - \mathbf{I}\|^2. \tag{3}$$

It essentially corresponds to a negative log-likelihood term for the case where the noise $\mathbf{n}$ in equation (1) is assumed as white Gaussian.

2.2 Blur kernel prior term $E_{\mathbf{kernel}}(\mathbf{k})$

For the blur kernel $\mathbf{k}$ we select a Laplacian prior to impose on it. This leads to utilizing an l_1-norm penalty term as $E_{\mathrm{kernel}}(\mathbf{k})$, *i.e.*,

$$E_{\mathrm{kernel}}(\mathbf{k}) = \rho \cdot \|\mathbf{k}\|_1. \tag{4}$$

Such a term is known to be sparsity-inducing, leading to kernels with few non-zero entries, which is an assumption that holds true in most of the cases encountered in practice. This is especially true for kernels due to camera shake, which is the most common example. Of course, employing alternative priors for $\mathbf{k}$ (that might be more appropriate for other cases) is also possible within the proposed framework.

2.3 Image prior term $E_{\mathbf{img}}(\mathbf{x}|\mathbf{e})$

For defining this term, we will draw upon recent work that shows the importance for blind deconvolution of imposing a discrete piecewise-constant MRF prior on image $\mathbf{x}$ (which inherently promotes sparsity). This, at first, means that the elements of image $\mathbf{x}$ are assumed to take values from a discrete label set, *i.e.*, $\mathbf{x} \in \mathcal{L}_n$, where $\mathcal{L}_n$ denotes the set of quantized images that contain a restricted number of at most n intensities or colors (n is supposed to be small).

Under this assumption, $E_{\mathrm{img}}(\mathbf{x}|\mathbf{e})$ is then given by the following formula

$$E_{\mathrm{img}}(\mathbf{x}|\mathbf{e}) = \sum_{(p,q)\in\mathcal{E}} w_{pq}(\mathbf{e})[x_p \neq x_q] + \lambda \sum_{(p,q)\in\mathcal{E}} (x_p - x_q)^2, \tag{5}$$

where $\mathcal{E}$ denotes the set of pairs of pixels that are considered to be neighbors in the MRF graph, and $[\cdot]$ equals 1 if the expression inside the brackets is true and zero otherwise.

The first term in eq. (5) above corresponds to a weighted discrete Potts model [2], which penalizes the assignment of different labels (*i.e.*, colors/intensities) to

neighboring image pixels. By definition, such a model promotes L_0 sparsity and its role here is to impose a piecewise-constant structure on image $\mathbf{x}$. The idea of applying such a prior term to blind image deconvolution has been introduced recently in [19], where it was shown its importance for avoiding trivial blur kernel solutions (such as the no-blur one) and for obtaining very high quality blur kernels.

The associated weights w_{pq} play an important role in this regard, as they are used for determining the amount of penalty that should be assigned to an intensity (or color) discontinuity across pixels p, q of image $\mathbf{x}$. In this work, given that the variables $\mathbf{e}$ should already be predicting which of the pixels of the deblurred image $\mathbf{x}$ belong to image edges, we choose to define the weights w_{pq} in terms of these edge-related variables $\mathbf{e}$. More specifically, we use the following formula[3] for setting these weights:

$$w_{pq}(\mathbf{e}) = \begin{cases} w_{\text{edge}}, & \text{if } p \text{ or } q \text{ is edge pixel based on } \mathbf{e} \\ w_{\text{non-edge}}, & \text{otherwise,} \end{cases} \qquad (6)$$

where w_{edge} and $w_{\text{non-edge}}$ are 2 parameters that satisfy $w_{\text{edge}} \ll w_{\text{non-edge}}$. The result of this is that label discontinuities in image $\mathbf{x}$ are penalized much less if there is evidence (according to variables $\mathbf{e}$) that there should actually exist an edge in the deblurred image. In doing so, the goal is to allow the edges of image $\mathbf{x}$ to be much better aligned with the true edges of the deconvolved image, which, as mentioned already, is important for high quality kernel estimation. In this manner we aim to be able to successfully transfer any edge-related prior knowledge (as encoded by variables $\mathbf{e}$) onto correctly estimating the structure of the image $\mathbf{x}$.

Last, concerning the term $\lambda \cdot \sum_{(p,q) \in \mathcal{E}} (x_p - x_q)^2$ also appearing in (5), its role here is to provide just a very small amount of extra regularization by penalizing large magnitude discontinuities (this can contribute a very slight image refinement in some cases). As a result, a small parameter λ, satisfying $\lambda \ll 1$ (*e.g.*, $\lambda = 10^{-3}$), should be used with it. We note that the role of this term is minor, and that the important term in (5) is the Potts term.

2.4   Edge prior term $E_{\mathbf{edge}}(\mathbf{e}|\mathbf{x})$

This term serves the purpose of allowing us to encode any available prior knowledge with regard to the edges of the latent image $\mathbf{x}$. The rationale behind its introduction is to help in the correct prediction of these edges, which is an important factor for the success of blind deconvolution. Importantly, this permits us to incorporate in a principled manner various types of such priors into our framework (ranging rom low-level to higher-level ones), where the precise specification and meaning of the corresponding variables $\mathbf{e}$ is to be updated accordingly in each case, *i.e.*, depending on the specific choice that has been made. In the next

---

[3] Other ways of expressing w_{pq} in terms of the variables $\mathbf{e}$ are also possible, but we found the above simple definition to be effective enough.

section, we describe one particular prior of this type that will be used, which is applicable to a wide variety of cases and provides strong and very useful high-level constraints.

3   Geometric parsing prior for blind deconvolution

Our motivation for introducing the scene-specific prior described in this section comes from the well known observation that many of the images today depict scenes exhibiting strong structural regularities. Here we wish to be able to successfully take into account as many of these regularities as possible, imposing at the same time assumptions that are as general as possible. To that end, we are going to rely on a geometric image parsing prior similar to the one used in recent work [35]. Such a prior is generic enough and has already been shown to successfully apply to a wide variety of cases. Essentially, the main assumptions[4] that we make about the depicted scenes, in this case, are that many of the edge pixels appearing in image $\mathbf{x}$ are part of line segments, many of these line segments can possibly be grouped into lines, and many of the resulting lines can possibly be grouped into parallel line families. These parallel line families, therefore, converge (in the image plane) into a set of vanishing points (including the so-called zenith vanishing point as well as a set of horizontal vanishing points that lie close to the horizon line).

As a result, the set of variables $\mathbf{e} = \{\mathbf{s}, \mathbf{l}, \mathbf{h}, z\}$ used in this case consists of the set of 2d line segments $\mathbf{s} = \{s_i\}$, which represent the edge segments of image $\mathbf{x}$, the set of lines $\mathbf{l} = \{l_i\}$, the set of horizontal vanishing points $\mathbf{h} = \{h_i\}$, as well as the zenith vanishing point z. Deciding, therefore, if p is an edge pixel according to these variables (as needed by the formula used for setting the weights in (6)) simply requires checking if p belongs to one of the segments in $\mathbf{s}$, *i.e.*

$$p \text{ is edge pixel (according to } \mathbf{e}) \Leftrightarrow p \in \cup s_i \ . \tag{7}$$

The corresponding prior $E_{\text{edge}}(\mathbf{e}|\mathbf{x})$ is then defined as

$$E_{\text{edge}}(\mathbf{e}|\mathbf{x}) = \sum_i E_{\text{pixel}}(p_i|\mathbf{s}, \mathbf{x}) + \sum_i E_{\text{segment}}(s_i|\mathbf{l}) + \sum_i E_{\text{line}}(l_i|\mathbf{h}, z)$$
$$+ \sum_{i,j} E_{\text{horizon}}(h_i, h_j|z) + E_{\text{prior}}(\mathbf{s}, \mathbf{l}, \mathbf{h}). \tag{8}$$

The individual terms appearing in (8) are defined similarly to [35], and essentially encode all the assumptions that we mentioned above regarding the regularities of the depicted scenes. We next briefly describe each of these terms (and we refer to [35] for a detailed explanation):

– $E_{\text{pixel}}(p_i|\mathbf{s}, \mathbf{x})$: this term encodes how well an edge pixel p_i of image $\mathbf{x}$ is explained by one of the edge segments included in set $\mathbf{s}$

---

[4] Note that these assumptions are more general than the so-called Manhattan world model.

$$E_{\text{pixel}}(p|\mathbf{s}, \mathbf{x}) = E_p(\mathbf{s}) \cdot \sum\nolimits_{q:(p,q)\in\mathcal{E}} [x_p \neq x_q],$$

where $E_p(\mathbf{s})$ is defined (for any pixel p) as

$$E_p(\mathbf{s}) = \min\Big(\theta_{\text{bg}},\ \min_i\ \theta_{\text{dist}}\cdot d(p, s_i) + \theta_{\text{grad}}\cdot d_{\text{angle}}(p, s_i)\Big).$$

In the above, $d(p, s_i)$ denotes the minimum distance from p to segment s_i, and $d_{\text{angle}}(p, s_i)$ denotes the angular difference between the local edge direction at p and the direction of s_i. The role of the term $\sum_{q:(p,q)\in\mathcal{E}} [x_p \neq x_q]$ is to ensure that that only an *edge* pixel p of the discrete image $\mathbf{x}$ can contribute (proportional to $E_p(\mathbf{s})$) to the E_{pixel} term.

– $E_{\text{segment}}(s_i|\mathbf{l})$: this terms aims at measuring how well segment s_i is explained by one of the lines in $\mathbf{l}$. It is defined as

$$E_{\text{segment}}(s|\mathbf{l}) = \min\Big(\mu_{\text{bg}}\cdot\text{length}(s),\ \min_i\ \mu_{\text{dist}}\cdot d_{\text{area}}(s, l_i)\Big),$$

where $d_{\text{area}}(s, l_i)$ measures the distance between segment s and line l_i as the area of the figure between the line and the segment divided by the cosine of the corresponding angle between the line and the segment.

– $E_{\text{line}}(l|\mathbf{h}, z)$: this encodes if the line l passes close to one of the vanishing points $\mathbf{h} \cup z$ as follows

$$E_{\text{line}}(l|\mathbf{h}, z) = \min\Big(\eta_{\text{bg}},\ \min_i(\eta_{\text{dist}}\cdot\phi(l, h_i), \eta_{\text{dist}}\cdot\phi(l, z))\Big),$$

where $\phi(l, h_i)$ measures the distance on the Gaussian sphere between the projection of l and the projection of a vanishing point h_i.

– $E_{\text{horizon}}(h_i, h_j|z)$: this term relates to measuring if the vanishing points in $\mathbf{h}$ lie close to a line in the image plane as they should. It is defined in the following manner

$$E_{\text{horizon}}(h_i, h_j|z) = \kappa_{\text{hor}}\cdot\tan\psi(h_i - h_j, L(z)),$$

where $L(z)$ is the line connecting the zenith and the principal point of the camera (assumed to lie at the center of image $\mathbf{x}$), and ψ is the absolute angle between $h_i - h_j$ and a perpendicular to $L(z)$.

– $E_{\text{prior}}(\mathbf{s}, \mathbf{l}, \mathbf{h})$: this term corresponds to an MDL-like prior that penalizes the number of lines segments $|\mathbf{s}|$ (taking also into account their length), the number of lines $|\mathbf{l}|$ and the number of vanishing points $|\mathbf{h}|$, aiming to favor explanations of the image edges of $\mathbf{x}$ involving as few elements as possible

$$E_{\text{prior}}(\mathbf{s}, \mathbf{l}, \mathbf{h}) = \lambda_{\text{line}}|\mathbf{l}| + \lambda_{\text{vp}}|\mathbf{h}| + \lambda_{\text{segm}}\sum_i\text{length}(s_i).$$

4   MAP-estimation inference

To perform blind image deconvolution, all we need is to optimize the energy function specified in the previous section. To that end, we follow a block coordinate descent approach by separately optimizing over $\mathbf{k}$, $\mathbf{x}$, and $\mathbf{e}$. We next describe the corresponding updates that result in such a process.

4.1   Optimizing over the kernel k

If variables $\mathbf{x}$ and $\mathbf{e}$ are kept fixed, the updating of kernel $\mathbf{k}$ corresponds to solving the following minimization task

$$\min_{\mathbf{k}} \|\mathbf{x} \otimes \mathbf{k} - \mathbf{I}\|^2 + \rho\|\mathbf{k}\|_1 = \|\mathbf{M_x}\mathbf{k} - \mathbf{I}\|^2 + \rho\|\mathbf{k}\|_1, \qquad (9)$$

where $\mathbf{M_x}$ denotes the matrix corresponding to a convolution by $\mathbf{x}$.

To efficiently compute a solution to the above problem, we resort to applying the Alternating Direction Method of Multipliers [36] in a manner similar to [19]. Essentially, this amounts to introducing a replicating variable together with a decoupling quadratic term for decomposing the problem into 2 subproblems and then applying alternating minimization between them.

4.2   Optimizing over the latent image

When $\mathbf{k}$ and $\mathbf{e} = \{\mathbf{s}, \mathbf{l}, \mathbf{h}, z\}$ are kept fixed, optimization of (2) over $\mathbf{x}$ reduces to the following problem

$$\min_{\mathbf{x} \in \mathcal{L}_n} E_{\text{data}}(\mathbf{k}, \mathbf{x}|\mathbf{I}) + E_{\text{img}}(\mathbf{x}|\mathbf{e}) + \sum_p E_{\text{pixel}}(p|\mathbf{s}, \mathbf{x}) =$$

$$\|\mathbf{k} \otimes \mathbf{x} - \mathbf{I}\|^2 + \lambda \cdot \sum_{(p,q)\in\mathcal{E}} (x_p - x_q)^2 +$$

$$\sum_{(p,q)\in\mathcal{E}} w_{pq}(\mathbf{e}) \cdot [x_p \neq x_q] + \sum_p \sum_{q:(p,q)\in\mathcal{E}} E_p(\mathbf{s}) \cdot [x_p \neq x_q].$$

To again decouple the above optimization task into easy-to-handle subproblems, we similarly introduce a replicating variable $\mathbf{x}'$ together with a quadratic penalty term $\beta\|\mathbf{x}' - \mathbf{x}\|^2 = \beta\sum_p(x'_p - x_p)^2$ (which penalizes deviations between $\mathbf{x}$ and $\mathbf{x}'$) [36], leading to the following objective function

$$\min_{\mathbf{x}'\in\mathcal{L}_n,\mathbf{x}} \|\mathbf{k} \otimes \mathbf{x} - \mathbf{I}\|^2 + \lambda \sum_{(p,q)\in\mathcal{E}} (x_p - x_q)^2 +$$

$$\sum_{(p,q)\in\mathcal{E}} \Big(w_{pq}(\mathbf{e}) + E_p(\mathbf{s}) + E_q(\mathbf{s}) \Big)[x'_p \neq x'_q] + \beta\|\mathbf{x}' - \mathbf{x}\|^2 \qquad (10)$$

Applying block coordinate descent to (10) with respect to $\mathbf{x}$ and $\mathbf{x}'$ leads to the 2 subproblems described next.

Optimizing over x The subproblem with respect to $\mathbf{x}$ involves minimizing the following least squares objective

$$\min_{\mathbf{x}} \|\mathbf{k} \otimes \mathbf{x} - \mathbf{I}\|^2 + \lambda \sum_{(p,q)\in\mathcal{E}} (x_p - x_q)^2 + \beta\|\mathbf{x} - \mathbf{x}'\|^2, \qquad (11)$$

which amounts to solving the linear system shown below

$$\left(\mathbf{M_k^T M_k} + \lambda(\mathbf{M_i^T M_i} + \mathbf{M_j^T M_j}) + \beta\right)\mathbf{x} = \mathbf{M_k^T I} + \beta\mathbf{x}',$$

where $\mathbf{M_k}$, $\mathbf{M_i}$, $\mathbf{M_j}$ denote the convolution matrices for filters $\mathbf{k}$, $\mathbf{i} = [1, -1]$ and $\mathbf{j} = [1, -1]^T$.[5]

A solution is efficiently computed through the following frequency-domain operations

$$\mathbf{x} = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\mathbf{k})} \circ \mathcal{F}(\mathbf{I}) + \beta\mathcal{F}(\mathbf{x}')}{|\mathcal{F}(\mathbf{k})|^2 + \lambda(|\mathcal{F}(\mathbf{i})|^2 + |\mathcal{F}(\mathbf{j})|^2) + \beta}\right), \tag{12}$$

where $\mathcal{F}^{-1}()$ and $\mathcal{F}()$ denote inverse and forward FFTs.

Optimizing over $\mathbf{x}'$ The subproblem with respect to $\mathbf{x}'$ corresponds to mini-mizing the energy of a discrete MRF Potts model

$$\min_{\mathbf{x}' \in \mathcal{L}_n} \sum_p V_p(x_p') + \sum_{(p,q) \in \mathcal{E}} V_{pq}(x_p', x_q'),$$

that has unary potentials $V_p(x_p') = \beta \cdot (x_p' - x_p)^2$, and Potts pairwise potentials $V_{pq}(x_p', x_q') = \left(w_{pq}(\mathbf{e}) + E_p(\mathbf{s}) + E_q(\mathbf{s})\right)[x_p' \neq x_q']$. Several off-the-shelf state-of-the-art optimizers exist for this task (in our experiments we have used the FastPD algorithm due to its efficiency [18]).

4.3   Optimizing over the edge-related variables e

With $\mathbf{k}$ and $\mathbf{x}$ being fixed, minimization of (2) over $\mathbf{e}$ reduces to

$$\min_{\mathbf{e}} E_{\text{edge}}(\mathbf{e}|\mathbf{x}) + \sum_{(p,q) \in \mathcal{E}} w_{pq}(\mathbf{e})[x_p \neq x_q]. \tag{13}$$

Here the additional term $\sum_{p,q} w_{pq}(\mathbf{e})[x_p \neq x_q]$ encourages the edge segments determined by variables $\mathbf{s}$ to agree with the current edges of image $\mathbf{x}$.

The above energy function (13) is of the same form as $E_{\text{edge}}(\mathbf{e}|\mathbf{x})$, with the only difference being that the term $\sum_p E_{\text{pixel}}(p|\mathbf{s}, \mathbf{x}) = \sum_{(p,q) \in \mathcal{E}} \left(E_p(\mathbf{s}) + E_q(\mathbf{s})\right)[x_p \neq x_q]$ is now replaced by $\sum_{(p,q) \in \mathcal{E}} \left(E_p(\mathbf{s}) + E_q(\mathbf{s}) + w_{pq}(\mathbf{e})\right)[x_p \neq x_q]$. Given that the weights $w_{pq}(\mathbf{e})$ depend only on the variables $\mathbf{s}$ (see (6)), and not on $\{\mathbf{l}, \mathbf{h}, z\}$, the resulting objective function is of similar form to the energy function used in [35], and can therefore be optimized using the same method, which is highly efficient.

---

[5] W.l.o.g. here we assume 4-connectivity for the edges $\mathcal{E}$ of the MRF graph, which is what leads to the use of the filters $\mathbf{i}$ and $\mathbf{j}$ in this case.

4.4   Multi-resolution inference

As is usually the case with blind deconvolution, optimization proceeds in a coarse-to-fine fashion. This is done in order to avoid bad local minima and to be able to deal with large kernels. We therefore use a multi-resolution image pyramid, and iterate the updates described in §4.1-§4.3 at each pyramid level. For efficiency and fast convergence, the variables $\mathbf{x}$ (§4.2), $\mathbf{x}'$ (§4.2), $\mathbf{e}$, $\mathbf{k}$ are maintained throughout the whole process, and are upsampled/upscaled when going from a coarser to a finer level in the pyramid. In practice, convergence at each level is very fast, with typically 3 iterations per level being enough on average.

Given the kernel $\mathbf{k}$, the final deconvolved image can be computed by applying (12) (using the final estimated image $\mathbf{x}'$ and a small β), or by using a more advanced non-blind deconvolution algorithm such as [38], [30].

5   Experimental results

We next test our method on a wide variety of cases and also provide comparisons with the state of the art. We note that for all the experiments (paper and suppl. material), we used uniform parameter settings for our method. More specifically, we set $w_{\mathrm{non-edge}} = 5 \cdot 10^2$, $w_{\mathrm{edge}} = 0.05 \cdot w_{\mathrm{non-edge}}$, $\lambda = 10^{-3}$, $n = 20$, while all parameters of energy E_{edge} were kept constant and set as in [35]. As already explained, λ should always be set to a small enough value (it is used only for adding a minor amount of regularization). Furthermore, it is enough that $w_{\mathrm{non-edge}}$ is an order of magnitude larger than w_{edge}, while any reasonably small value of n (*e.g.*, between 15 and 25) seems to suffice. In general, our method was quite robust (*i.e.*, not very sensitive) with respect to how its parameters are set. As a result, we expect the above settings to work well for any other case.

We applied our framework to a wide variety of test examples, including blurred images of scenes with structural regularities, as well as general scenes, while using a variety of blurred kernels, and also comparing with the current state-of-the-art. For the scenes with structural regularities, we made use of images from the publicly available "Eurasian cities" and "York Urban" datasets. We first show results on kernel estimation, which is the most critical part in blind deconvolution. To that end, we experimented with a wide range of challenging kernels, including ones with large sizes that introduce very significant blur. Fig. 2 compares our method with several state-of-the-art blind deconvolution algorithms. Even visually, it is clear that our estimated kernels match the ground truth much more closely (both for small and large kernel sizes). Fig. 8(a) also shows the corresponding average SSD errors with respect to the correct blur kernels, verifying the much superior performance of our method.

We next show results concerning the estimated deblurred images. Fig. 9 again compares several state-of-the-art techniques (additional results are included in the supplemental material due to lack of space). Thanks to its more accurate blur kernel estimation, our method manages to recover much better images with

| input image | true kernel | [4] | [20] | [38] | [19] | ours |
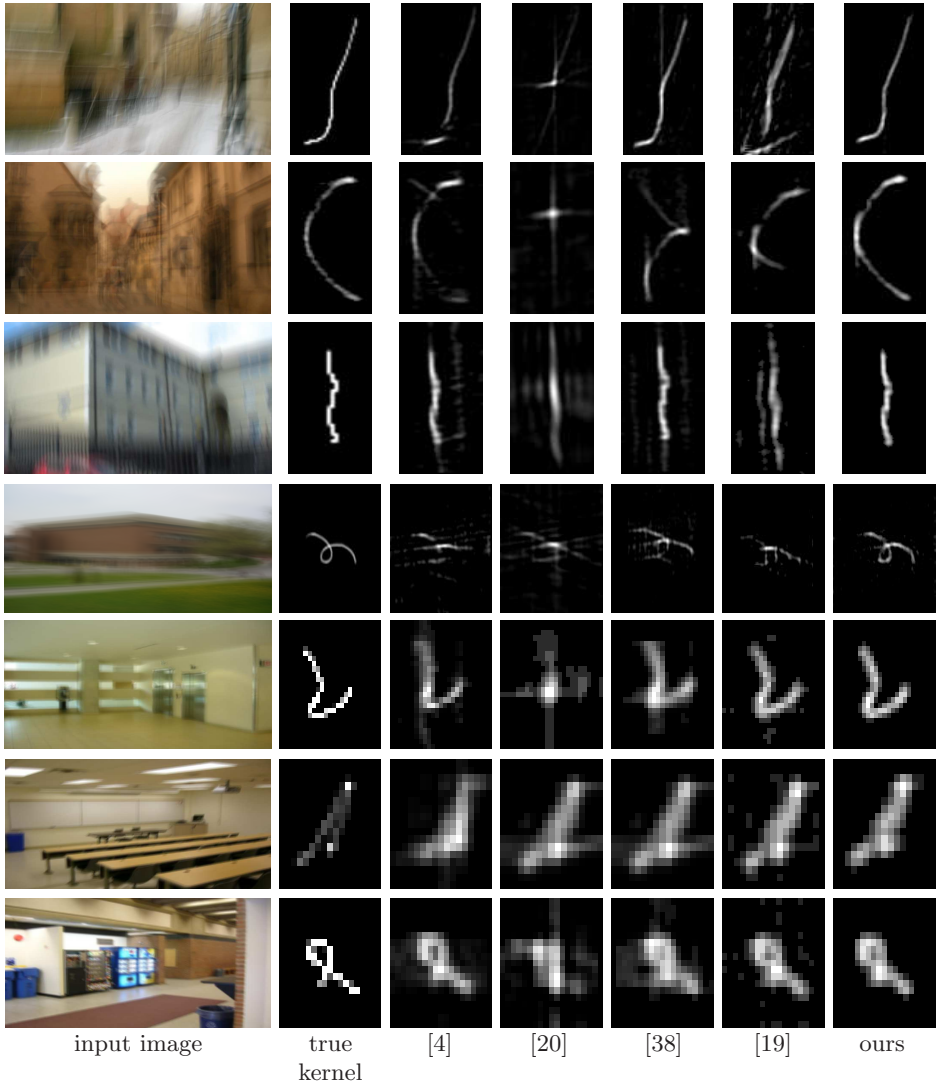
Fig. 2: Ground truth and estimated blur kernels by different deconvolution algorithms.

Fig. 3: More results by our method (input images [top], deblurred images [bottom]).

more fine details as well as less ringing artifacts. We report in Fig. 8(a) the average PSNR and SSIM scores over all the test images (including the ones in the supplementary material). We also visualize in Fig. 8(b) some of the corresponding edge segments **s** (as estimated during the course of our method for two test cases from Fig. 9). Fig. 3 contains additional deblurring results by our method (with even more being included in the supplemental material for a variety of scenes, both indoor and outdoor).

Note that [19], which is a method that also utilizes a discrete MRF image prior, has been included in all of our comparisons above. This is for illustrating the important gains coming from the incorporation of the geometric parsing prior. Fig. 4 also shows an example of the sparse discrete images $\mathbf{x}' \in \mathcal{L}_n$ that are estimated by [19] and our method. Notice, in our case, how much better $\mathbf{x}'$ can capture the true structure of the underlying latent image. To further illustrate the importance of the geometric parsing prior, we also show in Fig. 5 a comparison when this prior (*i.e.*, the term E_{edge}) is not used during blind deconvolution (in which case $w_{pq}(\mathbf{e})$ is set to a constant value).
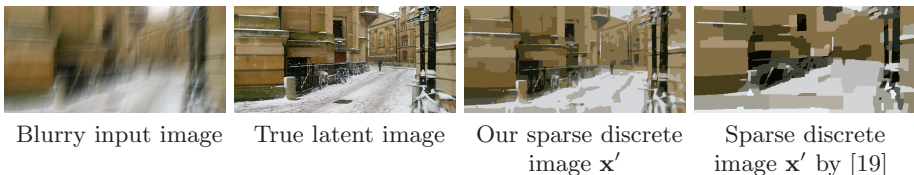

| Blurry input image | True latent image | Our sparse discrete image $\mathbf{x}'$ | Sparse discrete image $\mathbf{x}'$ by [19] |

Fig. 4: Sparse discrete images $\mathbf{x}'$ with our method and with method [19].

| Blurry input images | Results with the geometric parsing prior | Results without the geometric parsing prior |

Fig. 5: Results with and without the geometric parsing prior.

Due to the fact that our edge-related prior is incorporated into our framework in a robust and principled manner, our method is perfectly capable of handling not just scenes with structural regularities but also general scenes. To demon-
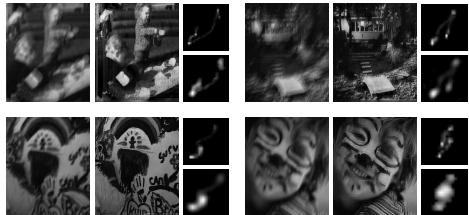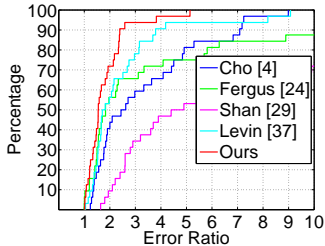
Fig. 6: Left: Error ratio evaluation for the benchmark dataset [24] (evaluation scores for other methods are taken from [23], [24]). **Right:** A few of our results for the dataset in [24] (for each one we show: input & deblurred images, true & estimated kernels).

strate that, we also apply it to the benchmark datasets [24] and [17]. Figures 6, 7 contain quantitative evaluations that verify our method's state-of-the-art performance on these datasets. We also show in Fig. 6 a few indicative deblurring results from dataset [24] (a full set of results can be found in supp. material).

Cho	Xu	Shan	Fergus	Krishnan	Whyte(lsq)	Whyte(Kri)	Hirsch	Ours
28.98	29.41	25.89	22.73	25.72	27.84	28.07	27.77	**29.71**

Fig. 7: Average PSNR values of deblurred images for various methods on dataset [17].

	[4]	[26]	[7]	[28]	Ours
kernel SSD error	32.6	64.6	46.6	53.0	**17.9**
image PSNR	22.5	21.1	22.9	24.1	**26.78**
image SSIM	0.69	0.62	0.67	0.69	**0.80**
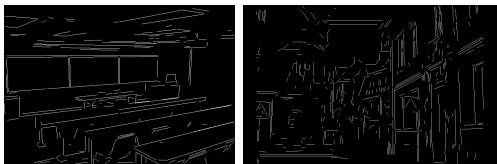
(a)   (b)

Fig. 8: In (a) the scores are computed using the best shifts for each of the estimated blur kernels and latent images. In (b) we show a visualization of the line segments **s** estimated during the course of our method for two of the test cases in Fig. 9.

6   Conclusions

We have proposed a sound MAP-estimation framework for blind image deconvolution, which uses a unified energy function that takes into account various types of constraints or priors, importantly including ones concerning the edges of the latent image. We have been able to successfully use this framework to take advantage of a very powerful prior of this type that relates to the strong structural regularities exhibited by many of the depicted scenes today. We have shown that such an approach improves the performance of blind deblurring in a wide variety of cases. Furthermore, due to the fact that the incorporation of the above prior into our framework takes place in a principled and robust manner, our method was also shown to be capable of handling even images that do not necessarily fully satisfy the above prior assumptions. More generally, we believe that our idea of utilizing scene-specific priors for improving blind deconvolution is one of great practical value, which can motivate further research and the development of new algorithms in this area.
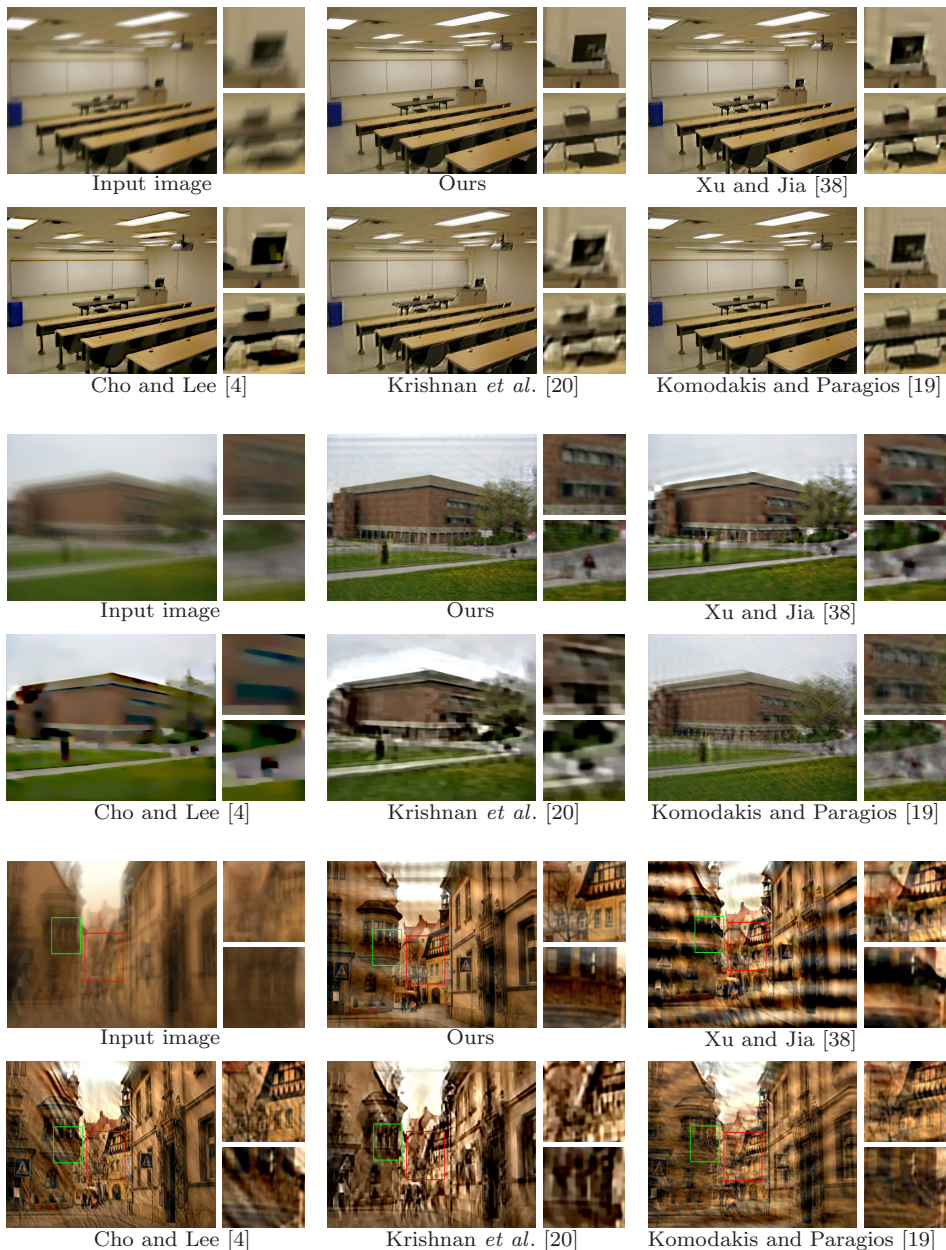
Fig. 9: Deblurred images (and corresponding close-ups) as estimated by different blind deconvolution methods. Additional results are included in the supplemental material.

References

1. Babacan, S.D., Molina, R., Katsaggelos, A.K.: Variational bayesian blind deconvolution using a total variation prior. IEEE Trans. on Image Processing (2009)
2. Blake, A., Rother, C., Brown, M., Pérez, P., Torr, P.H.S.: Interactive image segmentation using an adaptive gmmrf model. In: ECCV (2004)
3. Cho, H., Wang, J., Lee, S.: Text image deblurring using text-specific properties. In: ECCV (2012)
4. Cho, S., Lee, S.: Fast motion deblurring. In: SIGGRAPH ASIA (2009)
5. Coughlan, J.M., Yuille, A.L.: Manhattan world: Compass direction from a single image by bayesian inference. In: ICCV (1999)
6. Coughlan, J.M., Yuille, A.L.: The manhattan world assumption: Regularities in scene statistics which enable bayesian inference. In: NIPS (2000)
7. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. SIGGRAPH (2006)
8. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Manhattan-world stereo. In: CVPR (2009)
9. Gupta, A., Joshi, N., Zitnick, L., Cohen, M., Curless, B.: Single image deblurring using motion density functions. In: ECCV (2010)
10. Harmeling, S., Hirsch, M., Schlkopf, B.: Space-variant single-image blind deconvolution for removing camera shake. In: NIPS (2010)
11. Hirsch, M., Schuler, C.J., Harmeling, S., Schölkopf, B.: Fast removal of non-uniform camera shake. In: ICCV (2011)
12. Jia, J.: Single image motion deblurring using transparency. In: CVPR (2007)
13. Joshi, N., Kang, S.B., Zitnick, C.L., Szeliski, R.: Image deblurring using inertial measurement sensors. ACM Trans. Graph. 29, 30:1–30:9 (July 2010)
14. Joshi, N., Szeliski, R., Kriegman, D.J.: Psf estimation using sharp edge prediction. In: CVPR (2008)
15. Joshi, N., Zitnick, C.L., Szeliski, R., Kriegman, D.J.: Image deblurring and denoising using color priors. In: CVPR (2009)
16. Kim, T.H., Ahn, B., Lee, K.M.: Dynamic scene deblurring. In: ICCV (2013)
17. Köhler, R., Hirsch, M., Mohler, B., Schölkopf, B., Harmeling, S.: Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In: ECCV (2012)
18. Komodakis, N., Tziritas, G., Paragios, N.: Fast, approximately optimal solutions for single and dynamic MRFs. In: CVPR (2007)
19. Komodakis, N., Paragios, N.: MRF-based blind image deconvolution. In: ACCV (2012)
20. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: CVPR (2011)
21. Lee, H.S., Kwon, J., Lee, K.M.: Simultaneous localization, mapping and deblurring. In: ICCV (2011)
22. Lee, H.S., Lee, K.M.: Dense 3d reconstruction from severely blurred images using a single moving camera. In: CVPR (2013)
23. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Efficient marginal likelihood optimization in blind deconvolution. In: CVPR (2011)
24. Levin, A., Weiss, Y., Durand, F., Freeman, W.: Understanding and evaluating blind deconvolution algorithms. In: CVPR (2009)
25. Levin, A.: Blind motion deblurring using image statistics. In: NIPS (2006)

26. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Efficient marginal likelihood optimization in blind deconvolution. In: CVPR (2011)
27. Raskar, R., Agrawal, A., Tumblin, J.: Coded exposure photography: motion deblurring using fluttered shutter. SIGGRAPH pp. 795–804 (2006)
28. Raskar, R., Agrawal, A., Tumblin, J.: Coded exposure photography: motion deblurring using fluttered shutter. In: SIGGRAPH. pp. 795–804 (2006)
29. Schmidt, U., Rother, C., Nowozin, S., Jancsary, J., Roth, S.: Discriminative non-blind deblurring. In: CVPR (2013)
30. Shan, Q., Jia, J., Agarwala, A.: High-quality motion deblurring from a single image. SIGGRAPH (2008)
31. Shan, Q., Xiong, W., Jia, J.: Rotational motion deblurring of a rigid object from a single image. In: ICCV (2007)
32. Sun, L., Cho, S., Wang, J., Hays, J.: Edge-based blur kernel estimation using patch priors. In: ICCP (2013)
33. Tai, Y.W., Du, H., Brown, M.S., Lin, S.: Correction of spatially varying image and video motion blur using a hybrid camera. PAMI (2010)
34. Tai, Y., P.Tan, Brown, M.: Richardson-lucy deblurring for scenes under a projective motion path. PAMI (2011)
35. Tretyak, E., Barinova, O., Kohli, P., Lempitsky, V.: Geometric image parsing in man-made environments. IJCV 97(3) (2012)
36. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. SIAM Journal on Imaging Sciences (2008)
37. Whyte, O., Sivic, J., Zisserman, A., Ponce, J.: Non-uniform deblurring for shaken images. In: CVPR (2010)
38. Xu, L., Jia, J.: Two-phase kernel estimation for robust motion deblurring. In: ECCV (2010)
39. Yuan, L., Sun, J., Quan, L., Shum, H.Y.: Image deblurring with blurred/noisy image pairs. In: SIGGRAPH (2007)
40. Yuan, L., Sun, J., Quan, L., Shum, H.Y.: Progressive inter-scale and intra-scale non-blind image deconvolution. In: SIGGRAPH (2008)