

## Image Deblurring Via an Adaptive Dictionary Learning Strategy

Lei Li<sup>1</sup>, Ruiting Zhang<sup>2</sup>, Jiangmin Kan<sup>\*3</sup>, Wenbin Li<sup>4</sup>

<sup>1,3,4</sup> School of Technology, Beijing Forestry University, 100083, Beijing, China

<sup>1</sup> Institute of Atmospheric Physics, Chinese Academy of Sciences, 100029, Beijing, China

<sup>2</sup> Canvard College, Beijing Technology and Business University, 101118, Beijing, China

\*Corresponding author: kanjm@bjfu.edu.cn

### Abstract

Recently, sparse representation has been applied to image deblurring. The dictionary is the fundamental part of it and the proper selection of dictionary is very important to achieve super performance. The global learned dictionary might achieve inferior performances since it could not mine the specific information such as the texture and edge which is contained in the blurred image. However, it is a computational burden to train a new dictionary for image deblurring which requires the whole image (or most parts) as input; training the dictionary on only a few patches would result in over-fitting. To address the problem, we instead propose an online adaption strategy to transfer the global learned dictionary to a specific image. In our deblurring algorithm, the sparse coefficients, latent image, blur kernel and the dictionary are updated alternatively. And in every step, the global learned dictionary is updated in an online form via sampling only few training patches from the target noisy image. Since our adaptive dictionary exploits the specific information, our deblurring algorithm shows superior performance over other state-of-the-art algorithms.

**Keywords:** sparse representation, adaptive dictionary learning, image deblurring

### 1. Introduction

Image blur is usually caused by relative motion between the camera and the scene during the exposure time, e.g., camera shake. And this process can be modeled as the following equation:

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

Where  $x$  is the desired sharp image;  $y$  is the blurred image, that is, the degraded image;  $k$  is the blur kernel that is assumed to be linearly shift-invariant;  $n$  is the additive noise usually assumed as following Gaussian distribution;  $\otimes$  denotes the convolution operator. Image deblurring can be seen as an inverse problem of Eq.(1). Given only the degraded image as input, the goal of blind image deblurring is to inverse the above process and to recover both  $x$  and  $k$ , which are assumed to be unknown in this case. This is a so-called blind image deblurring which is a fundamental and challenging problem in the image and signal processing literature [1]-[3]. Image deblurring has many practical applications, such as removing the blur from consumer photographs, computational photography and astronomical imaging. However, it is still an open question to design an efficient deblurring algorithm.

Traditional approaches to remove image blur are always done by employing deconvolution method. However, the deconvolution is a severely ill-posed problem, that is, there can be exponentially many images  $x$  and kernels  $k$  that satisfy Eq.(1). To alleviate these issues, many prior assumptions and regularities are introduced on the structure of  $x$  and  $k$  [3]-[9]. The statistical priors of natural images are often utilized. And a commonly used prior on  $x$  is the heavy-tailed prior [4], that gradients of natural images follow a hyper-Laplacian distribution which was observed by Levin et al..

Recently, using redundant representations and sparse property in nature images has drawn a lot of research attention. And sparse representation has also been applied in image deblurring and achieved satisfactory performance [7],[10]. Cai et al. proposed a blind motion deblurring method by exploiting the sparseness of natural images in over-complete

predefined dictionary (e.g. wavelets, DCT) to help with kernel estimation and sharp image estimation [7]. Jia et al. proposed a new non-blind image deblurring method which jointly modeled the sparse representation of natural image patches and sparse gradient priors [11]. Li et al. proposed a blind image deblurring method by combining the sharp and blur dictionary pair and the sparse gradient prior, and estimating the blur kernel, sharp image and sparse coefficients respectively during the deblurring process [12]. The redundant and over-complete dictionary has been trained on image patches to help exploit the sparse prior of natural images [7],[11],[12].

Although deblurring with a prespecified dictionary is simple and fast, it results in low performance in most cases [13]. The reason is that the generic bases (i.e., pre-defined basis such as curvelets and framelets) do not exploit rich information contained in the blurred image or the global learned dictionary does not transfer well to different types of images. In [14], Hu et al. proposed a deblurring method that exploited sparse representation using a blurry-sharp dictionary pair learned directly from the blurred image. And the deblur kernel and sharp image were estimated iteratively.

However, the adaptive strategy [13] to learn a new dictionary is designed to work with overlapping patches (one per-pixel) of whole image. There are about 250000 patches (8x8) in a 512x512 image. Training with such redundant patches from a single image is time-consuming. Conversely, training with only a few patches leads to over-fitting. To balance the dilemma on how to quickly learn the dictionary and yet achieve higher accuracy, we propose a domain adaptation approach transferring the global learned dictionary to the specific image for deblurring. In our deblurring algorithm, the sparse coefficients, latent image, blur kernel and the dictionary are updated alternatively. And at every step, the global learned dictionary is updated in an online form via sampling only a few training patches from the blurred image, which made the updating process more efficient. The new learned dictionary exploits the specific information in the blurred image. Therefore, it can represent patches much better than the prespecified dictionary's. Compared to the deblurring methods [12],[15], our method shows the superior deblurring performance.

The rest of this paper is organized as follows. Our new image deblurring algorithm is described in detail in section 2. Experimental results and comparison with other state-of-the-art approaches are presented in section 3. Our work are summarized in section 4.

## 2. Image Deblurring Via Dictionary Adaption

### 2.1 Problem formulation

It is well-known that natural image patches can be modeled via sparse representation over an over-complete dictionary. Let  $x \in \mathbb{R}^n$  is an image patch,  $n$  is the dimension of the feature,  $D \in \mathbb{R}^{n \times k}$  is an over-complete dictionary with  $k$  atoms. Then the representation can be formulated as follow [13]:

$$x = D\alpha \quad (2)$$

Where the representation coefficient  $\alpha \in \mathbb{R}^k$  is sparse.

In the sparse representation,  $D$  is the fundamental part and a proper selection of  $D$  is very important. Popular selections for  $D$  are: Fourier, wavelet, wavelet packet, cosine packet, and curvelet etc. [16]. Nevertheless, the predefined dictionary has its advantages in some kind of problems. Therefore, no single method currently can represent an image desirably and generally. Learning the dictionary adaptively from sample images is another option. Nevertheless, the global learned dictionary does not exploit rich information contained in the blurred image, and therefore, it may not transfer well to the specific image.

To overcome this problem, we could apply a similar strategy like KSVD [13] to learn an adaptive dictionary on the blurred image. KSVD is a kind of generalization of the K-means algorithm. Given a set of image patches  $Y = [y_1, y_2, \dots, y_N]$ , KSVD learns the dictionary  $D$  by solving the following problem:

$$\arg \min_{D, \alpha_i} \sum_{i=1}^N \|y_i - D\alpha_i\|_2^2 \quad s.t. \quad \|\alpha_i\| \leq T_0, \quad \forall i. \quad (3)$$

Where  $N$  is the number of patches,  $T_0$  is the maximum of nonzero entries of  $\alpha_i$ . Since problem (3) is non-convex, the dictionary  $D$  and the coefficient  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$  are learned iteratively while the other variable is fixed.

However, KSVD is designed to learn the basis with overlapping patches (one per-pixel) of the whole image. The number of image patches is redundant and about 250000 patches (8x8) in a 512x512 image. Training with such many patches from a single image is time-consuming, as the author claimed that each iteration of their algorithm takes approximately 5 minutes on a windows PC of 2.67 GHz CPU and 4 GB RAM [14]. Conversely, training with a few patches leads to over-fitting. To balance the dilemma on how to quickly learn an adaptive dictionary and yet achieve higher accuracy, we propose our image deblurring algorithm via dictionary adaptation to a specific domain. And based on the above discussion, we propose our image deblurred model and concrete regularization terms:

$$\{\hat{\alpha}_i, \hat{x}, \hat{k}, \hat{D}\} = \arg \min_{\alpha_i, x, k, D} \|y - x \otimes k\|_F^2 + \sum_{i=1}^{T_0} (\|R_i y - D \otimes k \alpha_i\| + \lambda \|\alpha_i\|_1) \quad (4)$$

$$+ \tau \|G \otimes x\|^\omega + \beta \|k\|_2^2 + \gamma \|D - D_0\|_F^2$$

Where  $D_b = D \otimes k$ . Eq.(4) has four terms, the first term is the blurred image which can be sparsely represented under the blur dictionary  $D_b$ ; the second term is the sparse prior of the gradient image assumed to follow the hyper-Laplacian distribution [3], while  $G$  is the gradient extraction filters and  $\omega \in [0.5 - 0.8]$ ; the third term is a  $L_2$  norm regularization to the blur kernel estimation; and the last term is the regularization on the target dictionary  $D$  with  $D_0$ , which controls the complexity of the target domain to avoid the over-fitting problem and makes it possible with a few training data for dictionary learning.

## 2.2 Optimization Algorithm

Obviously, the object function in problem (4) is non-convex. It can be easily proved that problem (4) is jointly convex with respect to all of its variables. And we propose an algorithm that alternately optimizes one variable while the others are fixed. The method is described in Algorithm 1.

**1) Sparse representation sub-problem:** At the beginning of each iteration, we fix  $k$  and  $D$  to estimate the coefficient  $\alpha$  of each image patch, the problem is then transformed to:

$$\{\hat{\alpha}_i\} = \arg \min_{\{\alpha_i\}} \sum_{i=1}^{T_0} (\|R_i y - D_b^k \alpha_i\|_2^2 + \lambda \|\alpha_i\|_1) \quad (5)$$

Where  $D_b^k = D_b \otimes k^k$ . Obviously, we could decompose the Eq. (5) for each image patch separately, and solve each problem as follow:

$$\{\hat{\alpha}_i\} = \arg \min_{\alpha_i} \|R_i y - D_b^k \alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \quad (6)$$

There is only one unknown variable in the above optimization problem, which can be solved efficiently via the basis pursuit algorithm [17] or the accelerated proximal gradient (APG) methods [18],[19].

**2) Latent image estimation sub-problem:** Similar to the ideas proposed in [20], we assume the sparse coefficients  $\alpha_i$  of the latent image patch  $R_i x$  with  $D$  are the same as the one of blurred image  $R_i y$  relative to  $D_b$ . Thus, when the current estimation  $\alpha_i$  is known, we can reconstruct latent image  $R_i x$  via the following problem:

$$\hat{x} = \arg \min_x \|y - x \otimes k\|_F^2 + \sum_i (\|R_i k - D \alpha_i\|_2^2) + \tau \|G \otimes x\|^\omega \quad (7)$$

The solution to problem (7) can be obtained quickly according to [10]. In practice, we set the  $G$  as first-order gradient filters  $G_1 = [1, -1]$ ,  $G_2 = G_1^T$  and set  $\omega = 0.5$  the same as configurations in [10].

**3) Blur kernel estimation sub-problem:** As the other sub-problems, all other variables except  $k$  are fixed. The minimization of model (4) reduces to the following problem:

$$\hat{k} = \arg \min_k \|y - x \otimes k\|_F^2 + \beta \|k\|_F^2 \quad (8)$$

This is a least square problem with Tikhonov regularization, which is easily solved in the frequency-domain.

**4) Dictionary adaption sub-problem:** In this sub-problem, we update the dictionary of the latent image  $x$  which has been estimated in the latent image estimation sub-problem. We set other variables except  $D$  fixed, and dictionary updating degrades to the following problem:

$$\begin{aligned} \hat{D} &= \arg \min_D \sum_{i=1}^M (\|R_i x - D^k \alpha_i\|_2^2 + \gamma \|D - D_0\|_F^2) \\ &= \arg \min_D (\text{tr}(D^T D (C C^T + \gamma I)) - 2 D^T (C^T X + \gamma D_0)) \\ &= \arg \min_D \text{tr}(D^T D A) - 2 \text{tr}(D^T B) \end{aligned} \quad (9)$$

Where  $I$  is the identity matrix,  $X = [R_1 x, R_2 x, \dots, R_M x]$ ,  $C = [\alpha_1, \alpha_2, \dots, \alpha_M]$ ,  $M$  is the number of patches used to update the dictionary, and  $A = C C^T + \gamma I$  and  $B = C^T X + \gamma D_0$ . Problem (9) can be solved in the following closed formulation:

$$D = (C C^T + \gamma I)^{-1} (C^T X + \gamma D_0) \quad (10)$$

However, the inverse of a matrix is time-consuming. We utilized the stochastic gradient descent [21] to update the dictionary effectively and efficiently. This on-line method is effective and easy to implement. The algorithm for updating the dictionary is presented in Algorithm 2.

**Algorithm 1** Blind image deblurring via dictionary adaption.

- Input: blurred image  $y$ , sharp dictionary  $D_0$ , kernel size, iteration number  $T$ , the number of patches  $M$ .
- Output: Estimated latent image  $x$  and blur kernel  $k$ .
- Initialization:  $k$
- 1. For:  $t = 1, 2, \dots, T$ ;
- 2. Sparse representation: update sparse coefficients  $\{\alpha_i\}$  via minimizing the sub-problem (5);
- 3. Latent image estimation: update the latent image  $x$  via minimizing the sub-problem (7);
- 4. Blur kernel estimation: update the blur kernel  $k$  via minimizing the sub-problem (8);
- 5. Dictionary Updating: update the sharp dictionary  $D$  via minimizing the sub-problem (9);
- 6. end For.

**Algorithm 2** Optimization for Problem (9)

- Input:  $D_0$ , latent image  $x$ , sparse coefficients  $\{\alpha_i\}$ , the number of patches  $M$ .
- Output: Dictionary  $D_t$ .

1. Randomly draw  $M$  patches from latent image  $x$  to form  $X=[R_1x, R_2x, \dots, R_Mx]$  and their corresponding sparse coefficients  $C=[\alpha_1, \alpha_2, \dots, \alpha_M]$ .
2. Calculate  $A=CC^T + \gamma I$  and  $B=C^T X + \gamma D_0$ ;
3. For:  $j=1$  to  $k$  do
4. Update the  $j$ -th column via optimizing Problem (9):  $u_j = \frac{1}{A[j, j]}(b_j - D\alpha_j) + d_j$ ,

$$d_j = \frac{1}{\max(\|u_j\|_2, 1)} u_j$$

5. end For.

### 3. Experiments

The proposed algorithm which is implemented and corresponding experiments to verify its effectiveness are carried out in MATLAB. In all our experiments, we set  $\lambda=0.01$ ,  $\tau=1500$ ,  $\beta=0.01$ ,  $T=5$ ,  $\gamma=0.1$ ,  $M=2000$  empirically. The initial sharp dictionary  $D_0$  is learned from the sharp images using KSVD [13]. The initial kernel  $k$  is set to be the Gaussian kernel with  $\sigma=1$ . Each iteration of our algorithm takes approximately 2 minutes on a windows PC of 2.50GHz CPU and 4 GB RAM. Blurred images for experiment are synthesized with three different kinds of blur kernel, including motion blur kernel (direction 45 degree and length 5 pixels), Gaussian blur kernel (standard deviation 5 pixels) and average blur kernel (5 pixels), and then additive Gaussian noise following standard deviation of 0.01 are added to the blurred image. For color images, we directly split them into three channels, RGB and deblur each channel respectively. We compared our result with the state-of-the-art algorithm [12],[15],[23] which is a blind image deblurring method based on deconvolution method. The final results are evaluated in terms of PSNR (peak signal to noise ratio) and SSIM (structural similarity index) [22]. Because the authors did not publish their code, we directly quote the deblurring results for the gray images in [12]. For the color images, we only compare our algorithm with the ones in [15].

Table 1. quantitative comparative evaluation. PSNR is chosen as performance measure

File	Blur type	PSNR			
		Dilips's method	Li's method	Dong's method	Our method
house.tif	motion	27.96	29.01	26.16	<b>31.07</b>
	gaussian	27.06	29.26	30.60	30.05
	average	28.21	29.29	31.04	<b>31.09</b>
boats.tif	motion	27.13	28.55	25.52	<b>29.95</b>
	gaussian	26.73	27.11	29.13	<b>29.27</b>
	average	26.43	27.98	29.01	<b>29.07</b>
cameraman.tif	motion	25.83	26.22	25.86	<b>28.71</b>
	gaussian	21.67	25.59	25.56	<b>26.50</b>
	average	23.77	<b>26.18</b>	24.32	26.15
barbara.tif	motion	25.92	25.99	26.36	<b>28.71</b>
	gaussian	21.76	26.02	25.29	<b>26.50</b>
	average	24.97	24.67	24.82	<b>26.15</b>
baboon.tif	motion	23.54	24.44	25.43	<b>25.71</b>
	gaussian	22.89	<b>24.99</b>	24.98	24.12
	average	22.51	23.58	23.63	<b>24.02</b>

\*Best results are noted in bold.

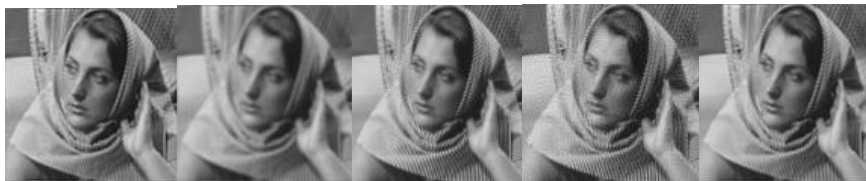
The experimental results of deblurred image in terms of PSNR and SSIM are presented in Table 1 and Table 2 respectively. From the two tables, we can conclude that our algorithm shows superior performance than Dilips's on both PSNR and SSIM. From the deblurred image in Figure 1 and Figure 2, there exists noise amplification in smooth regions in Dilips's method, such as face region in Barbara, background region in House and Cameraman. In contrast, our algorithm is able to get more acceptable results on the deblurred images in these smooth and

textured regions with fewer ringing effects. One possible explanation is that the assumed prior distribution of edges used in [12],[15],[23] does not hold well whereas our method exploits the information from the blurred image.

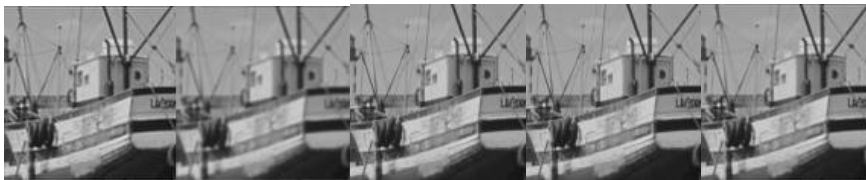
Table 2. quantitative comparative evaluation. SSIM is chosen as performance measure

File	Blur type	PSNR			
		Dilips's method	Li's method	Dong's method	Our method
house.tif	motion	0.816	0.832	0.7144	<b>0.864</b>
	gaussian	0.768	0.787	<b>0.8383</b>	0.829
	average	0.807	0.821	0.7984	<b>0.854</b>
boats.tif	motion	0.845	0.836	0.6736	<b>0.895</b>
	gaussian	0.821	0.842	0.7840	<b>0.865</b>
	average	0.804	0.801	0.7575	<b>0.857</b>
cameraman.tif	motion	0.847	0.848	0.7608	<b>0.849</b>
	gaussian	0.695	0.728	0.7169	<b>0.809</b>
	average	0.708	<b>0.797</b>	0.6519	0.772
barbara.tif	motion	0.837	0.841	0.7282	<b>0.855</b>
	gaussian	0.723	0.732	0.6634	<b>0.808</b>
	average	0.754	<b>0.801</b>	0.6273	0.789
baboon.tif	motion	0.787	0.772	0.7946	<b>0.855</b>
	gaussian	0.747	0.825	0.8060	<b>0.858</b>
	average	0.718	0.816	0.7961	<b>0.837</b>

\*Best results are noted in bold.



(a) (b) (c) (d) (e)



(f) (g) (h) (i) (j)



(k) (l) (m) (n) (o)

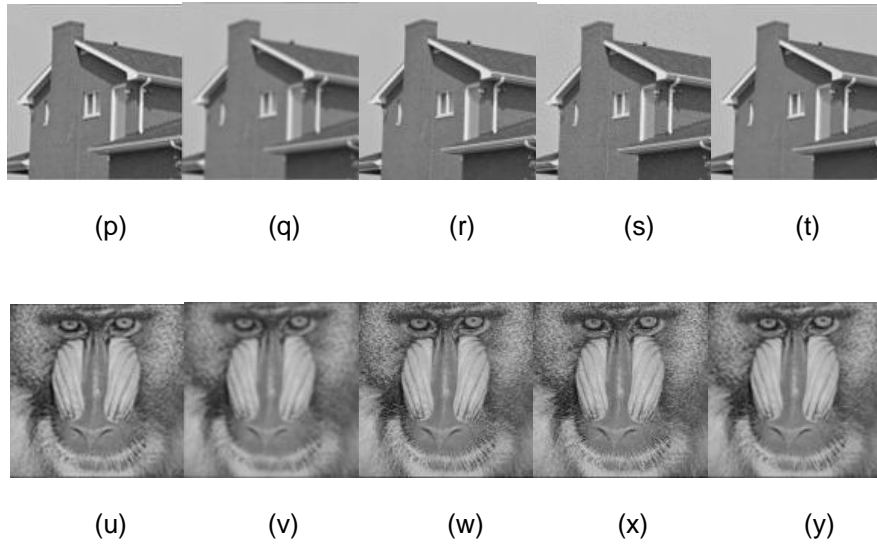


Figure 1. Deblur results





Figure 2. Color Deblur results

From Figure 2 and Table 3, we can see that our method is much more effective deblurring method. The deblurring results with competing methods are compared in Figure 2. We can see that there are many noise residuals and artifacts around edges in the deblurred images in Dilips's method [15]. Our method leads to the best visual quality. It not only can remove the blurring effects and noise, but also can reconstruct more and sharper image edges than Dilips's methods.

Table 3. PSNR and SSIM results of deblurring images.

File	Blur type	PSNR		SSIM	
		Dilips's method	Our method	Dilips's method	Our method
Parrot.jpg	motion	22.56	<b>25.76</b>	0.738	<b>0.835</b>
	gaussian	23.76	<b>25.89</b>	0.771	<b>0.834</b>
	average	22.44	<b>25.34</b>	0.705	<b>0.809</b>
Lena.jpg	motion	25.98	<b>28.61</b>	0.782	<b>0.851</b>
	gaussian	25.23	<b>26.93</b>	0.722	<b>0.783</b>
	average	24.41	<b>26.74</b>	0.651	<b>0.757</b>

\*Best results are noted in bold.

We have also done the deblurring experiments on Barbara with the global learned dictionary and the adaptive dictionary, as shown in Fig.3. The PSNR and SSIM are 25.76dB and 0.7884 respectively without dictionary adaption while the counterparts are 26.50dB and 0.808.



Deblurring with adaptive dictionary is showed the superior performance because it exploited the specific information in the image.

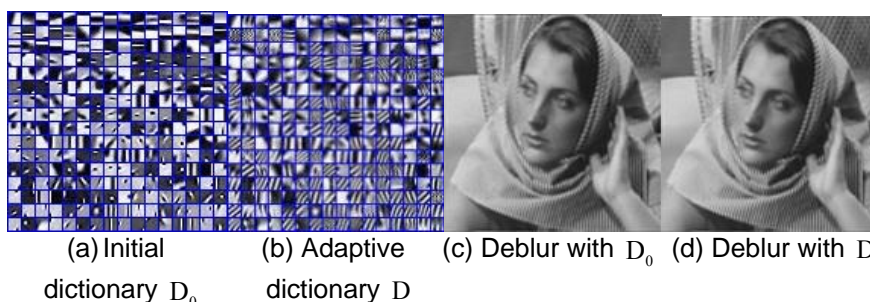


Figure 3. Comparison of deblurring results on Barbara with the global learned dictionary and the adaptive dictionary. The PNSR are 25.76dB and 26.50dB respectively and the SSI are 0.7884 and 0.808

#### 4. Conclusion

In this paper, we proposed an online adaptive dictionary strategy for image deblurring algorithm based on the sparse representation. Our strategy transfers a global learned dictionary to a specific image to make the trade-off between computational cost and accuracy. The new learned dictionary exploited the specific information such as the texture and edge information in the blurred image. Therefore, it can represent patches better than the prespecified dictionary and the global learned dictionary. The approach taken is based on sparse and redundant representations in overcomplete dictionaries learned. With only a few training samples, our method can speed up the learning process via domain adaptation. Furthermore, the domain adaptation allows us to circumvent the overfitting problem effectively. Also, the adaptive dictionary exploited the specific information in the image, which makes our deblurring algorithm more effective. Experiments have demonstrated that our image deblurring method yields superior performances over other methods.

#### Acknowledgements

This work was supported Beijing Higher Education Young Elite Teacher Project (NO. YETP1949), Fundamental Research Funds for the Central Universities under Grant No. TD2013-4, and National Natural Science Foundation of China under Grant No. 30901164.

#### References

- [1] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics*. 2006; 25: 787–794.
- [2] A. Levin. Blind motion deblurring using image statistics. *Advances in Neural Information Processing Systems (NIPS)*. 2006; 2: 4.
- [3] Q. Shan, J. Jia, A. Agarwala. High-quality motion deblurring from a single image. *ACM Transactions on Graphics*. 2008; 27(73):10.
- [4] A. Levin, R. Fergus, F. Durand, W.T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*. 2007; 26: 70-71.
- [5] J. Jia. *Single image motion deblurring using transparency*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2007: 1–8.
- [6] N. Joshi, R. Szeliski, D. Kriegman. *PSF estimation using sharp edge prediction*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1–8.
- [7] J.-F. Cai, H. Ji, C. Liu, Z. Shen. *Blind motion deblurring from a single image using sparse approximation*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009: 104–111.
- [8] A. Levin, Y. Weiss, F. Durand, W.T. Freeman. *Understanding and evaluating blind deconvolution algorithms*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009:

- 1964–1971.
- [9] N. Joshi, C. Zitnick, R. Szeliski, D. Kriegman. *Image deblurring and denoising using color priors*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009: 1550–1557.
  - [10] D. Krishnan, R. Fergus. Fast image deconvolution using hyperlaplacian priors. *Advances in Neural Information Processing Systems (NIPS)*. 2009; 22: 1–9.
  - [11] C. Jia, B. Evans. *Patch-based image deconvolution via joint modeling of sparse priors*. Proceedings of the 18th IEEE International Conference on Image Processing (ICIP). 2011: 681–684.
  - [12] H. Li, Y. Zhang, H. Zhang, Y. Zhu, J. Sun. *Blind image deblurring based on sparse prior of dictionary pair*. Proceedings of the 21th International Conference on Pattern Recognition (ICPR). 2012: 3054–3057.
  - [13] M. Aharon, M. Elad, A. Bruckstein. K-SVD: An algorithm for designing over complete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*. 2006; 54: 4311–4322.
  - [14] Z. Hu, J.B. Huang, M.H. Yang. *Single image deblurring with adaptive dictionary learning*. Proceedings of the 17th IEEE International Conference on Image Processing (ICIP). 2010: 1169–1172.
  - [15] K. Dilip, F. Tay, T. and Rob. *Blind deconvolution using a normalized sparsity measure*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2011: 233–240.
  - [16] J.L. Starck, M.K. Nguyen, F. Murtagh. Wavelets and curvelets for image deconvolution: a combined approach. *Signal Processing*. 2003; 83: 2279 – 2283.
  - [17] S.S. Chen, D.L. Donoho, M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*. 1998; 20: 33–61.
  - [18] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *SIAM Journal on Optimization*. 2008.
  - [19] S. Ji, J. Ye. *An accelerated gradient method for trace norm minimization*. Proceedings of International Conference on Machine Learning. 2009: 457-464.
  - [20] Y. Lou, A. L. Bertozzi, S. Soatto. Direct sparse deblurring. *Journal of Mathematical Imaging and Vision*. 2011; 39: 1–12.
  - [21] J. Mairal, F. Bach, J. Ponce, G. Sapiro. *Online dictionary learning for sparse coding*. Proceedings of the 26th Annual International Conference on Machine Learning. 2009: 689–696.
  - [22] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004; 14: 600–612.
  - [23] Weisheng Dong, Lei Zhang, Guangming Shi, Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE Trans. on Image Processing*. 2013; 22(4): 1620-1630.