

# Review of Recent Developments in Image Blind Deconvolution

J. Kotera

Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic.

F. Šroubek

Institute of Information Theory and Automation, AS CR, Prague, Czech Republic.

**Abstract.** Blind deconvolution (or deblurring) attempts to recover the original sharp photograph from a blurred one. Such task is very underdetermined and is therefore usually solved as maximum a posteriori probability (MAP) estimation, incorporating prior information on natural images. We provide a detailed introduction to the subject, formulate blind deconvolution as MAP estimation, and describe how individual components appeared in recent literature. We pay special attention to the distinction between traditional simultaneous estimation of blur kernel and sharp image as opposed to estimation of the kernel alone by taking into account all possible sharp images. These two branches emerged in recent papers and understanding their relationship seems to be the way towards successful deconvolution.

## Introduction



**Figure 1.** (a) A sharp photograph and (b) a photograph of the same scene corrupted with motion blur, (c) is the corresponding blur kernel (out of proportion).

Camera shake during longer exposure causes undesirable blur of the resulting photograph. Other than camera motion, there are numerous other causes of possible blur – out of focus blur, blur caused by non-stationary environment between camera and the photographed object (atmospheric blur known to astronomers), or the imaging device itself (microscope, scanner, ...) may introduce its own intrinsic blur. Regardless of the exact cause, the quality of the resulting photograph is often severely compromised and data may be (seemingly) lost. It is the aim of image blind deconvolution to reverse this process and recover the sharp image of the original photographed scene given a single blurred observation. The important fact here is that the blur is considered unknown, which is what distinguishes blind deconvolution from regular (non-blind) deconvolution – a well developed and significantly easier problem. Also, (single-channel) blind deconvolution requires no extra information like multiple photographs of the same scene (so-called multi-channel blind deconvolution, Šroubek and Milanfar [2012]; Šroubek and Flusser [2003, 2005]), constraints on the blur shape, or nonstandard hardware setup (camera with motion sensors, Joshi *et al.* [2010]), so it is perfectly applicable to everyday photography.

To make this difficult task at least slightly easier, it is almost always assumed, that the blur is linear (i.e. image-independent) and so-called space-invariant. The latter means that a single pixel in the sharp image has the same blurred response in the observed image regardless of where it is located in the image. This assumption is very often violated, for example any camera rotation about the  $z$ -axis (optical axis) results in space-variant blurring, non-planar scenes introduce another cause for non-stationarity. Even

linearity may not hold, because imaging devices can typically handle only finite interval of intensities and clip intensity values that fall outside this interval (too bright or too dark pixels), in which case the blurring is no longer signal-independent. Nevertheless, real blurs can be more or less approximated by linear space-invariant blurs and the deconvolution task is difficult enough without considering arbitrary blurs.

Let us formulate the model and resulting problem mathematically. Linear space-invariant blurring can be denoted by convolution of the original image by a blur kernel (called point spread function, PSF). Let  $x$  be the original sharp image,  $k$  the unknown blur kernel,  $y$  the observed blurred image, and  $n$  the measurement noise. Then our blurring model is

$$y = k * x + n. \quad (1)$$

The problem of recovering  $x$  from  $y$  is severely underdetermined as there is infinite number of pairs  $(x, k)$  explaining any given  $y$ . This can be more easily demonstrated in the Fourier domain, where convolution translates to (pixel-wise) multiplication. Let  $X$ ,  $K$ ,  $Y$ , and  $N$  denote respectively the Fourier transforms of  $x$ ,  $k$ ,  $y$ , and  $n$ . Then (1) translates to

$$Y = K X + N.$$

Omitting the noise for the moment, we see that for any given  $K$  with nonzero values wherever  $Y$  is nonzero, we can perform a simple division and come up with  $X$  such that  $Y = K X$ , or  $y = k * x$ . One particular trivial solution is the so-called no-blur solution, where  $y = x$ ,  $k = \delta$  (Dirac delta function, identity with respect to convolution).

It is therefore necessary to introduce additional assumptions on  $x$  and  $k$  and employ proper estimation techniques to turn blind deconvolution into a well-posed problem.

## MAP estimation

Most widely used technique is the Bayesian *maximum a posteriori* (MAP) estimation. It consists of maximizing the posterior probability of the unknown parameters for the given input. For a nice general introduction to probabilistic/energy based methods in image restoration see for example Chan and Shen [2005], the most relevant details are however explained below. In our setting, we maximize  $p(x, k|y)$  for the given input  $y$ <sup>1</sup>,  $p(\cdot)$  is the probability density. This is sometimes referred to as MAP <sub>$x, k$</sub>  approach because it simultaneously estimates both  $x$  and  $k$ . Another possible approach will follow.

By Bayes' formula (and assuming  $x$  and  $k$  independent), we have

$$p(x, k|y) = \frac{p(y|x, k)p(x)p(k)}{p(y)}. \quad (2)$$

Let us have a look at the individual terms on the right-hand side.  $p(y|x, k)$  is the distribution of  $y$  for given  $x, k$ . Since  $y - k * x = n$ , it is exactly the distribution of measurement noise. The noise is assumed to follow zero-mean Gaussian distribution, pixel-by-pixel independent

$$p(n) = p(y|x, k) \propto e^{-\frac{1}{2\sigma^2} \sum_i (y_i - (k * x)_i)^2} = e^{-\frac{1}{2\sigma^2} \|y - k * x\|_2^2}$$

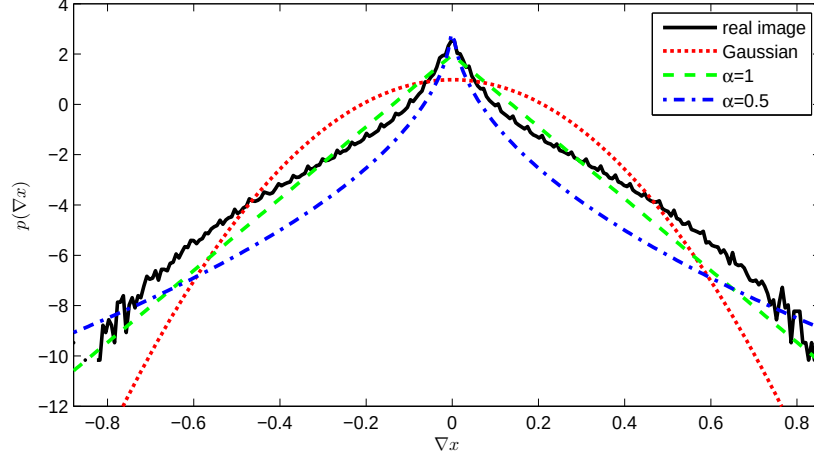
(subscript  $i$  denotes  $i$ -th pixel).

The term  $p(x)$  is the so-called image prior and it states, which sharp images  $x$  are a priori (i.e. independently of the observation  $y$ ) more likely. This key component of MAP is supposed to serve as the missing regularization for our ill-posed problem and drive us towards the right solution. However, as we will see, it often fails to do so. The image prior is always expressed in terms of image derivatives rather than intensity values, because while intensities of different natural photographs can follow very different distributions, the distributions of the image derivatives tend to be similar and so-called heavy-tailed, meaning that high values are far more likely than in case of Gaussian distribution (see Fig. 2). Such heavy-tailed distribution can be approximated by generalized Gaussian distribution with exponent  $\alpha \leq 1$ , as in Levin *et al.* [2011b]. Denoting  $\partial_x x_i$ ,  $\partial_y x_i$  the derivatives of the image  $x$  at pixel  $i$ , the prior  $p(x)$  can be formulated as

$$p(x) \propto e^{-\frac{1}{2\sigma^2} \sum_i |\partial_x x_i|^\alpha + |\partial_y x_i|^\alpha}. \quad (3)$$

---

<sup>1</sup>Estimating the kernel  $k$  is not theoretically necessary but in practice it cannot be avoided as every numerical algorithm requires the estimation of  $k$  in order to compute  $x$ .



**Figure 2.** The distribution of derivatives in a natural image (black) and its approximation by Gaussian (red) and sparse (green, blue) priors.

For  $\alpha < 1$  this prior is called *sparse*, since it favors few high values and zero elsewhere rather than many low nonzero values. With  $\alpha = 2$  this prior is obviously called Gaussian and although rather inaccurate, it is interesting theoretically and numerically because it allows for closed form solutions of some deconvolution approaches. According to Levin *et al.* [2011b], natural images are typically best approximated by  $\alpha \in [0.5, 0.8]$ .

Other priors are possible, for example Fergus *et al.* [2006] used sum of Gaussians, which can be awkward for some deconvolution approaches because after taking the logarithm, the prior becomes the logarithm of sum of product of Gaussians (as opposed to logarithm of product of Gaussians as in (3)), which is difficult to handle. Authors in Shan *et al.* [2008] used piecewise-defined function

$$\phi(x) = \begin{cases} -k|x| & |x| \leq T \\ -(ax^2 + b) & |x| > T \end{cases}$$

and then  $p(x) \propto e^{\phi(x)}$  approximates the natural image prior.

The kernel prior,  $p(k)$  receives far less attention. In the simplest case, it can be considered uniform (i.e. ignored). Some authors (e.g. Shan *et al.* [2008]) use sparse prior in the form

$$p(k) \propto e^{-\tau \sum_i k_i}.$$

It is unclear how this prior really helps the deconvolution process because while it is true that blur kernels are usually sparse, this prior too much favors the delta kernel in the trivial no-blur solution, which we want to avoid. Regardless of the particular prior, other hard constraints are necessary, such as positivity of kernel values and limited support much smaller than the size of the input image. Also, the condition  $\|k\|_1 = 1$  or  $\sum_i k_i = 1$  is typically considered, it means that the blurring process leaves the overall image intensity unchanged.

The denominator term  $p(y)$  serves merely as a normalization constant (for  $y$  given) which can be disregarded. After taking the logarithm, it reduces to an additive constant in the energy which we want to minimize, so it has no effect on the minimizers.

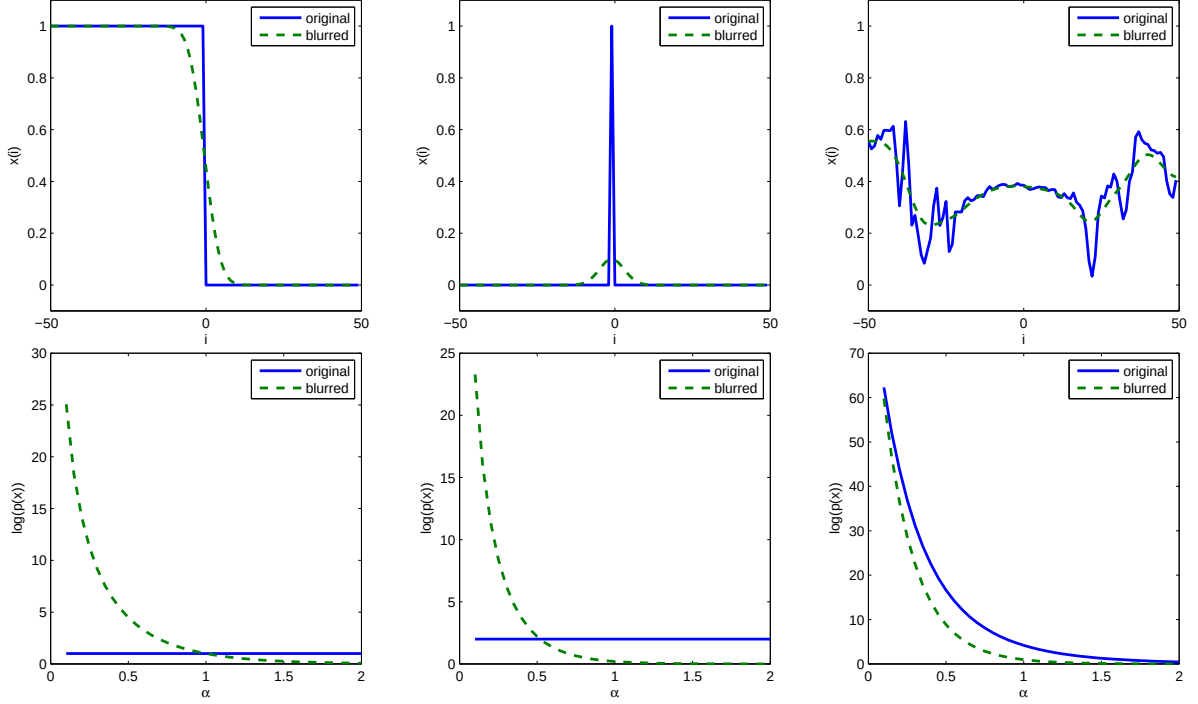
Let us state the overall  $\text{MAP}_{x,k}$  problem (for case of sparse prior (3) and uniform kernel prior). After taking the negative logarithm of (2) we strive to minimize the energy

$$E_{x,k}(x, k) = \lambda \|k * x - y\|_2^2 + \sum_i |\partial_x x_i|^\alpha + |\partial_y x_i|^\alpha \quad (4)$$

with respect to  $(x, k)$  ( $\lambda$  depends on numerous variances and other constants in (2)). In the case of Gaussian image prior, this simplifies to

$$E_2(x, k) = \lambda \|k * x - y\|_2^2 + \|\nabla x\|_2^2,$$

a well-known functional in image restoration (typically considered with  $k$  known, however).



**Figure 3.** Top row: Example 1D signals in solid blue (step edge, spike, and row from natural image) and their blurred versions in dashed green. Bottom row:  $\log(p(x))$  score given by (3) (“the higher the better”) of the signals from above for varying values of  $\alpha$  (degree of sparsity).

### Limitations of $\text{MAP}_{x,k}$

Directly maximizing  $p(x, k)$  (or equivalently minimizing  $E_{x,k}$ ) usually does not work as expected and instead of estimating the correct kernel  $k$  it results in no-blur solution ( $x = y$ ,  $k = \delta$ ). The reason for this is the fact that image prior  $p(x)$ , even if perfectly accurate, very often favors blurred images to sharp ones. Let us demonstrate this on model 1D signals, see Figure 3. The top row consists of three model signals (blue/solid line) – step edge, spike, and row extracted from a natural image. Each of these signals was blurred by discrete Gaussian kernel with  $\sigma = 4$  (green/dashed line). The bottom row shows  $\log(p(x))$  score (negative energy) of these signals given by prior (3) for varying values of  $\alpha$  (i.e. varying degree of sparsity). In case of no measurement noise this is the whole  $\text{MAP}_{x,k}$  score since the first (noise) term in (4) vanishes. We see that for step edge the prior works as expected, i.e.  $\alpha < 1$  the prior correctly gives higher score to the sharp edge while Gaussian prior incorrectly favors blurred edge. For sharp spike (second column) sharp signal wins only for very low  $\alpha$ , in which case the prior starts to behave like a binary penalty (distinguishing only zero/nonzero values). In the third case of natural image, the blurred signal wins everywhere.

The effect of blur on image statistics is twofold: 1) It reduces the sparsity of derivatives which decreases the image MAP score and 2) it reduces the individual values of derivatives which increases the score. In special cases (like ideal step edge) the first effect wins and sparse prior correctly assigns higher score to the sharp image. In most other cases, however, the second effect is stronger and can be compensated only by unreasonably low values of  $\alpha$ , otherwise the blurred image gets higher score. The situation is same for other choices of  $p(x)$  based on gradient distribution, because such prior inevitably prefers low  $\nabla x$ , in which case the blurred image usually wins.

The global minimum of (4) is thus close to the trivial no-blur solution  $(x, k) = (y, \delta)$ , depending on noise intensity. This complicates the deconvolution task because we must alter the minimization procedure such that it finds the local minimum with correct  $(x, k)$ , which is inconvenient because the minimization may easily either fall into the global no-blur minimum or end up in a different local minimum. Several possibilities emerged in the literature. Shan *et al.* [2008] start the optimization with kernel far from the delta solution (user specifies the approximate kernel shape). Jia [2007]; Joshi *et al.* [2008]; Xu and Jia [2010] explicitly detect edges and reweight their contribution (as our analysis shows, sparse priors work much better around step edges than rugged texture with narrow spikes). Shan *et al.* [2008] and others reweight individual energy terms during optimization process (suppressing the data-

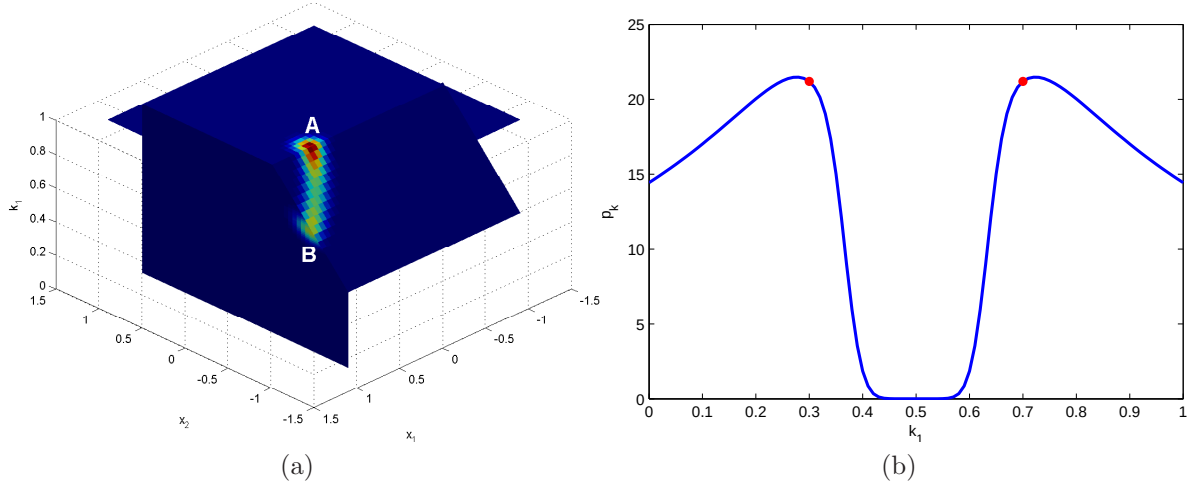
fitting (noise) term early in the optimization produces rough image estimate with strong edges and dense kernel, slowly increasing the weight of the data-fitting term then improves the estimate while hopefully avoiding the delta solution).

### MAP<sub>k</sub> estimation

Another possibility, in the context of image processing first utilized by Miskin and MacKay [2000], Fergus *et al.* [2006] and later further explored by Levin *et al.* [2011b], is the so-called MAP<sub>k</sub> approach. Simply put, instead of estimating  $(x, k)$  simultaneously by maximizing the joint distribution  $p(x, k|y)$  we estimate only the kernel  $k$  by maximizing  $p(k|y)$ , which is  $p(x, k|y)$  marginalized over all images  $x$ :

$$p(k|y) = \int p(x, k|y) dx. \quad (5)$$

The sharp image  $x$  is then estimated using any non-blind deconvolution technique. The important difference here is that this approach dramatically decreases the number of unknowns, because the size of blur kernel is typically much smaller than the size of the image. The price to pay is the obvious difficulty of calculating the integral (5). Let us demonstrate the usefulness of this approach on a minimalistic



**Figure 4.** (a) Values of  $p(x, k|y)$  on plane connecting the no-blur solution (A) and the correct solution (B). (b) Marginalization of  $p(x, k|y)$  over  $x$ , dots indicate correct solution (probability density is not normalized).

1D deconvolution problem. Let  $x = (x_1, x_2)$  be an unknown vector with independent Gaussian prior  $x_i \sim N(0, \sigma^2)$  for each component, similarly let  $n = (n_1, n_2)$  be noise vector  $n_i \sim N(0, \eta^2)$ . We wish to decompose the given two-component vector  $y$  as  $y = k * x + n$  (we used circular convolution in our experiment) for some  $k = (k_1, k_2)$ ,  $k_1 + k_2 = 1$ . The MAP <sub>$x, k$</sub>  probability is as follows

$$p(x, k|y) \propto e^{-\frac{1}{2\eta^2}((x_1 k_1 + x_2 k_2 - y_1)^2 + (x_1 k_2 + x_2 k_1 - y_2)^2)} e^{-\frac{1}{2\sigma^2}(x_1^2 + x_2^2)} \quad (6)$$

Since  $k_1 + k_2 = 1$ , we really have three unknowns,  $x_1$ ,  $x_2$ , and  $k_1$ . The dependance of  $p(x, k|y)$  on  $x_1, x_2, k_1$  is difficult to visualize because it fills a 3D volume, Figure 4a shows a slice through this volume along planes  $k_1 = 1$ ,  $x_1 = x_1^{\text{true}}$ , and plane connecting the delta solution and the desired (true) solution. Inspection of this plot (and underlying data) reveals, that

- The distribution  $p(x, k|y)$  has a strong global maximum in the no-blur solution  $x = y, k = (1, 0)$ .
- The distribution  $p(x, k|y)$  has a weak local maximum close to the true solution (as a particular example, for true solution  $x = (.8, -.8)$ ,  $k = (.7, .3)$  and no noise we have a maximum at  $x = (.55, -.55)$ ,  $k = (.79, .21)$ ).
- These two maxima are connected by a ridge of high values ( $p(x, k|y)$  is negligible elsewhere), so optimization procedure may easily miss the correct local maximum and “fall” to the global maximum because the two are not well separated.

Figure 4b shows the result of marginalization of (6) over  $x$ . The situation has changed dramatically because now we have a nice global maximum away from the delta solution (the presence of two maxima is consequence of the fact that deconvolution cannot distinguish between shifted kernel  $k = (k_2, k_1)$  with appropriately shifted image  $x = (x_2, x_1)$ ). The dots show the position of true solution and we see that the estimation is almost correct (this of course strongly depends on the noise intensity  $\eta$  and how well the values of  $x$  fit the prior  $p(x)$ ). This experiment shows significant advantage of  $\text{MAP}_k$  over  $\text{MAP}_{x,k}$ , which would be even more apparent if we used  $x$  (and  $y$ ) vectors with more components (as in real world images) because while for  $\text{MAP}_{x,k}$  the number of unknowns grows with image size, for  $\text{MAP}_k$  grows only amount of data, which improves the estimation.

The drawback of  $\text{MAP}_k$  approach lies in the apparent difficulty of computing the marginalization integral (5). For Gaussian prior  $p(x)$  it has a closed form solution but in general case, approximations are needed. Several methods appeared in the literature and their analysis is beyond the scope of this article. A brief overview can be found in Levin *et al.* [2011b]. Fergus *et al.* [2006] used the so-called variational Bayes, approximating the joint distribution by the “most fitting” (hence the name variational) product of simpler distributions  $p(x, k|y) = q(x)q(k)$ . Another possibility explained at length in Levin *et al.* [2011a] is quadratic approximation of  $p(x, k|y)$  at its mode (accurate for Gaussian prior, which is quadratic). Interestingly, the resulting energy functional is not altogether different from  $\text{MAP}_{x,k}$  energy functional, it merely contains an additional term which favors kernels with more blur. This relationship suggests that simpler  $\text{MAP}_{x,k}$  approach could benefit from the increased estimation performance of  $\text{MAP}_k$ , if it were altered to contain an additional regularization term penalizing small blurs.

## Conclusion

We have outlined the main building blocks of both the straightforward  $\text{MAP}_{x,k}$  deconvolution and the more sophisticated  $\text{MAP}_k$  deconvolution as they appeared in the recent literature. While (suitably altered) simple  $\text{MAP}_{x,k}$  deconvolution method provides acceptable results, our discussion (and similar analyses from referenced articles) suggest that there is more to be gained from considering the posterior distribution  $p(x, k)$  as a whole instead of just looking at its mode. We therefore believe that more research in this direction will bring not only better computational results but better understanding of image deconvolution in general.

## References

- T. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Society for Industrial and Applied Mathematics, 2005.
- R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *ACM Trans. Graph.*, 25(3):787–794, July 2006.
- J. Jia. Single image motion deblurring using transparency. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.
- N. Joshi, R. Szeliski, and D.J. Kriegman. PSF estimation using sharp edge prediction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.
- N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski. Image deblurring using inertial measurement sensors. *ACM Trans. Graph.*, 29:30:1–30:9, July 2010.
- A. Levin, Y. Weiss, F. Durand, and W.T. Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2657–2664, June 2011.
- A. Levin, Y. Weiss, F. Durand, and W.T. Freeman. Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2354–2367, Dec. 2011.
- J. Miskin and D. J. C. MacKay. Ensemble learning for blind image separation and deconvolution. In M. Girolani, editor, *Advances in Independent Component Analysis*. Springer–Verlag, 2000.
- F. Šroubek and J. Flusser. Multichannel blind iterative image restoration. *IEEE Transactions on Image Processing*, 12(9):1094–1106, Sept. 2003.
- F. Šroubek and J. Flusser. Multichannel blind deconvolution of spatially misaligned images. *IEEE Transactions on Image Processing*, 14(7):874–883, July 2005.
- F. Šroubek and P. Milanfar. Robust multichannel blind deconvolution via fast alternating minimization. *IEEE Transactions on Image Processing*, 21(4):1687–1700, April 2012.
- Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. *ACM Trans. Graph.*, 27(3):73:1–73:10, Aug. 2008.
- L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*, volume 6311 of *Lecture Notes in Computer Science*, pages 157–170. Springer Berlin / Heidelberg, 2010.