Image Priors for Image Deblurring with Uncertain Blur

Daniele Perrone[1]
perrone@iam.unibe.ch

Avinash Ravichandran[2]
http://cis.jhu.edu/~avinash/

René Vidal[3]
http://cis.jhu.edu/~rvidal/

Paolo Favaro[1]
paolo.favaro@iam.unibe.ch

[1] Universität Bern,
Bern, Switzerland

[2] UCLA VisionLab
University of California,
Los Angeles, CA, USA

[3] Center for Imaging Science
Johns Hopkins University,
Baltimore, MD, USA

Abstract

We consider the problem of non-blind deconvolution of images corrupted by a blur that is not accurately known. We propose a method that exploits dictionary-based image priors and non Gaussian noise models to improve deblurring accuracy in the presence of an inexact blur. The proposed image priors express each image patch as a linear combination of atoms from a dictionary learned from patches extracted from the same image or from an image database. When applied to blurred images, this model imposes that patches that are similar in the blurred image retain the same similarity when deblurred. We perform image deblurring by imposing this prior model in an energy minimization scheme that also deals with outliers. Experimental results on publicly available databases show that our approach is able to remove artifacts such as oscillations, which are often introduced during the deblurring process when the correct blur is not known.

1 Introduction

Image deblurring is the problem of recovering a sharp image from a blurred one. This problem recurs often in photography due to camera shake or long exposures with moving objects. When only the blurry image is given, one needs to solve a blind-deconvolution problem and recover not only a sharp image, but also a characterisation of the blur. A common scheme for most blind-deconvolution algorithms [2, 7, 13, 24] is to alternate between estimating the blur given the sharp image and estimating the sharp image given the blur. In particular, the latter step is called non-blind deconvolution. A common choice for non-blind deconvolution algorithms is to use methods that rely on an error-free blur estimate. However, small errors in the blur estimate result in visible artifacts in the restored image, which may not be removed by future iterations (see Fig. 1).

A few methods have considered uncertainty in the blur estimate. These methods typically rely on the use of robust norms in the data term [23, 24], or detect outliers by using an iterative approach [4]. However, their main limitation is that they use an image prior that
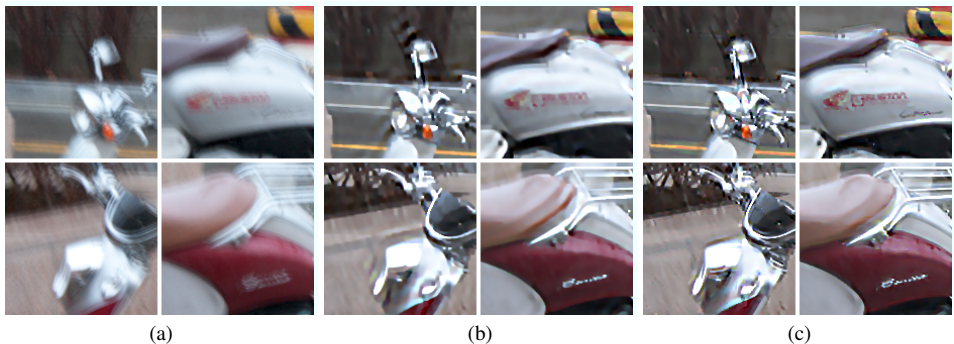
Figure 1: a) 2x2 selected areas from a real blurred image; b) Deblurring results with the algorithm of Cho *et al.* [5]; c) deblurring results with our algorithm.

does not discourage the presence of artifacts in the restored images (see Fig. 1). For example, priors learned from dictionary-based methods [5] by alternating between the dictionary learning step and the deblurring step, may build dictionaries that incorporate artifacts with high gradients (*e.g.*, piecewise constant oscillations).

In this work we propose a novel image prior to remove artifacts introduced by blur errors. To achieve this goal we also use a dictionary-based prior learned only from the input blurred image and a database of images. However, we propose a method to prune ambiguities in the prior due to blur (see Section 4.1). We will show experimentally how our method effectively improves the accuracy of the restored image and visibly reduces artifacts due to blur errors.

2 Prior Work

Blind and non-blind deconvolution are highly ill-conditioned problems: Small errors in the blur estimate often produce noticeable artifacts in the reconstructed image. To overcome this issue, the typical approach is to restrict the class of possible solutions by using regularization [20]. The most prominent regularization methods capture and exploit prior knowledge about the reconstructed image. Total Variation (TV) has been among the most successful image priors and it has been used with good results in many deblurring and denoising algorithms [7, 10, 17, 18, 24]. TV is based on the idea that natural images are usually piecewise constant, an assumption that follows from recent work on image statistics [19]. However, since TV is a global prior, it does not account for the fact that sub-regions in an image can have different statistics [4], *e.g.*, the texture of a tree is different from that of a house.

To overcome the limitations of TV, some methods relax the global prior assumption and impose that texture statistics should change smoothly across an image [18, 21]. Some methods use wavelet bases that are specifically chosen to represent a natural image [13, 16]. However, such methods strongly rely on the chosen wavelet basis for successfully restoring the images. Other methods instead use dictionaries to represent local statistics. Dictionary-based methods have become popular for denoising [5, 14, 15], and deblurring [6, 12]. These methods avoid choosing a predefined basis set, but instead learn a dictionary (usually over-complete) from a dataset of images. Then, each image patch is denoised or deblurred by expressing it as a linear combination of patches from the dictionary. Methods differ in the choices for learning the dictionary and determining the linear combination coefficients. In general, local priors considerably enhance the accuracy of the reconstructed images. How-

ever, the prior imposed by such methods is limited to the statistics found in a small neigh-borhood of a pixel. To deal with the limitations of local priors, non-local approaches can be used [1, 5, 15]. These methods determine correspondences between pixels with a simi-lar local appearance in large areas, and not just in a neighborhood of a pixel. While these methods can greatly improve the quality of the restored images, when they are applied to noisy and blurry patches they produce artifacts or are less effective due to ambiguities in the correspondences.

Our method combines dictionary-based and non-local approaches. We address image deblurring with an uncertain blur by building a dictionary of patches and by finding the non-local correspondences directly on the blurry image. To deal with the errors in the non-local correspondences, we exploit the partial knowledge of the blur. We propose a consensus strat-egy to separate correspondences due to genuine correspondences in the sharp image from correspondences due to replicas of the same patch generated by blur (see Sections 3 and 4).

3 Image Deblurring

Let g be an observed degraded image

$$g = k * f + n, \tag{1}$$

where $*$ is the convolution operator, k is a blur kernel, or point spread function (PSF), f is the noiseless and sharp image, and n is additive noise generated during the acquisition process. The aim of image deblurring is to estimate the noiseless image f given the noisy image g and the kernel k. What makes this problem challenging is the fact that even with no noise, eq. (1) is typically not invertible and is satisfied by infinitely many solutions. Furthermore, in most practical cases the blur k is known only up to some error.

To address these challenges, we introduce priors on the image f. We begin by writing the imaging model (1) as a matrix-vector operation. To do so, we rearrange the pixels in the sharp image f as a column vector $\vec{f}$. Similarly, we rearrange the pixels of the blurred image g and the noise image n as vectors $\vec{g}$ and $\vec{n}$, respectively. Then, we can write eq. (1) as

$$\vec{g} = K\vec{f} + \vec{n}, \tag{2}$$

where K is a matrix operator performing the same discrete convolution as the blur kernel k. Eq. (2) is a linear system where we need to recover $\vec{f}$ given $\vec{g}$ and K. Because of the noise n the system cannot be solved by a simple matrix inversion. One approach is to ignore noise in correspondence of large singular values of K, and to discard equations otherwise. This is equivalent to a rank deficient linear system in $\tilde{K}$ with no noise, which is known to have infinitely many solutions.

One way to obtain a unique solution is to introduce additional linear equations, which we call *image priors*, via a matrix A and a vector $\vec{b}$

$$A\vec{f} = \vec{b} \tag{3}$$

such that the stacked matrix $[\tilde{K}^T \ A^T]^T$ has rank equal to the length of $\vec{f}$. The matrix A encodes our prior knowledge of what sharp textures look like. For instance, images typically enjoy some spatial regularity due to the finite extension of smooth objects, which can be imposed by approximating the gradient operator with A as the finite difference operator.

To enforce this regularity, we consider applying the same linear constraints to all pixels in patches of $L \times L$ pixels. For this purpose, we extract patches of $L \times L$ pixels centered at each pixel of the image f, rearrange the pixel intensities of each patch as a column vector and collect all such vectors into a matrix $F \in \mathbf{R}^{L^2 \times N}$, where N is the number of pixels in f. We can then write our prior as

$$F = DC, \tag{4}$$

where the columns of $D \in \mathbf{R}^{L^2 \times M}$ represent patches (dictionary) and the columns of $C \in \mathbf{R}^{M \times N}$, which we call the *correspondence matrix*, specify the weights used to express a patch in F as a linear combination of the patches in D. In our formulation the design of the image prior is done by choosing D first and then *learning C*.

Denoising methods such as [14] use a dictionary learnt from sharp patches of natural images ($D = D_0$). One then represents each patch in the given image using this dictionary with the constraint that the coefficients of the linear combination must be sparse. One drawback of this approach is that the dictionary is typically optimized to perform well on average and not on the specific input image, therefore we do not use this case in our approach. Other methods such as non-local means [1] use the model with $D = F$, and define the entries of C via a kernel between the patches of the image itself. The method we use to learn the matrix C is inspired by this approach. However, we extend it to the case where the dictionary is a mix of both an external dictionary D_0 and the image itself, *i.e.*, $D = [D_0 \ F]$.

4   Learning the Correspondence Matrix C

One of the most important components of the proposed model is the correspondence matrix C, which needs to be learned from data according to the model chosen in Section 3. In all the cases, learning C can be achieved by solving the equation $F = DC$ with respect to C for some given F and D. To learn C, we face two important challenges: The first challenge is that F is typically not available and the second is that we do not have enough equations to obtain a unique matrix C.

To deal with the first challenge, we extract noisy and blurred patches G_i from the image $g = k * f + n$. Let B be the matrix of patches extracted from the blurred noiseless image $b = k * f$. Since $B = KF = KDC$, we can express the blurred patches in B in terms of the blurred dictionary $E \triangleq KD$ using the *same* correspondence matrix C. When we use the image itself as a dictionary, the blurred dictionary is given by the blurred image patches, *i.e.*, $E = KF = B$. Since B is unknown (unless the image is noiseless), we approximate the blurred patches in B by the blurred and noisy patches in G, *i.e.*, we choose $E = G$. When we use a dictionary composed of both blurry image patches and a blurred dictionary, we will have $E = [G \ E_0]$, where $E_0 \triangleq KD_0$. The basic idea here is that the linear combination of an over-complete basis that yields a sharp image is the same as the one that yields a blurred version of the image with respect to the blurred basis. Lou et al. [12] have introduced this model, but only with $E = E_0$.

The second challenge, *i.e.*, the non uniqueness of the matrix C is due to the overcompleteness of the dictionary D. We introduce additional constraints on C by exploiting image self-similarities. As in [1], we consider a patch at pixel i, B_i, found as a weighted average of similar patches D_j extracted from either the same image or from a dictionary of patches.

Table 1: Correspondences accuracy from blurred images.

	false positives	false negatives
Non-Local means on blurred images	38.1%	8.8%
Correlation-based correction	20.6%	18.5%

Specifically, if we consider all the patches as vectors in $\mathbf{R}^{L^2}$, then we have

$$B_i = \sum_{j=1}^{M} D_j \frac{\phi(D_i, D_j)}{\sum_{\ell=1}^{M} \phi(D_i, D_\ell)}, \quad i = 1, \dots, N, \tag{5}$$

where ϕ is a positive semi-definite kernel that measures the similarity between two patches. In our work we use the following kernel,

$$\phi(D_i, D_j) = \begin{cases} 1 & \|D_i - D_j\|_2^2 \leq \varepsilon^2 \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where ε is proportional to the standard deviation of noise. It is easy to see that (5) can be written in matrix form as

$$B = DC^{nlm}, \tag{7}$$

where D is a matrix of patches ($D = [G \; E_0]$ in our case) and $\{C^{nlm}\}_{j,i} = \frac{\phi(D_i, D_j)}{\sum_{\ell=1}^{M} \phi(D_i, D_\ell)}$ is the correspondence matrix obtained from the procedure in eq. (5).

Given a noisy patch G, where the noise is i.i.d. Gaussian, we know that the denoised patch will lie at the center of a hypersphere and that the noisy patch will lie on its hypersurface. Hence, we can take the previous estimate $B^{nlm} = DC^{nlm}$ and project it on the center of the hypersphere as

$$B^{anlm} = L\sigma \frac{B^{nlm} - G}{\|B^{nlm} - G\|} + G \tag{8}$$

where B^{anlm} is the final adjusted estimate, and σ is the standard deviation of noise. This operation can also be written directly in terms of the images as

$$b^{anlm} = L\sigma \frac{b^{nlm} - g}{\sqrt{(b^{nlm} - g)^2 * w}} + g, \tag{9}$$

where b^{nlm} is the filtered image, g is the input noisy and blurry image, and w is a unit-normalized kernel that defines a window of the size of an $L \times L$ patch. As it is easy to see, the adjustment method adds very little to the computational load of non-local means.

4.1 Self-Similarity in Blurred Images

The non-local means procedure allows us to build a matrix C by finding patches that are self-similar via eq. (6). However, when we apply this procedure to a blurred image, incorrect correspondences may be generated. We distinguish two types: *false negatives*, i.e., correspondences found in the sharp image, but not in the blurred one, and *false positives*, i.e., correspondences present in the burred image, but not in the sharp one. In the first row of Table 1 we show the percentages of false positives and false negatives found while learning the
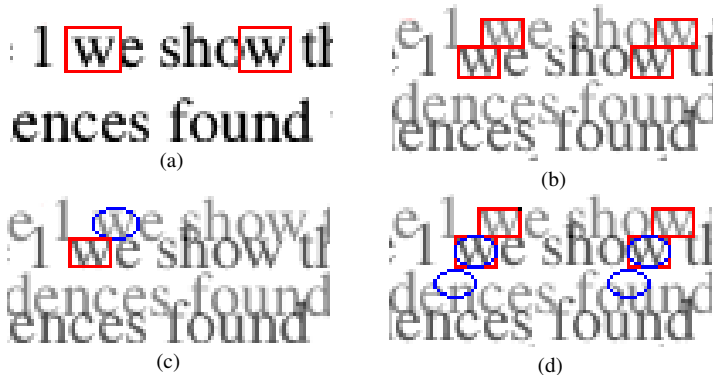
Figure 2: **Example of ambiguous correspondence correction.** In this toy example we show the correction performed by the correlation-based method. The PSF consists of only two peaks, which result in an overlap of two copies of the sharp image with two relative shifts. **(a)** Sharp image of text, and two correct correspondences (red squares) of the character 'w'. **(b)** Blurred version of the previous image which leads to additional incorrect correspondences. **(c)** Selected patch (red square) and the area corresponding to the second peak of the PSF (blue circle). **(d)** The correlated patches shown in (a) are obtained by overlapping the correspondence sets of the two patches (blue circles and red squares).

matrix C from blurred images via non-local means. As one can see there is a relatively low number of correspondences that have been missed (false negatives), but a very high number of additional incorrect correspondences (false positives). While false negatives have the effect of reducing the amount of regularization that could be imposed during the reconstruction to counterbalance the artifacts due to the blur uncertainty, false positives directly introduce artifacts in the reconstruction. Hence, in this section we describe a consensus technique to drastically reduce the number of false positives while keeping the false negatives small. Notice that this aspect has not been addressed by prior patch-based approaches.

In Fig. 2 we provide a synthetic example to illustrate why blur generates false positives. On the top-left (a) we show a sharp image of text and on the top-right (b) we show the same image after some motion blur with two dominant peaks. If we run non-local means in the sharp image at the patch containing the 'w' letter, we find only 2 correspondences (a). However, when the same algorithm is run on the blurry image, 4 correspondences (red boxes) are found (b). We can interpret blurring as a process that generates "copies" of a patch with relative displacements due to the nonzero components of blur and contrast proportional to their intensity. Such copies could mislead non-local means into finding additional correspondences. Notice that these copies are overlaid and averaged onto other texture in the image and thus it could happen that some of the true correspondences are instead lost.

To reduce the false positives we suggest using our (partial) knowledge of the blur. Based on our interpretation of blur as a *copy operator*, the majority of these errors depends on the largest nonzero components of the blur, so that a small uncertainty will not hinder our strategy. Formally, let $C_p = \{j \in \mathbb{Z} : ||B_p - B_{p+j}||^2 \leq \varepsilon^2\}$ be the set of correspondences for the pixel p learned from the blurred image, and $\mathcal{K} = \{i \in \mathbb{Z} : |\max(k) - k_i| \leq \tau\}$ be the set of non-zero entries of the PSF k, where τ is a threshold based on blur noise. For the sake of simplicity, consider the example in Fig. 2, where a PSF consisting of only two strong peaks is used. Such PSF generates a blurred image (b) that is made of two shifted and overlapping versions of the same sharp image. For each patch centered at p, we enforce that its improved set of correspondences $\hat{C}_p$ be the intersection of all the correspondence sets of patches at pixels with relative displacement given by the main PSF peaks, *i.e.*, where the *consensus* is

unanimous: $\hat{C}_p = \bigcap_{i \in \mathcal{K}} C_i$. In Fig. 2 (c) the PSF peaks determine only two candidate patches for copies: one denoted by a red square and one denoted by a blur circle. When we overlap the correspondences due to both patches by centering them on the same patch (d), we obtain the correct set (a). We can see in Table 1 how this method reduces false positives when run on a database of images. In Section 6 we will also show experimentally how the prior learned with this consensus strategy is effective in reducing artifacts due to blur uncertainties.

5 Image Deconvolution and Outlier Rejection

Once C has been learned, it provides a constraint for one of the image models in Section 3. We then pose the problem of recovering the sharp image f via the following convex optimization problem

$$\min_{f,n,e} \quad \frac{1}{2}\|A\vec{f} - \vec{b}\|_2^2 + \beta\|\nabla f\|_2 + \frac{\lambda}{2}\|n\|_2^2 + \gamma\|e\|_1 \tag{10}$$
$$\text{subject to} \quad g = h * f + n + e$$

Here the constants β, λ, $\gamma > 0$ determine the smoothness of the solution, the Gaussian and uniform noise levels. We consider four regularization terms: the image prior (learned from the blurry input and a dictionary), total variation, small Gaussian noise energy and sparsity of errors in the model. The image prior is enforced via the matrix A and the vector $\vec{b}$, which are obtained from the model $F = DC$. When the dictionary D is built from the image itself ($D = F$), we have $A = I_d - C^T$ and $\vec{b} = 0$. In the case $D = [D_0 \ F]$ we define $A = I_d - C_2^T$ where the matrix $C^T = [C_1^T \ C_2^T]$, C_1 applies to the dictionary D_0 and C_2 to the matrix F.

Notice that the constraint provided by C (or, equivalently, by A) may not be sufficient to regularize the optimization problem. For example, if there are no similar patches in the model $F = FC$, then $C = I_d$ and since $A = I_d - C^T = 0$ and $\vec{b} = 0$, the first term in eq. (10) will be always 0. To avoid this issue, we use the total variation term $\|\nabla f\|_2$ where the symbol ∇ denotes the gradient operator. Finally, we impose sparsity in e by penalizing its ℓ_1 norm.

To solve the problem in eq. (10), we write n in terms of the image model and substitute its expression in the energy. Then, we minimize the energy by using the following gradient iteration

$$f^{t+1} = \arg\min_f \frac{1}{2}\|A\vec{f} - \vec{b}\|_2^2 + \beta\|\nabla f\|_2 + \frac{\lambda}{2}\|g - h*f - e^t\|_2^2 \tag{11}$$
$$e^{t+1} = \mathcal{T}_{\frac{\gamma}{\lambda}}(g - h*f^t), \quad e^0 = 0$$

where f^t and e^t are the t-th iteration estimates of f and e respectively, and $\mathcal{T}_\alpha(x)_i \doteq (|x| - \alpha)_+ \text{sign}(x_i)$ is the shrinkage operator. The first minimization problem can be solved by computing the Euler-Lagrange equations and then by linearizing them around the current solution. We obtain the following linear system in f

$$A^T(A\vec{f} - \vec{b}) - \beta\nabla \cdot \frac{\nabla f}{\|\nabla f^t\|_2} - \lambda h^* * (g - h*f - e^t) = 0, \tag{12}$$

which can be solved efficiently via conjugate gradient descent. Finally, to deal with saturated pixels, we simply set n to zero and $e = g - h*f$ at the pixels in the image model where the blurred image g is equal to either the minimum or the maximum value in the range.

6   Experiments

In this section we compare our approach with the state-of-the-art deblurring algorithms. We will show that our approach is effective when the blur kernel is not known with high accuracy. In all experiments we use patches of size 5×5 and build the dictionary using patches from the Caltech 101 dataset [6].

We first present quantitative results using the blur database introduced in [11]. This database consists of 32 images (4 images with 8 different blurs) of size 255×255 pixels. We add different levels of Gaussian noise to the ground-truth PSF (from 0% to 5%) and then evaluate the reconstructions obtained by the different algorithms. Although in general real PSFs are affected by non Gaussian noise, the artifacts produced by PSFs corrupted by Gaussian or non Gaussian noise are similar.

We report the performance using two metrics, namely, structured similarity index (SSIM) and peak signal-to-noise ratio (PSNR). The PSNR is a classic metric used to measure image quality, but it does not match well human-perceived image quality [8]. To overcome this limitation we use also the SSIM metric recently proposed by Wang *et al.* [22].

As all the methods depend on tuning parameters, we thoroughly examined two ways of setting the parameters. In the first setting we tested the performance of each individual algorithm for a given PSF noise level on several tunings and picked the best parameter based on the average of all the metrics (normalized between 0 and 1). This test shows how well an algorithm can work when the PSF noise is known. In general, however, one does not know the noise level on the PSF. Hence, in the second setting we repeat the same tuning optimization, but look for the best tuning across all the PSF noise levels. We tuned the parameters on a separate blur database of 12 images with 7 different blurs (84 samples).

Table 2 shows the quantitative performance of our method using the imaging models $D = [D_0 \ F]$ and $D = F$ compared with other deconvolution algorithms using different metrics. In this table the best performing method is highlighted in bold. While for low noise levels our algorithm is close to the best performing method, as the noise level increases we see that our method outperforms the other methods across all metrics. Also, shown in this table is the case when we set the parameters using the second tuning method. For this case we see that our method outperforms or matches the best method. We repeated the same experiment after adding 2.5% noise to our images. The results in this case are shown in Table 3 with the same arrangement as in the previous table. Notice that, as in Table 2, our algorithm yields the best performance when the PSF is noisy.

We also performed experiments on real images. In this case we used the following parameters; $\beta = 0.5$, $\lambda = 1000$, $\gamma = 0.001$ and $\varepsilon = [0.0118, 0.0196]$. Fig. 4 and Fig. 3 compare qualitatively results obtained on blurry images taken at night. In night images blurring is a common problem due to the longer exposures necessary for low-light conditions. Since this image was not synthetically blurred or calibrated, we do not have access to the original sharp image or the PSF. Hence, quantitative results are not shown on these images. Furthermore, the original implementation of Welk *et al.* [23] is not available and the results shown in Fig. 4 were kindly provided to us by the authors. Consequently, we do not include this method in our previous quantitative comparisons. In Fig. 1, Fig. 4 and Fig. 3 we show details of reconstructed images from real images and blur. In such images, our method effectively removes artifacts and performs better than algorithms that use classical sparse image priors or non Gaussian noise models.

Table 2: Deblurring performance comparison with image noise of 0 %.

	PSNR					SSIM				
	$[D_0 F]$	F	[▯]	[▯]	[▯]	D_0F	F	[▯]	[▯]	[▯]
Fixed parameters										
PSF Noise 0%	37.29	37.28	**37.65**	33.75	36.19	0.980	0.980	**0.983**	0.953	0.980
PSF Noise 3%	**26.70**	26.64	26.43	26.31	26.01	**0.827**	0.824	0.820	0.816	0.819
PSF Noise 5%	**22.60**	22.53	22.27	22.49	22.03	**0.697**	0.692	0.683	0.695	0.684
Noise-adaptive parameters										
PSF Noise 0%	**38.90**	**38.90**	**38.90**	33.88	36.19	**0.989**	**0.989**	**0.989**	0.952	0.980
PSF Noise 3%	**27.20**	27.07	26.68	26.66	26.61	**0.840**	0.834	0.831	0.827	0.836
PSF Noise 5%	**24.40**	24.24	23.64	24.02	23.31	**0.764**	0.756	0.747	0.753	0.743

Table 3: Deblurring performance comparison with image noise of 2.5 %.

|	PSNR					SSIM				
	$[D_0 F]$	F	[▯]	[▯]	[▯]	D_0F	F	[▯]	[▯]	[▯]
Fixed parameters										
PSF Noise 0%	28.80	**28.94**	28.93	28.17	28.65	0.851	**0.854**	0.819	0.826	0.843
PSF Noise 3%	**26.44**	26.32	25.56	26.12	25.05	**0.801**	0.797	0.753	0.781	0.768
PSF Noise 5%	**24.28**	24.01	22.89	24.03	22.47	**0.748**	0.737	0.679	0.724	0.692
Noise-adaptive parameters										
PSF Noise 0%	**28.95**	**28.95**	28.91	27.62	28.63	0.830	0.829	0.819	0.781	**0.843**
PSF Noise 3%	**26.22**	26.21	**26.22**	26.19	25.55	**0.785**	0.784	0.784	0.784	0.764
PSF Noise 5%	**24.74**	24.62	24.59	23.88	24.06	**0.745**	0.742	0.740	0.721	0.721
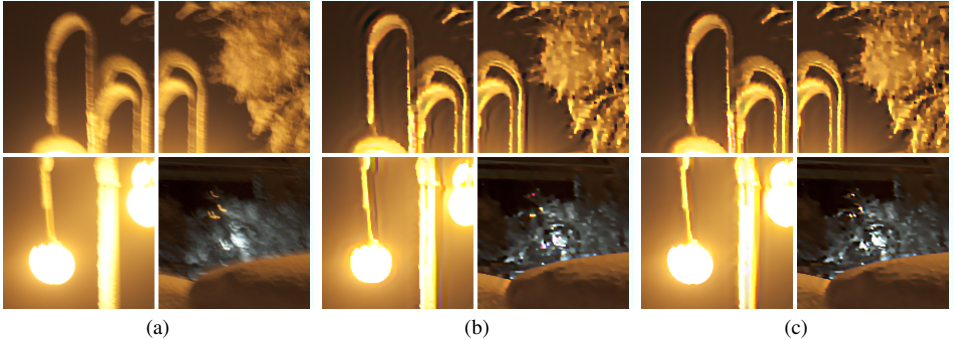
(a)   (b)   (c)

Figure 3: a) 2x2 selected areas from a real blurred image; b) Deblurring results with the algorithm of Cho *et al.* [▯]; c) deblurring results with our algorithm.
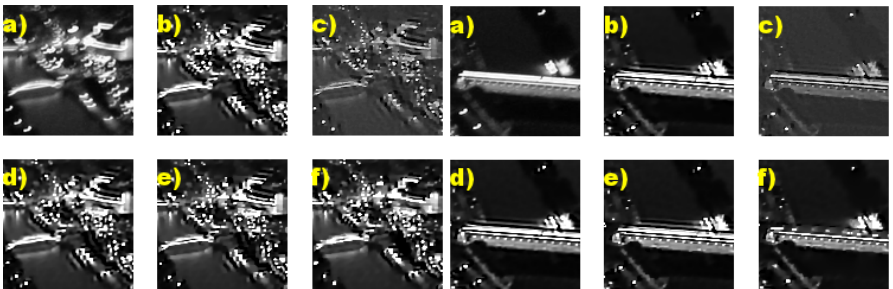

Figure 4: **Example of image deblurring with uncertain non-uniform blur and saturated pixels.** a) Input blurred image. b) Krishnan et al. [▯], c) Welk et al. [▯], as reported in their paper, d) Levin et al. [▯], e) Cho et al. [▯], f) Our method.

7 Conclusions

In this paper we have proposed a novel method using dictionary-based and self-expressing priors that is used for image deblurring when the blur is uncertain. We have introduced a prior that is learned from the blurred image and a dictionary of patches in order to avoid ringing artifacts. Our experimental results show that our performance is overall better than the state-of-the-art methods when the blur kernel is noisy. As future work we will investigate different ways to learn image priors from the blurred image.

8 Acknowledgments

References

[1] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *CVPR*, pages 60–65, 2005.

[2] S. Cho and S. Lee. Fast motion deblurring. *ACM Trans. Graph.*, 28(5):1–8, 2009. ISSN 0730-0301.

[3] S. Cho, J. Wang, and S. Lee. Handling outliers in non-blind image deconvolution. In *ICCV*, pages 1–8, 2011.

[4] T.S. Cho, N. Joshi, C.L. Zitnick, S.B. Kang, R. Szeliski, and W.T. Freeman. A content-aware image prior. In *CVPR*, pages 169–176, 2010.

[5] W. Dong, L. Zhang, and G. Shi. Centralized sparse representation for image restoration. In *ICCV*, pages 1259–1266, 2011.

[6] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, volume 2, pages 264–271, Madison, Wisconsin, June 2003.

[7] 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, 2006. ISSN 0730-0301.

[8] B. Girod. Digital images and human vision. chapter What's wrong with mean-squared error?, pages 207–220. MIT Press, Cambridge, MA, USA, 1993. ISBN 0-262-23171-9.

[9] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-Laplacian. In *NIPS*, 2009.

[10] A. Levin. Blind motion deblurring using image statistics. In *NIPS*, 2006.

[11] A. Levin, R. Fergus, F. Durand, and W.T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26, 2007. ISSN 0730-0301.

[12] Y. Lou, A. L. Bertozzi, and S. Soatto. Direct sparse deblurring. *J. Math. Imaging Vis.*, 39:1–12, 2011. ISSN 0924-9907.

[13] S. Lyu, E.P. Simoncelli, and H. Hughes. Statistical modeling of images with fields of gaussian scale mixtures. In *NIPS*, pages 416–423. MIT Press, 2006.

[14] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53 –69, 2008. ISSN 1057-7149.

[15] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *CVPR*, pages 2272 –2279, 2009.

[16] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338 – 1351, 2003. ISSN 1057-7149.

[17] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physics D*, 60:259–268, 1992. ISSN 0167-2789.

[18] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *ACM Trans. Graph.*, pages 1–10, 2008. ISBN 978-1-4503-0112-1.

[19] A. Srivastava, A.B. Lee, E. P. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18:17–33, January 2003. ISSN 0924-9907.

[20] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems*. Vh Winston, 1977.

[21] C. Wang, L. Sun, Z. Chen, J. Zhang, and S. Yang. Multi-scale blind motion deblurring using local minimum. *Inverse Problems*, 26(1), 2010.

[22] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 13(4):600–612, 2004.

[23] M. Welk, D. Theis, and J. Weickert. Variational deblurring of images with uncertain and spatially variant blurs. In *DAGM-Symposium*, pages 485–492, 2005.

[24] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, 2010.