Blind Deconvolution with Re-weighted Sparsity Promotion

Dilip Krishnan[1], Joan Bruna[2], and Rob Fergus[2]

[1]CSAIL, Massachusetts Institute of Technology, MA
[2]Courant Institute, New York University, NY

December 16, 2013

Abstract

Blind deconvolution has made significant progress in the past decade. Most successful algorithms are classified either as Variational or Maximum a-Posteriori (MAP). In spite of the superior theoretical justification of variational techniques, carefully constructed MAP algorithms have proven equally effective in practice. In this paper, we show that all successful MAP and variational algorithms share a common framework, relying on the following key principles: sparsity promotion in the gradient domain, l_2 regularization for kernel estimation, and the use of convex (often quadratic) cost functions. Our observations lead to a unified understanding of the principles required for successful blind deconvolution. We incorporate these principles into a novel algorithm that improves significantly upon the state of the art.

1 Introduction

Starting with the influential work of [7], the state of the art in blind deconvolution has advanced significantly. For blurred images involving camera translations or rotations, impressive performance levels have been achieved by a number of algorithms [5, 24, 8, 20, 15, 22, 9, 25, 23].

The simplest form of the blind deconvolution problem arises from the following formation model:

$$y = x_0 \star k_0 + n \tag{1}$$

where y is the observed blurred and noisy image, x_0 the unknown sharp image and k_0 is the unknown blur kernel. The noise n is assumed IID Gaussian noise with unknown variance σ^2. Blind deconvolution is the problem of recovering x_0 and k_0, given only the observation y. The model in 1 assumes spatially uniform blur, and can be extended to non-stationary blurs due to in-plane rotations [22]. If k_0 is known, then the problem reduces to that of non-blind deconvolution [10, 12].

Blind deconvolution is ill-posed since neither the sharp image x_0, the blur kernel k_0 or the noise variance are known. To alleviate these issues, prior assumptions on the structure of x_0 and k_0 must be employed. A commonly used prior on x_0 is the heavy-tailed prior [12], motivated from the observation that gradients of natural images follow a hyper-Laplacian distribution. Using this prior leads to good results in many applications such as non-blind deconvolution [10], super-resolution [19] and transparency separation [13]. If $\nabla x = (\nabla x)_i$, the heavy-tailed distributions used are of the form $\mathbf{p}(x) = \prod_i p(\nabla x_i)$ with $p(z) \propto e^{-|z|^\alpha}$ The exponent α is typically in the range of 0.6 to 0.8 [12]. Priors on the kernel k_0 have received lesser attention, but they usually tend to work on the sparsity of the kernel for motion blurs, such as the l_1 norm $\|k\|_1$ [18], or sparsity of coefficients under a curvlet transform [2].

Unfortunately, using the above priors in a naive alternating minimization (AM) framework leads to the trivial solution $\widehat{x_0} = y, \widehat{k_0} = \delta$, where δ is the Dirac. In [14], Levin *et al* analyze the reasons behind this phenomenon, when the heavy-tailed prior is used. The fundamental reason is quite simple: the probability of a sharp image x is lower under the commonly used heavy-tailed prior, with exponent in the range of 0.6-0.8. In their paper, Levin *et al* also identified a workaround that still used the heavy-tailed priors. The original version of this theoretically sound algorithm was proposed in [7] and called the variational method.

A simpler family of algorithms such as those of [24, 5] are categorized as Maximum a-Posteriori (MAP). The chief distinction between the variational and MAP algorithms is the use of probability distributions in the former, as opposed to point estimates in the latter. The kernel estimate $\widehat{k_0}$ is thus obtained by marginalizing the posterior distribution over all possible images x. This Bayesian approach is usually seen as a strong advantage for the variational methods since the uncertainty of an estimate is taken into account. Indeed, they perform well empirically. However, in practice, the marginalization is intractable and a series of approximations are performed to realize a practical algorithm. MAP formulations, on the other hand, use AM updates on $\widehat{x_0}$ and $\widehat{k_0}$, resulting in non-convex optimizations. In spite of this seemingly inferior formulation, in practice the best MAP formulation techniques have proven as effective as variational methods. The key to their performance is the use of additional steps to complement the AM iterations.

The main contribution of this paper is to show that the use of approximations in the variational method and various non-naive approaches in MAP methods lead to essentially the same framework. We show that sparsity inducing regularizations are the key ingredient, irrespective of whether they provide good image gradient priors or not. This helps explain why the top-performing methods all achieve similar performance. We develop a simple algorithm that is based on our insights and show that it achieves state of the art results but has only a single user-defined parameter. This is unlike other algorithms which have a number of user-defined parameters, making them hard to use in practice.

Our work has shared ground with that of [23], who also seek to explain the reasons behind the success of the variational approach. We show that most successful algorithms (not just variational) follow similar principles. Our resulting recipes are conceptually simpler than that suggested by [23], and we also provide directions for future improvements.

The variational and MAP paradigms do not cover all deconvolution algorithms. Notably, the spectral analysis based algorithm of [8] and the Radon transform based method of [26] are two examples where our current analysis does not hold.

Notations: We denote by $\mathcal{F}(x)$ the Fourier transform of x. $\nabla x = (\partial_h x, \partial_v x)$ denotes the gradient of a two-dimensional signal.

2 Review of Variational and MAP Approaches

In this section, we review the variational algorithm of [15] and [23], and the MAP algorithms of [25], [24] and [5]. These algorithms are all considered state of the art, and perform very well on the benchmark dataset of [14].

The above algorithms work in the gradient domain for the kernel estimation. Since convolution commutes with derivatives, this does not change the form of the cost function 1. The gradient space is used to determine a kernel $\widehat{k_0}$, and the final sharp image $\widehat{x_0}$ is typically recovered with a non-blind deconvolution algorithm such as [10].

2.1 Naive MAP

The naive MAP algorithm that is prone to poor solutions solves the following cost function:

$$(\widehat{\nabla x_0}, \widehat{k_0}) = \arg\min_{\nabla x, k} \lambda \|\nabla y - \nabla x \star k\|^2 + \sum_i |\nabla x_i|^\alpha \ . \tag{2}$$

Alternating minimization is usually employed: given a current estimate k_n, a new update ∇x_{n+1} is computed, and vice-versa. The regularizer on ∇x is a heavy tailed prior [14] with $\alpha < 1$. It has been shown in [14] that this cost function leads to the trivial solution $\widehat{x_0} = y, \widehat{k_0} = \delta$. This is because the trivial solution achieves the lowest cost for both the likelihood term $\|\nabla x \star k - \nabla y\|^2$ and the regularizing term $\sum_i |\nabla x|_i^\alpha$. 1 shows this phenomenon for the 32 blurred images from the dataset of [14] for values of $\alpha = 0.5$ and $\alpha = 0.8$. Heavy-tailed priors give a lower cost to the blurred image is because the blurring operation reduces the overall gradient variance, which reduces $\sum_i |\nabla x_i|^\alpha$. On the other hand, because zero gradients near strong edges become non-zero due to blur, an opposite effect is that $\sum_i |\nabla x_i|^\alpha$ is increased by blurring. For $\alpha = 0.5$ or larger , the former effect dominates and this causes the measure to prefer the blurred image. It is shown in [23], that for very small α values, the situation may be reversed. However, the resulting cost functions are unstable and difficult to minimize.

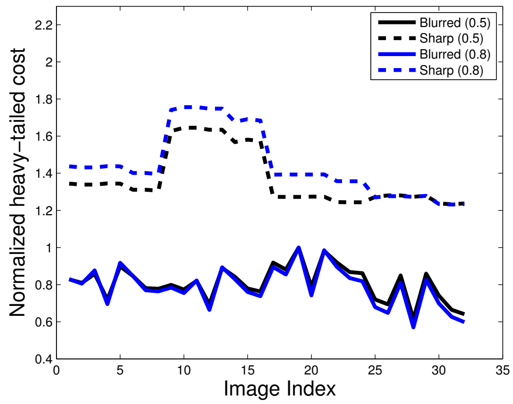


Figure 1: Comparison of costs of blurred and sharp images under heavy-tailed prior: 32 images from the dataset of [14] for $\alpha = 0.5$ and $\alpha = 0.8$: gradients of blurred images have lower cost.

2.2 Successful MAP Methods

In [5], alternating x and k updates are performed using the following equations:

$$
\begin{aligned}
x_{n+1} &= \arg\min_x \sum_j \|\partial_j x \star k_n - \partial_j y\|^2 + \alpha \|\partial_j x\|^2 \\
k_{n+1} &= \arg\min_k \sum_j \|\partial_j x_{n+1} \star k - \partial_j y\|^2 + \beta \|k\|^2 \ ,
\end{aligned}
\tag{3}
$$

where j indexes a set of partial derivative filters - in their implementation [5] use 6 filters [1]. Clearly, due to the phenomenon seen in 1, this simple formulation has little hope of succeeding since the quadratic regularization forces x_n towards the blurred image y. Therefore, [5] introduce an additional step to promote sparsity in $\{\partial_\gamma x\}$. This additional step is a shock filter [16], which suppresses gradients of small magnitude and boosts large magnitude gradients. This shock filtering step is performed after the x update step, and prior to the k estimation, thereby preventing a drift towards the trivial solution.

[24] also use a shock filter, and additionally an importance map, which is designed to down weight the importance of low magnitude gradients as well as isolated spikes. The k update step is identical to that of [5], and is given in 3, also using an l_2 (quadratic) norm on k.

The very recent work of [25] employs an ℓ_0-like prior on ∇x. The cost functions that

[1]The filters are first-order and second-order derivative filters in horizontal, vertical and diagonal directions.

they solve to update x and k are given by:

$$\begin{aligned} x_{n+1} &= \arg\min_x \|y - x \star k_n\|^2 + \lambda \Phi(\nabla x) \\ k_{n+1} &= \arg\min_k \|y - x_{n+1} \star k\|^2 + \gamma \|k\|^2 \ , \end{aligned} \tag{4}$$

where Φ is a function that approximates $\|\nabla x\|_0$. The x update step involves a series of quadratic relaxations that progressively approximate the ℓ_0 function more closely, thereby imposing sparsity on the gradients ∇x. The above papers, [5, 24, 25] and other MAP methods, periodically enforce non-negativity and sum-to-1 constraints of the entries of k. Generally, this is done after a k-update step.

2.3 Variational Methods

The variational algorithm was introduced to blind deconvolution in [7] and a simpler version of it in [15]. [7] was variational in both x and k, whereas [15] was variational only in x. Under a probabilistic interpretation of blind deconvolution, a variational estimation of k is given by:

$$\hat{k} = \arg\max_k p(k|y) = \arg\max_k \int p(y|k,x)p(x)dx \ . \tag{5}$$

However, 5 is computationally intractable, and a series of approximations are made in [7, 15] to realize a practical algorithm.

One can show [23, 15] that the final form of the resulting algorithm has the general form

$$\begin{aligned} \nabla x_{n+1} &= \arg\min_x \frac{1}{\eta_n^2} \|\nabla y - x \star k_n\|^2 + \sum_i (w_{i,n} x)^2 \\ k_{n+1} &= \arg\min_k \|\nabla y - \nabla x_{n+1} \star k\|^2 + \lambda_n \|k\|^2 \ , \end{aligned} \tag{6}$$

where η_n refers to a noise level parameter and the weights $w_{i,n}$ evolve dynamically to penalize current estimates $\nabla x_{i,n}$ of low gradient amplitudes and to "protect" large gradients. The resulting iterative minimization therefore favors a sparse ∇x_{n+1}. On the other hand, the k step consists of a ridge regression, where the parameter $\lambda_n = \mathrm{Tr}(\Sigma_n^{-1})$ and Σ_n is a diagonal covariance of ∇x_n estimated from the previous x-step [15]. As a result, the regularization strength is a measure of the overall variance in the estimate of ∇x_{n+1}.

3 The Common Components

This section explains why sparsity promoting regularizations play a central role for blind deconvolution. We argue that the main reason is not related to the prior distribution of image gradients.

3.1 Sparsity Promotion

The total variation has been extensively used as an efficient regularizer for several inverse problems [3, 17], including denoising and non-blind deconvolution. It corresponds to the ℓ_1 norm computed on image gradients ∇x, which is known [17] to promote solutions whose gradients are sparse.

This suggests that a similar sparsity-promoting prior will also be useful for the blind-deconvolution inverse problem. For that purpose, several authors [18, 4] suggested using $\|\nabla x\|_p$ with $p \leq 1$ as a prior. Similarly, all variational approaches are based on sparsity promoting priors [23]. Since the derivative is a linear, translation invariant operator, we have $\nabla y = (\nabla x_0) \star k_0 + \nabla n$. This results in a cost function of the form

$$\|\nabla y - x \star k\|^2 + \Phi(x) \ ,$$

where Φ is a sparsity-promoting function. Since natural images typically have a spectrum decaying as $\sim \omega^{-2}$ and $\mathcal{F}(\partial x)(\omega) = i\omega\mathcal{F}(x)(\omega)$, it results that the likelihood term expressed in the gradient domain is simply a reweighted ℓ_2 norm with equalized frequencies.

However, the blind deconvolution inverse problem requires not only the estimation of x_0 but also estimating the kernel k_0. We argue that enforcing sparsity of ∇x is a regularizer for $\widehat{k_0}$ which is highly efficient, even when input images do not have sparse gradients.

We shall consider a ridge regression (l_2 norm) on the kernel. Let us concentrate on the case of spatially uniform blur of Eq (1), and let us suppose the kernel k_0 has compact support of size S. The following proposition, proved in Appendix A, shows that if one is able to find an approximation of ∇x_0 which has small error in *some* neighborhood Ω of the image domain, then setting to zero ∇x outside Ω yields a good approximation of k_0. We denote $dist(i, \Omega) = \inf\{|i - j| \, , \, j \in \Omega\}$.

Proposition 3.1 *Let $y = x_0 \star k_0 + n$, with $\sum_i k_{0i} = 1$. For a given x and a given neighborhood Ω, let*

$$\begin{aligned} \epsilon^2 &= \|x - x_0\|^2_{\Omega,S} := \sum_{dist(i,\Omega) \leq S} |x_i - x_{0i}|^2 \ , \\ \gamma^2 &= \|x_0\|^2_{\Omega,S} \ , \end{aligned} \tag{7}$$

and let us assume that the matrix A whose columns are

$$(A)_j = \{x_{0j-i} \, ; |i| \leq S\}_{j \in \Omega}$$

satisfies $\lambda^2_{\min}(A) = \inf_{\sum_i y_i = 0 \, , \, \|y\| = 1} A(y) = \delta > 0$. Then, by setting

$$\tilde{x}_i = \begin{cases} x_i & \textit{if } i \in \Omega \ , \\ 0 & \textit{otherwise} \ , \end{cases} \tag{8}$$

the solution of

$$\widehat{k_0} = \arg \min_{k \, s.t. \sum_i k_i = 1} \|y - \tilde{x} \star k\|^2 + \lambda \|k\|^2 \tag{9}$$

satisfies

$$\|\widehat{k_0} - k_0\| \leq C\|k_0\| + c \ , \tag{10}$$

where $C = O(\max(\epsilon\gamma\delta^{-1}, \lambda))$ *and* $c = O(\|n\|_\Omega \gamma\delta^{-1})$.

This proposition shows that in order to recover a good estimation of the kernel, it is sufficient to obtain a good estimation of the input gradients on a certain neighborhood Ω. Sharp geometric structures and isolated singularities are natural candidates to become part of Ω, since they can be estimated from y by thresholding the gradients. This partly explains the numerical success of shock filtering based methods such as those in [5, 24]. Promoting sparsity of the image gradients thus appears as an efficient mechanism to identify the support of isolated geometric features, rather than a prior for the distribution of image gradients. In particular, Proposition 3.1 shows that images having textured or oscillatory regions do not necessarily increase the approximation error, as long as they also contain geometric features. Proposition 3.1 gives a bound on the estimation error of k_0 given a local approximation of x_0. The error is mainly controlled by ϵ, the approximation error of x_0 on the active set Ω, and δ, which depends upon the amount of diversity captured in the active set. The so-called *aperture problem* corresponds to the scenario $\delta = 0$, in which k_0 can be recovered only on the subspace spanned by the available input data.

Finally, let us highlight the connection between this result and the recent work of Ahmed *et al.* In [1], the authors show that under certain identifiability conditions, one can recover x_0 and k_0 by solving a convex program on the outer product space. In this sense, the sparsity enhancement of x helps identify a subspace Ω such that the restrictions $y\big|_\Omega$, $x\big|_\Omega$ satisfy better identifiability conditions.

3.2 ℓ_2 norm on k

The inverse problem of Eq (1) requires regularisation not only for the unknown image but also for the unknown kernel. It is seen from 2 that all the top-performing methods use an ℓ_2 ridge regression on the kernel k, which regularises the pseudo inverse associated to

$$\min_k \|\nabla y - \nabla \hat{x} \star k\|^2 \ .$$

An ℓ_2 norm gives lower cost to a diffuse kernel, which helps to push away from the trivial solution $k = \delta$. Moreover, the previous section showed that the necessary sparse regularisation of the x-step may cause the regression to be ill-conditioned due to the aperture problem.

Since the ridge regression only contains Euclidean norms, one can express it in the Fourier domain

$$\min_k \|y_f - x_f \cdot \mathcal{F}(k)\|^2 + \lambda \|\mathcal{F}(k)\|^2 \ ,$$

where y_f and x_f are respectively the Fourier transforms of ∇y and $\widehat{\nabla x}$ computed at the resolution of the kernel. It results in the well-known Wiener filters, in which frequencies with low energy in the current estimate $\widehat{\nabla x}$ are attenuated by the ridge regression. This may create kernels with irregular spectra, which translates into slow spatial decay, thus producing diffused results. In order to compensate for this effect, some authors [15] introduced a sparsity-promoting term in the estimation of k as well. Since we assume positive kernels with constant DC gain (set to 1 for simplicity), $\|k\|_1 = 1$ by construction, thus requiring a regulariser of the form $\|k\|_p$ with $p < 1$ in practice.

3.3 Convex Sub-problems

A notable aspect of the successful algorithms is the use of quadratic cost functions for both the x and k sub-problems (even though the joint problem is non-convex). Quadratic cost functions are especially simple to optimize when convolutions are involved: fast FFT or Conjugate Gradient methods may be used. For non-quadratic convex cost functions, iteratively reweighed least squares [6] may be used.

When using a convex sparsity-promoting regularizer for ∇x, one may compromise the sparsity promotion ability. However, this must be balanced against the fact that for a non-convex regularizer, it can be hard to achieve a sparse enough solution, as seen in the results of [11], which uses a non convex regulariser.

The tradeoff between sparsity-promotion and the solvability of a regularizer is therefore an important design criterion. The re-weighted methods of [15] and [25] seem to strike a good balance by solving convex (quadratic) cost functions. In our experiments with the publicly released code of [15], we found that solving each sub-problem to a high level of accuracy was crucial to the performance of the method. For example, reducing the number of conjugate gradients iterations in the ∇x update of (6) caused the performance to be much poorer. This is due to the lack of sufficient level of sparsity in the resulting ∇x.

3.4 Multi-scale Framework

Due to the non-convex nature of the blind deconvolution problem, it is easy to get stuck at a local minimum. A standard mechanism to overcome this is to use a coarse-to-fine framework for estimating the kernel. This coarse-to-fine scheme is used by all successful algorithms. At each scale in the pyramid, the upsampled kernel from the coarser level, and the downsampled blurred image from the finest level are used as an initialization. At the coarsest level, a simple initialization away from the δ kernel is used, such as a 2-pixel horizontal or vertical blur.

4 Our New Algorithm

We combine the principles described above into a new algorithm that performs above the state of the art on the benchmark dataset of [14]. In addition to the high performance, an advantage of our method is that it has only a single user-defined parameter that determines the l_2 regularization level on the estimation of k. This is in contrast with methods such as [25, 5] which have a few parameters whose settings can be hard to estimate.

We work in derivative space, using horizontal, vertical and diagonal derivative filters. As argued in section 3.1, our x update step is given by a reweighted least squares formulation which promotes solutions with isolated geometric structures, whereas the k update solves a least squares regression using ℓ_2 and ℓ_p regularisation, as discussed in section 3.2:

$$\begin{aligned} \nabla x_{n+1} &= \arg\min_x \|\nabla y - x \star k_n\|^2 + \sum_i w_{i,n} \cdot (x_i)^2 \ , \\ k_{n+1} &= \arg\min_k \|\nabla y - \nabla x_{n+1} \star k\|^2 + \lambda_1 \|k\|_2 + \lambda_2 \|k\|_{0.5} \ . \end{aligned} \tag{11}$$

The weights $w_{i,n}$ at each iteration are based on the current estimate ∇x_n. They are designed to select the regions of ∇x_n with salient geometrical features while attenuating the rest. Let $p_{i,n}$ be the patch of size R centered at pixel i of ∇x_n. We consider

$$w_{i,n} = \frac{\eta}{\eta + |\nabla x_{i,n}| \cdot \|p_{i,n}\|_2} \ . \tag{12}$$

The values of $w_{i,n}$ range between 0 and 1, and they are inversely proportional to $|\nabla x_{i,n}|$. Small gradients will have a larger regularization weight (close to 1), and as a result these small gradients will tend to be shrunk towards 0 in (11). However, point-wise reweighting does not have the capacity to separate geometrically salient structures, such as edges or isolated singularities, from textured regions. Proposition 3.1 showed that isolated gradients, corresponding to those salient geometric features, provide better identifiability than regions with dense large gradients. In order to perform this geometric detection, it is thus necessary to consider non point-wise weights. Eq (12) considers the local ℓ_2 norm $\|p_{i,n}\|_2$ over a neighbourhood at each given location. Isolated features have large local energy relative to non-sparse, textured regions. Therefore, $w_{i,n}$ will tend to attenuate those textured regions in favour of salient geometry. In our experiments, we set $R = 5$ and $\eta = \|\nabla y - \nabla x_n \star k_n\|^2$ to progressively anneal the offset in Eq 12.

Our k-update step uses ridge regression with $\lambda_1 = 3$ and $\lambda_2 \|k\|_{0.5}$ with $\lambda_2 = 3 \cdot 10^{-3}$, assuming ℓ_2 normalized input gradients: $\|\nabla y\|_2 = 1$. We solve (11) by performing 30 iterations of Conjugate Gradient, which achieves high accuracy owing to its quadratic formulation. The kernel update in (11) is solved using IRLS. After every k update, we set negative elements of k to 0, and normalize the sum of the elements to 1. We embed the entire framework in a multi-scale framework and perform 20 alternating iterations of x and k at each level.

5 Experimental Results

In this section, we compare our algorithm to that of [5, 15], and [25]. Our algorithm parameters are fixed to the values given in 4 for the tests on the dataset of [14]. For the other images in this section the weight on the l_2 regularizer for the kernel estimation was increased from 3 to 10 to prevent excessively sparse kernels.

We start with the test dataset of [14]. This consists of 4 images blurred with 8 motion blur kernels, giving rise to 32 blurred image-kernel pairs. The standard method of comparison is to compute the ratio of the mean square error of the recovered image with the mean square error of the blurred image deconvolved with the ground-truth kernel, which is known. For all comparisons in this section, we use the sparsity based non-blind deconvolution method of [15] to perform the final non-blind deconvolution step. We use the executable downloaded from the website of the authors of [25] and used existing results for [5] (provided with the code of [15]. We used the same non-blind deconvolution technique provided with the code of [14] with the same parameter settings.

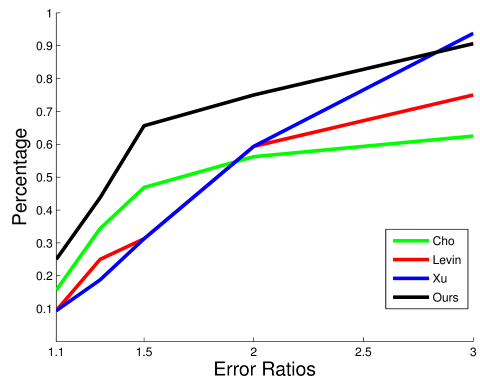
Error ratios less than 3 are considered visually good. 2 shows the cumulative error ratios and our recovered kernels for the different algorithms. It is seen that our algorithm outperforms the other methods, with 75% of the images achieving an error ratio less than 2. However, all the algorithms perform quite well. This is to be expected since each of these methods does promote sparsity of the gradients. The kernels we recover, shown in the last row, are very close to the ground-truth kernels.

Next, we compare with some real-world examples. In 3, we compare methods on an example from [25] (distributed as part of their package). We show here the output of the executable of [25], which appears somewhat inferior to the result in their paper (nevertheless still being quite good).

In 4, we use an image from [8]. The algorithm in that paper is based on spectral arguments, and so does not fall under the variational or MAP categories. Our method, [5] and [25] perform well. The output of [15] results has artifacts around the edges.

In a recent paper, [26] proposed a new algorithm to handle deblurring in the case of very high noise levels. We show that our proposed algorithm is quite robust to such situations by using an example from their paper (5). The algorithm of [25] produces significant ringing. These could possibly be reduced by parameter adjustments, but no parameters are exposed in their executable. Note that unlike the conclusions of [26], we find that [15] works quite well on this example.

The code of [23] is not available. However, we note that our method seems to perform as well as theirs on the dataset of [14]. Finally, by modifying the likelihood term using the ideas in [21], our method can be extended to the case of blur due to camera in-plane rotation.



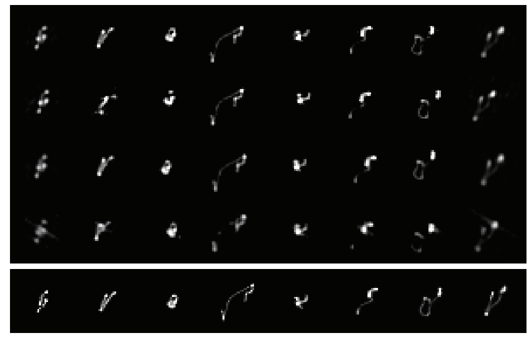


Figure 2: Left: Performance on dataset of [14]. We compare the following methods: the variational algorithm of Levin *et al.*[15]; the MAP algorithms of Cho and Lee [5] and Xu *et al.*[25]; and our new algorithm. Our algorithm is the top-performing. Right: our recovered kernels are shown: the top 4 rows correspond to the 4 images and the 8 columns correspond to the kernels we recover for each image. The last row shows the 8 ground truth kernels.

Figure 3: A real-world example from [25].

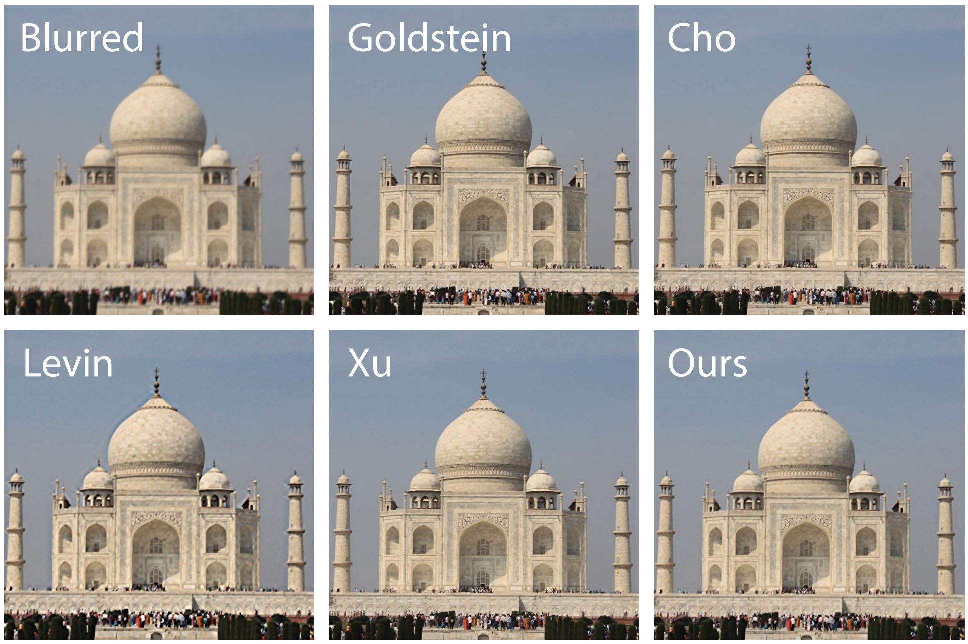


Figure 4: An example from [8]. We also include their result for comparison.

Figure 5: A real-world example from [26] that exhibits blur and high noise levels. Note that unlike [26], we find that the method of [15] also performs well. The results of [25] exhibits significant ringing.

6 Discussion

In this paper, we have discussed a number of common properties of successful blind deconvolution algorithms, with sparsity promotion being the most important. In spite of the good performance of existing methods, a number of open problems remain.

The original formulation 1 is non-convex, and alternating minimization schemes are only guaranteed to reach a local minimum. The use a multi-scale pyramid improves the numerical convergence, but it is quite possible to get stuck in sub-optimal solutions even in that scenario. These problems tend to be exacerbated in large images with many levels in the pyramid, where errors from the coarse to fine scheme may gradually accumulate. Therefore, other minimization strategies such as the convex programming based approach of [1] may prove to be better initialization strategies than the multi-scale scheme.

Existing sparsity promoting schemes are not consistent estimators of the blurring kernel k_0 because as the size of the input y increases, they are penalised by estimation errors on the x_0. Consistent estimators may be obtained by extracting stable geometric structures, using non-local regularisation terms, such as those presented in (12). Highly oscillatory textures do not corrupt the estimation of k_0, thus showing that sparsity can be highly efficient even when input images do not have sparse gradients. Reweighting schemes provide efficient algorithms for that purpose, although its mathematical properties remain an open issue.

A Proof of Proposition 3.1

Given Ω, we define $\Omega_S = \{i \ s.t. \ dist(i, \Omega) \leq S\}$, and we decompose the likelihood term as

$$\|y - \tilde{x} \star k\|^2 = \|y - \tilde{x} \star k\|^2_{\Omega_S} + \|y - \tilde{x} \star k\|^2_{\Omega_S^c} \ . \tag{13}$$

Since $x\big|_{\Omega^c} \equiv 0$, and k has compact support smaller than S, it results that

$$\|y - \tilde{x} \star k\|^2 = \|y - \tilde{x} \star k\|^2_{\Omega_S} + \|y\|^2_{\Omega_S^c} \ ,$$

and hence

$$\begin{aligned} \widehat{k_0} &= \arg\min_k \|y - \tilde{x} \star k\|^2 + \lambda \|k\|^2 \\ &= \arg\min_k \|y - \tilde{x} \star k\|^2_{\Omega_S} + \lambda \|k\|^2 \ . \end{aligned} \tag{14}$$

Since $\sum_i \widehat{k_0}_i = \sum_i k_{0i} = 1$ by construction, we shall restrict ourselves to the subspace $\{k \, ; \, \langle k, \mathbf{1} \rangle = 0\}$. If $y = x_0 \star k_0 + n$ and $e = x_0 - x$, it follows that

$$\widehat{k_0} = \arg\min_k \|x_0 \star (k_0 - k) + n - e \star k\|^2_{\Omega_S} + \lambda \|k\|^2 \ .$$

By denoting by A and $\tilde{A}$ the linear operators

$$A(y) = P_{\Omega_S}(x_0 \star y) \ , \ \tilde{A}(y) = P_{\Omega_S}(e \star y) \ ,$$

it results from (14) that

$$\begin{aligned}
\widehat{k_0} &= \left((A+\tilde{A})^T(A+\tilde{A})+\lambda\mathbf{I}\right)^{-1}[(A+\tilde{A})^T Ak_0 + (A+\tilde{A})^T n] \\
&= \left(\overline{A}+F\right)^{-1}\left(\overline{A}k_0+f\right) \ ,
\end{aligned}$$

with $\overline{A} = A^T A$, $F = A^T\tilde{A} + \tilde{A}^T A + \tilde{A}^T\tilde{A} + \lambda\mathbf{I}$ and $f = \tilde{A}^T Ak_0 + (A+\tilde{A})^T n$. Since $\delta > 0$, it results that $\overline{A} = A^T A$ is invertible in the subspace of 0-mean vectors. Since

$$\left(\overline{A}+F\right)^{-1}\left(\overline{A}k_0+f\right) = (\mathbf{1}+\overline{A}^{-1}F)^{-1}k_0 + \overline{A}^{-1}f \ ,$$

it follows that

$$\begin{aligned}
\|\widehat{k_0}-k_0\| &\leq \|\left(\mathbf{I}+\overline{A}F\right)^{-1}-\mathbf{I}\|\|k_0\| + \delta\|f\| \\
&\leq \frac{\|\overline{A}F\|}{1-\|\overline{A}F\|}\|k_0\| + \delta^{-1}(\epsilon\|k_0\|\gamma + (\gamma+\epsilon)\|n\|_\Omega) \\
&\leq O(\max(\epsilon\delta^{-1/2}, \epsilon\gamma\delta^{-1}, \lambda))\|k_0\| + O((\gamma+\epsilon)\delta^{-1}\|n\|_\Omega) \ \ \square \ .
\end{aligned}$$

References

[1] Ali Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *arXiv:1211.5608*, 2012. 7, 13

[2] Jian-Feng Cai, Hui Ji, Chaoqiang Liu, and Zuowei Shen. Blind motion deblurring from a single image using sparse approximation. In *CVPR*, pages 104–111. IEEE, 2009. 2

[3] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004. 6

[4] Tony F Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375, 1998. 6

[5] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. *SIGGRAPH ASIA*, 28(5), 2009. 1, 2, 3, 4, 5, 7, 9, 10, 11

[6] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *CPAM*, 63(1):1–38, 2009. 8

[7] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W.T. Freeman. Removing camera shake from a single photograph. *SIGGRAPH*, 25:787–794, 2006. 1, 2, 5

[8] Amit Goldstein and Raanan Fattal. Blur-kernel estimation from spectral irregularities. In *ECCV*, pages 622–635. Springer, 2012. 1, 3, 10, 12

[9] Michael Hirsch, Christian J Schuler, Stefan Harmeling, and Bernhard Scholkopf. Fast removal of non-uniform camera shake. In *ICCV*, pages 463–470. IEEE, 2011. 1

[10] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *NIPS*, 2009. 2, 3

[11] Dilip Krishnan, Terence Tay, and Rob Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, pages 233–240. IEEE, 2011. 8

[12] A. Levin, R. Fergus, F. Durand, and W.T. Freeman. Image and depth from a conventional camera with a coded aperture. *SIGGRAPH*, 26(3):70, 2007. 2

[13] A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. *PAMI*, 29(9):1647–1654, Sept 2007. 2

[14] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009. 2, 3, 4, 9, 10, 11

[15] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *CVPR, 2011*, pages 2657–2664. IEEE, 2011. 1, 3, 5, 8, 10, 11, 12

[16] Stanley Osher and Leonid I Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27(4):919–940, 1990. 4

[17] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. 6

[18] Q. Shan, J. Jia, and A. Agarwala. High quality motion deblurring from a single image. *SIGGRAPH*, 27, 2008. 2, 6

[19] Marshall F. Tappen, Bryan C. Russell, and William T. Freeman. Exploiting the sparse derivative prior for super-resolution and image demosaicing. In *SCTV*, 2003. 2

[20] Chao Wang, Yong Yue, Feng Dong, Yubo Tao, Xiangyin Ma, Gordon Clapworthy, Hai Lin, and Xujiong Ye. Nonedge-specific adaptive scheme for highly robust blind motion deblurring of natural images. *IEEE Trans. on Image*, 22(3), 2013. 1

[21] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *CVPR*, 2010. 10

[22] Oliver Whyte, Josef Sivic, and Andrew Zisserman. Deblurring shaken and partially saturated images. In *ICCV Workshops, 2011*, pages 745–752. IEEE, 2011. 1, 2

[23] David Wipf and Haichao Zhang. Revisiting bayesian blind deconvolution. *arXiv:1305.2362*, 2013. 1, 2, 3, 5, 6, 10

[24] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. *ECCV*, pages 157–170, 2010. 1, 2, 3, 4, 5, 7

[25] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural L_0 sparse representation for natural image deblurring. In *CVPR*, 2013. 1, 3, 4, 5, 8, 9, 10, 11, 12

[26] Lin Zhong, Sunghyun Cho, Dimitris Metaxas, Sylvain Paris, and Jue Wang. Handling noise in single image deblurring using directional filters. In *CVPR*, 2013. 3, 10, 12