Handling noise in image deconvolution with local/non-local priors

Hicham Badri, Hussein Yahia

HANDLING NOISE IN IMAGE DECONVOLUTION WITH LOCAL/NON-LOCAL PRIORS

Hicham Badri[1,2] *and Hussein Yahia*[1]

INRIA Bordeaux Sud-Ouest, France[1]
University Mohammed V Rabat, LRIT Associated Unit to CNRST-URAC 29, Morocco[2]

ABSTRACT

Non-blind deconvolution consists in recovering a sharp latent image from a blurred image with a known kernel. Deconvolved images usually contain unpleasant artifacts due to the ill-posedness of the problem even when the kernel is known. Making use of natural sparse priors has shown to reduce ringing artifacts but handling noise remains limited. On the other hand, non-local priors have shown to give the best results in image denoising. We propose in this paper to combine both local and non-local priors to handle noise. We show that the blur increases the self-similarity within an image and thus makes non-local priors a good choice for denoising blurred images. However, denoising introduces outliers which are not Gaussian and should be well modeled. Experiments show that our method produces a better image reconstruction both visually and empirically compared to methods some popular methods.

***Index Terms*—** Image deconvolution, deblurring, non-local prior, self-similarity, sparsity.

1. INTRODUCTION

Image deblurring consists in reconstructing a true image x from a degraded image y with a kernel k :

$$y - x \otimes k = \eta, \tag{1}$$

where $\otimes$ is the convolution operator that we consider spatially invariant in this paper and η a Gaussian noise. The problem is called *blind deconvolution* [1, 2, 3, 4, 5, 6] if both the kernel and the latent image are unknown, or *non-blind deconvolution* [7, 8, 9, 10, 11] if the kernel is known. Image deconvolution is one of the most common operations in image processing, it is extensively used in computational photography [1, 12, 7]. Unfortunately, several disturbing artifacts are produced even in the non-blind deconvolution case. These artifacts are mainly due to the near-sparsity of the kernel in the frequency domain which produces a large magnitude when performing inverse filtering, resulting in amplified signal and noise [13]. Ringing is the most noticeable artifact which consists in oscillations near edges. Methods such as [7, 8] use natural sparse priors to regularize the problem and have become very popular due to their simplicity. This regularization has shown to reduce the ringing artifacts but cannot efficiently handle high noise levels . On the other hand, non-local priors [14, 15, 16, 17], which take into account the self-similarities within images, have shown to be particularly efficient for image denoising.

We propose in this paper to use both non-local and local priors for image deconvolution. First, we show that the blur increases the self-similarities within images, which makes low-rank estimation a favorable choice to denoise blurred images. However, after denoising, the blur model is no longer Gaussian. We show that the denoising operation introduces Laplacian outliers. We propose a MAP formulation with the l_1-norm on the data term to handle outliers and a $\log(l_p)$ prior for the derivatives. The $\log(l_p)$ function promotes more sparsity than the l_p norm and helps to get rid of artifacts caused by the outliers. We show that the proposed method improves the image reconstruction quality.

2. PROBLEM FORMULATION

In the MAP optimization framework, image deconvolution can be formulated using the Bayes rule :

$$p(x|y) \propto p(y|x)p(x). \tag{2}$$

The likelihood term $p(y|x)$ is based on the convolutional model (1). The noise η is modeled as a set of independent and identically distributed random variables for all pixels, each of which follows a Gaussian distribution. The prior $p(x)$ regularizes the problem and is usually a sparse prior modeling the natural distribution of the derivatives of natural images [7, 8]. The model that we propose introduces an intermediate denoised image y' to be deconvolved :

$$p(x|y') \propto p(y'|x)p(x). \tag{3}$$

A non-local prior is used to produce the image y' while a local prior is used to reconstruct the final image x given the intermediate image y'. Note that this time, the likelihood $p(y'|x)$ is different from the one of model (2). We will see that this likelihood is in fact Laplacian and no longer Gaussian. The proposed method consists in two steps : 1) estimating the variable y' given the blurred and noisy image y, 2) estimating the latent image x given both y' and k.

This work is supported by an INRIA (Direction of Research) CORDI-S grant.

2.1. Step 1 : Estimating the Blurred Noise-Free Image y'

The first step of the method consists in denoising the blurred and noisy image y to produce a new image y', which is blurred but noise-free. In order to perform this task, we might use a method which takes into account the self-similarities within the blurred image, without using any specific prior to images as blurred images are hard to model. Given a set of similar patches[1], we want to extract their noise-free versions using only the self-similarities within this group of patches. One way to think about it is to study the singular values ; a group of similar patches stacked in one matrix $\mathbf{Y}$ should minimize the rank of $\mathbf{Y}$. As the rank is a difficult measure to deal with, we rather use the nuclear norm which is the convex relaxation of the rank ($||Y||^* = \sum_{i=1}^{K} \lambda_i$, where λ_i are the singular). **Fig. 1** shows the singular values distribution of all the matrices $\mathbf{Y}$ (using overlapping patches) within a natural image (in black) and within its blurred version (in red) with a real camera shake kernel. As can be seen, the distribution of the singular values of the matrices $\mathbf{Y}$ in the blurred image tends to be more compact and sparser compared to the natural image in black. This means that the patches within a self-similar group are much more similar in a blurred image (the distribution would be Dirac-like in a blank image). As the noise increases the rank of the matrices $\mathbf{Y}$, using this self-similarity as a prior can help recovering the blurred image from its blurred and noisy version without using any natural prior specific to blurred images. Denoising becomes a low-rank approximation problem :

$$\text{minimize } ||\mathbf{Y}'||_* + \tau||\mathbf{E}||_2^2 \text{ s.t. } \mathbf{Y} = \mathbf{Y}' + \mathbf{E}, \qquad (4)$$

where $\mathbf{Y}'$ is a matrix containing patches of the image y' and the matrix $\mathbf{E}$ represents the noise. Low-rank approximation problems can be solved with iterative singular value thresholding such as the methods in [18]. The operation is repeated for each group of similar patches and the final image y' is reconstructed by averaging overlapping patches. A similar low-rank approach was recently used to perform restoration of natural images in [17], where the iterative regularization is performed on the whole image and not on the patch matrices. In practice, both methods have shown to work well when used with the proposed MAP framework.

2.2. Step 2 : Estimating the Deblurred Image x

Once the blurred but noise-free image y' is estimated, we want to recover the natural image x, given the kernel k. We formulate the problem as a MAP estimation :

$$p(x|y') \propto p(y'|x)p(x). \qquad (5)$$

In order to use this model, we need to model both :
1) The distribution of the residual error $y' - x \otimes k$.
2) The prior information about the natural image x.

[1] Similar patches can be found using a simple Block Matching as used in various non-local methods.

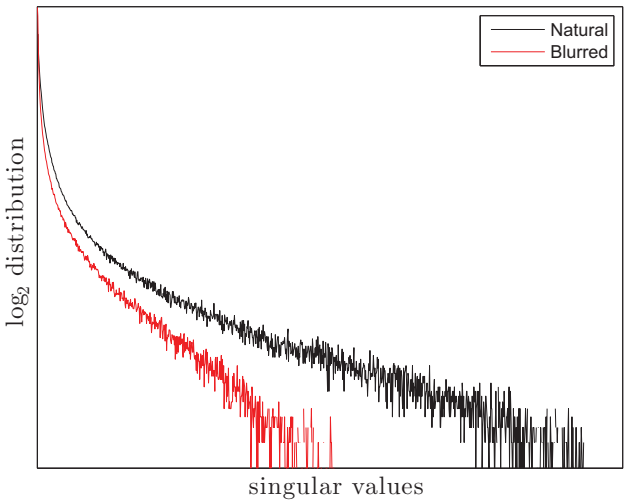


Fig. 1: Distribution of the singular values of the matrices containing similar patches in a natural image (in black) and its blurred version (red). The self-similarity in the blurred image is much more present compared to its natural version.

2.2.1. *Modeling the residual error* $p(y'|x)$

Fig 2. shows the distribution of the residual error between the denoised image y' and the convolution $x \otimes k$. As can be seen, the residual error is no longer Gaussian as it was in the blur model (1) but rather Laplacian. This is due to the fact that denoising a blurred image reduces the noise but introduces outliers as the denoising operation does not take into account the blur kernel. In order to deal with these outliers, we use the l_1-norm on the data term ($p(y'|x) \propto e^{-\gamma|x \otimes k - y'|}$). It can be shown also empirically that the residual of the first and the second derivatives follow a hyper-Laplacian law ($p(\partial y'|\partial x) \propto e^{-\gamma|\partial x \otimes k - \partial y'|^{\alpha<1}}$). However, using this prior together with the sparse natural prior $p(x)$ requires solving a problem with multiple sparse functions which is harder to solve. Instead, for the derivatives residual error, we use a Gaussian prior instead of the hyper-Laplacian for the sake of simplicity.

2.2.2. *Modeling the prior* $p(x)$

Due to the ill-posedness of the deconvolution operation, a prior on the latent image x is necessary. This prior is usually considered on the derivatives and modeled using a hyper-Laplacian law [7, 8]. However, due to the outliers in our case, we found that a norm promoting more sparsity than the l_p-norm is required to get rid of some artifacts while preserving fine structures. To do so, we use the $\log(l_p)$ norm. The $\log$ is a concave function that helps promoting more sparsity when used as a surrogate function on the l_p-norm. On the other hand, solving efficiently a problem using the $\log(l_p)$ norm is more challenging. We will see in this paper how to use this function via a simple proximal method.

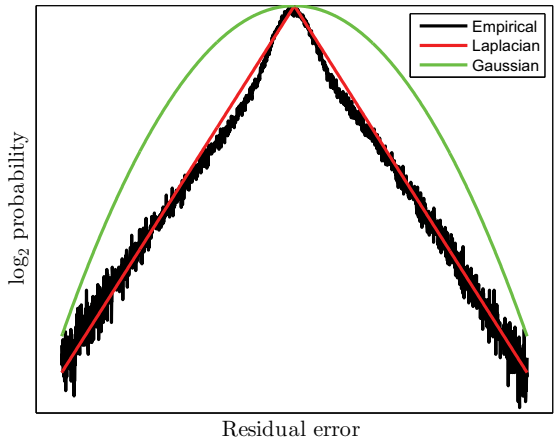


Fig. 2: The Laplacian distribution models well the residual error between the image y' and the convolution $x \otimes k$.

2.2.3. *Optimization*

The deconvolution method that we propose consists in solving the following optimization problem :

$$\begin{aligned}\underset{x}{\text{argmin}} \sum_i^N \gamma|(x \otimes k)_i - y_i^{'}| + \sum_t^T \rho_t \log(|(\partial_t x)_i|^p + \epsilon) \\ + \sum_s^S \frac{\alpha_s}{2} \left((\partial_s x \otimes k)_i - (\partial_s y')_i\right)^2,\end{aligned} \tag{6}$$

where γ, ρ_t, and α_s are positive regularization terms. The first term consists in the l_1 norm to handle outliers due to the denoising step, the second term consists in the sparse prior and the third term models the residual error of the derivatives. As we said before : 1) we use the $\log(l_p)$ norm which promotes more sparsity in order to get rid of some artifacts due to the presence of outliers, 2) we use the l_2 norm instead of the l_p norm to model the derivatives residual error for the sake of simplicity. While the l_2 norm does not fit the real model of this residual, we found that it slightly improves the deconvolution result. As the problem is non-convex, we use a half-quadratic approach [19] two times :

$$\begin{aligned}&(\mathbf{p_1}) : q^{(l_1+1)} \leftarrow \underset{q}{\text{argmin}} \sum_i^N |q_i| + \frac{\beta_1}{2} \left((x^{(l_1)} \otimes k - y')_i - q_i\right)^2 \\ &(\mathbf{p_2}) : x^{(l_1+1)} \leftarrow \underset{x}{\text{argmin}} \sum_i^N \frac{\gamma\beta_1}{2} \left((x \otimes k)_i - (y_i^{'} + q_i^{(l_1+1)})\right)^2 + \\ &\sum_{t=1}^T \rho_t \log(|(\partial_t x)_i|^p + \epsilon) + \sum_{s=1}^S \frac{\alpha_s}{2} \partial_s \Big((x \otimes k)_i - (y_i' + q_i^{(l_1+1)})\Big)^2,\end{aligned} \tag{7}$$

where β_1 is a new regularization term. Then problem $(\mathbf{p_2})$ is split into two sub-problems :

$$\begin{aligned}&(\mathbf{p_{21}}) : v_t^{(l_2+1)} \leftarrow \underset{v_t}{\text{argmin}} \sum_i^N \log(|v_{t,i}|^p + \epsilon) + \frac{\beta_2}{2} \left(\partial_t x_i^{(l_2)} - v_{t,i}\right)^2 \\ &(\mathbf{p_{22}}) : x^{(l_2+1)} \leftarrow \underset{x}{\text{argmin}} \sum_i^N \gamma\beta_1 \left((x \otimes k)_i - (y_i^{'} + q_i^{(l_1+1)})\right)^2 + \\ &\sum_{t=1}^T \rho_t \beta_2 \left(\partial_t x_i - v_{t,i}^{(l_2+1)}\right)^2 + \\ &\sum_{s=1}^S \alpha_s \partial_t \left((x \otimes k)_i - (y_i' + q_i^{(l_1+1)})\right)^2,\end{aligned} \tag{8}$$

where β_2 is another regularization term. Problems $(\mathbf{p_1})$ and $(\mathbf{p_{21}})$ are in the proximal form. The solution of the l_1 norm in the proximal form is known to be a soft-thresholding operation, thus the solution to problem $(\mathbf{p_1})$ is given as follows

$$q^{(l_1+1)} = \max\left\{0, |(x^{(l_1)} \otimes k - y')| - \frac{1}{\beta_1}\right\} \text{sign}(x^{(l_1)} \otimes k - y'). \tag{9}$$

However solving problem $(\mathbf{p_{21}})$ is not straightforward due to the non-convexity of the proximal operator. We use a first order approximation to estimate a solution. As the minimization of v_t is in the proximal form :

$$\text{prox}_{bh}(a) = \underset{v_t}{\text{argmin}}\, h(v_t) + \frac{1}{2b}||v_t - a||_2^2, \tag{10}$$

it can be shown that the first order approximation is given as follows [20, 21] :

$$\text{prox}_{bh}(a) \approx a - b\nabla h(a). \tag{11}$$

As the shrinkage operator influences only the magnitude, we rewrite equation (11) as follows :

$$\text{prox}_{bh}(a) = \max(0, |a| - b\nabla h(|a|))\text{sign}(a). \tag{12}$$

Now setting $h(|a|) = \log(|a|^p + \epsilon)$ and replacing $a = \partial_t x^{(k)}$, $b = \frac{1}{\beta_2}$, the solution $v_t^{(l_2+1)}$ is given pointwise as follows :

$$v_t^{(l_2+1)} = \max\left(0, |\partial_t x^{(l_2)}| - \frac{1}{\beta_2} \frac{p|\partial_t x^{(l_2)}|^{p-1}}{|\partial_t x^{(l_2)}|^p + \epsilon}\right) \text{sign}(\partial_t x^{(l_2)}), \tag{13}$$

where ϵ is a small parameter set to 0.001 to offer stability. The remaining problem $(\mathbf{p_{22}})$ is quadratic and easy to solve using the Fourier transform $\mathcal{F}$:

$$\begin{aligned}x^{(l_2+1)} &= \mathcal{F}^{-1}\left\{\frac{\mathcal{F}(y' + q^{(l_1+1)}) \circ \Gamma_1 + \sum_{t=1}^T \Gamma_2 \circ \mathcal{F}(v_t^{(l_1+1)})}{\mathcal{F}(k) \circ \Gamma_1 + \sum_{t=1}^T \Gamma_2 \circ \mathcal{F}(\partial_t)}\right\} \\ \Gamma_1 &= \overline{\mathcal{F}(k)} \circ \left(\gamma\beta_1 + \sum_{s=1}^S \alpha_s \overline{\mathcal{F}(\partial_s)} \mathcal{F}(\partial_s)\right) \\ \Gamma_2 &= \rho_t \beta_2 \overline{\mathcal{F}(\partial_t)},\end{aligned} \tag{14}$$

where $\circ$ denotes a pointwise multiplication. In all the experiments, we do not use more than 13 inner iterations of $(\mathbf{p_2})$ and 2 outer iterations of $(\mathbf{p_1})$.

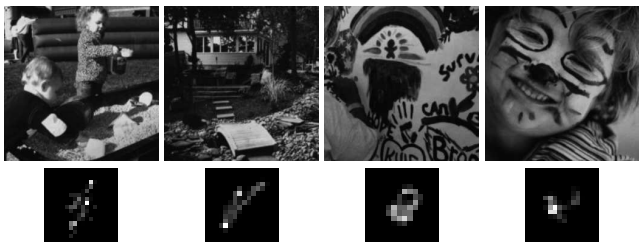


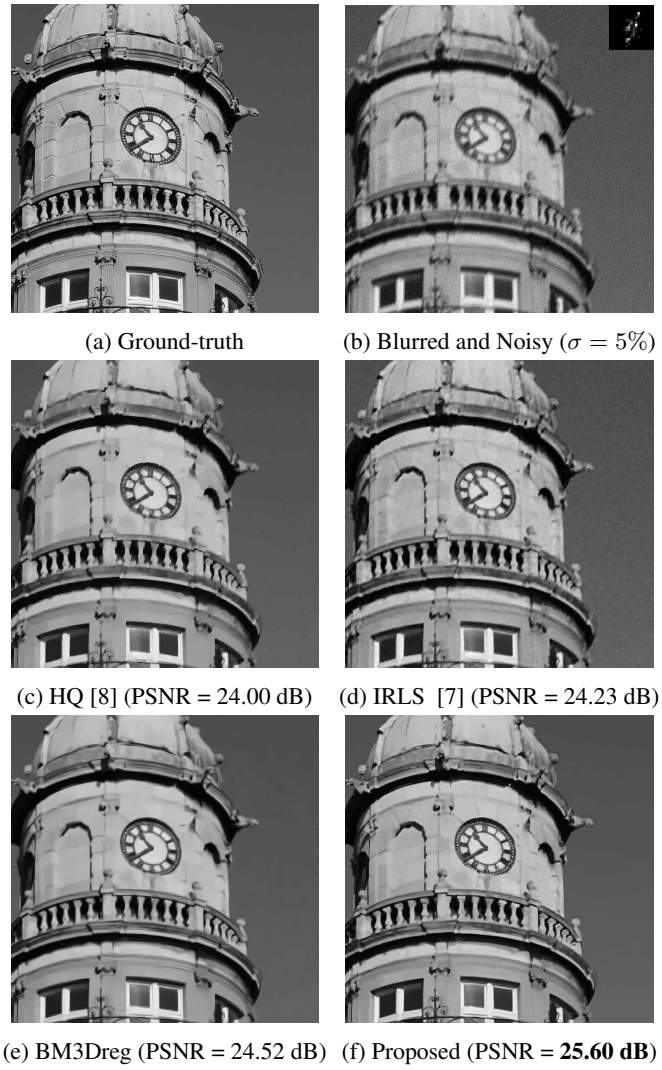
Fig. 3: Dataset from [2] used in our experiments with 5 noise levels, resulting in 80 test images.

3. EXPERIMENTS

In this section, we evaluate the proposed method and compare it with three methods : two popular methods that use a hyper-Laplacian sparse gradient prior only [7, 8], and a third method that uses BM3D [15] regularization instead of the local sparse gradient prior, similar to [9, 22]. We use 4 images and 4 real-world blur kernels from the standard benchmark dataset of [2] (see **Fig. 3**) for 5 synthetic Gaussian noise levels $\sigma = 0.25, 0.5, 1, 2, 5\%$, resulting in a total of 80 test images. The PSNR and SSIM [23] of the experiments are presented in **Table 1**. As can be seen, the proposed method performs well in both low and high level noise situations in terms of the PSNR and SSIM. The improvement is most noticeable for high noise levels. It is worth noting that even though the IRLS method [7] leads to higher PSNR and SSIM compared to the HQ approach of [8], it does not necessarily lead to a better visual quality. In order to show the visual quality of the proposed method, we run experiments on a 800×800 real-world image blurred with a 19×19 real camera shake kernel (both from [8]). The results are presented in **Fig. 4**. As can be seen, the proposed method (f) leads to a better reconstruction result compared to the three other methods (c), (d), (e). The BM3D regularization (e) seems to perform much better than the sparse gradient prior methods (c) and (d), but introduces oversmoothing and some visible artifacts. Note that BM3D regularization is relatively slower compared to simple sparse gradient regularization. In the latter case, the processing consists in simple pixelwise gradient shrinkage operations. Our method performs denoising only once and benefits from the computational efficiency of sparse gradient methods. As we perform only two outer iterations of problem $(\mathbf{p_2})$, our MAP deconvolution method is only around two times slower than [8]. The low-rank denoising step can be parallelized as denoising the groups is independent.

	Method	HQ [8]	IRLS [7]	BM3Dreg [15]	Proposed
$\sigma = 0.25\%$	PSNR	34.10 dB	34.91 dB	35.02 dB	**35.33 dB**
	SSIM	0.953	0.958	0.959	**0.963**
$\sigma = 0.5\%$	PSNR	32.72 dB	33.71 dB	34.02 dB	**34.32dB**
	SSIM	0.932	0.941	0.942	**0.947**
$\sigma = 1\%$	PSNR	30.70 dB	32.03 dB	32.18 dB	**32.83 dB**
	SSIM	0.895	0.907	0.909	**0.920**
$\sigma = 2\%$	PSNR	28.82 dB	29.45 dB	29.70 dB	**30.62 dB**
	SSIM	0.845	0.852	0.862	**0.881**
$\sigma = 5\%$	PSNR	26.11 dB	26.68 dB	26.88 dB	**27.85 dB**
	SSIM	0.752	0.773	0.777	**0.805**
Mean	PSNR	30.49 dB	31.36 dB	31.56 dB	**32.19 dB**
	SSIM	0.875	0.886	0.890	**0.903**

Table 1: Experiments results conducted on the dataset **Fig. 3**. The proposed method performs well for both low and high noise levels.



(a) Ground-truth (b) Blurred and Noisy ($\sigma = 5\%$)

(c) HQ [8] (PSNR = 24.00 dB) (d) IRLS [7] (PSNR = 24.23 dB)

(e) BM3Dreg (PSNR = 24.52 dB) (f) Proposed (PSNR = **25.60 dB**)

Fig. 4: Various deconvolution results in the case of high noise level ($\sigma = 5\%$). The proposed method produces a better reconstruction with less visible artifacts.

4. CONCLUSION

We present a new noise-aware non-blind deconvolution algorithm. The new approach consists in combining non-local and local priors. We show that the blur increases the self-similarities within images which makes low-rank approximation a good tool to perform denoising of blurred images. However, denoising a blurred image introduces outliers in the convolution model. We show that the residual error after denoising is no longer Gaussian but rather Laplacian. We present a new MAP deconvolution formulation which takes into account these outliers together with an efficient solver based on two half-quadratic splits. Experiments show that the proposed method produces a better image reconstruction both visually and empirically while benefiting from the computational efficiency of sparse gradient methods.

5. REFERENCES

[1] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, July 2006.

[2] Anat Levin, Yair Weiss, Frédo Durand, and William T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, 2011.

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

[4] Anat Levin, Yair Weiss, Frédo Durand, and William T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *CVPR*, 2011, pp. 2657–2664.

[5] Dilip Krishnan, Terence Tay, and Rob Fergus, "Blind deconvolution using a normalized sparsity measure," in *CVPR*, 2011, pp. 233–240.

[6] Li Xu, Shicheng Zheng, and Jiaya Jia, "Unnatural l0 sparse representation for natural image deblurring," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2013, CVPR '13, pp. 1107–1114, IEEE Computer Society.

[7] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, July 2007.

[8] Dilip Krishnan and Rob Fergus, "Fast image deconvolution using hyper-laplacian priors," in *NIPS*, 2009, pp. 1033–1041.

[9] Aram Danielyan, Vladimir Katkovnik, and Karen Egiazarian, "Bm3d frames and variational image deblurring," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1715–1728, 2012.

[10] Uwe Schmidt, Carsten Rother, Sebastian Nowozin, Jeremy Jancsary, and Stefan Roth, "Discriminative non-blind deblurring," in *CVPR*, 2013, pp. 604–611.

[11] Uwe Schmidt and Stefan Roth, "Shrinkage fields for effective image restoration," in *CVPR*, 2014, to appear.

[12] Ramesh Raskar, Amit Agrawal, and Jack Tumblin, "Coded exposure photography: Motion deblurring using fluttered shutter," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 795–804, July 2006.

[13] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum, "Progressive inter-scale and intra-scale non-blind image deconvolution," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 74:1–74:10, Aug. 2008.

[14] Antoni Buades and Bartomeu Coll, "A non-local algorithm for image denoising," in *In CVPR*, 2005, pp. 60–65.

[15] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen O. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[16] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman, "Non-local sparse models for image restoration," in *ICCV*, 2009, pp. 2272–2279.

[17] Weisheng Dong, Guangming Shi, and Xin Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 700–711, 2013.

[18] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.

[19] Donald Geman and Chengda Yang, "Nonlinear image recovery with half-quadratic regularization.," *IEEE Transactions on Image Processing*, vol. 5, no. 7, pp. 932–946, 1995.

[20] Neal Parikh and Stephen Boyd, "Proximal algorithms," *Foundations and Trends in Optimization, 2013.*

[21] Hicham Badri, Hussein Yahia, and Driss Aboutajdine, "Fast multi-scale detail decomposition via accelerated iterative shrinkage," in *SIGGRAPH Asia 2013 Technical Briefs*, New York, NY, USA, 2013, SA '13, pp. 33:1–33:4, ACM.

[22] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen O. Egiazarian, "Image restoration by sparse 3d transform-domain collaborative filtering," in *Image Processing: Algorithms and Systems*, 2008, p. 681207.

[23] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.