PATCH-BASED IMAGE DECONVOLUTION VIA JOINT MODELING OF SPARSE PRIORS

Chao Jia and Brian L. Evans

Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, Texas 78712
cjia@mail.utexas.edu, bevans@ece.utexas.edu

ABSTRACT

Image deconvolution aims to recover an image that has been degraded by a linear operation such as blurring during image acquisition. Deconvolution based on maximum a-posteriori (MAP) estimation requires the global prior probability of the original image. Conventional methods usually model the image priors by uniformly characterizing the statistical properties of either some forward measurements of images or the representation coefficients in frames, neglecting the local image statistics. In this paper, we adopt local sparse representation in image deconvolution. Our contributions include proposing (1) a joint model of natural images combining sparse representation of image patches and sparse gradient priors, and (2) an efficient iterative algorithm to infer the MAP estimate of image deconvolution using the proposed model. Experiments indicate that the proposed method can recover the original image with high peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index compared with state-of-the-art methods.

Index Terms— sparse representation, image priors, image deconvolution.

1. INTRODUCTION

Image deconvolution is a kind of inverse problem in image processing that aims to recover the original image $\mathbf{X}$ given an observation

$$\mathbf{Z} = \mathbf{H}\mathbf{X} + \mathbf{n} \tag{1}$$

where $\mathbf{H}$ is the matrix form of the convolution kernel (a linear shift-invariant operator) for vectorized images and $\mathbf{n}$ is additive noise. In this paper we restrict our discussion to Gaussian noise and assume $\mathbf{H}$ is given. This problem is known as non-blind deconvolution and the most common way to solve it is to adopt MAP estimation, which requires the prior probability of $\mathbf{X}$ to bias the estimation process to generate more likely results.

Existing image priors usually model images either in an analysis way or a synthesis way [1] and in both kinds of priors sparsity is widely used. Analysis-based priors model the probability of various forward measurements of images such as gradients [2], while synthesis-base priors decompose each image into a sparse linear combination of atoms in a dictionary and directly model the probability distribution of combination coefficients [3].

Recent research in visual cortex [4] and signal processing [5] reveals that image patches can be sparsely represented over a learned overcomplete dictionary (also known as "sparse coding") better than pre-defined frames like wavelets. This particular kind of synthesis-based priors has been successfully applied in image denoising and super-resolution with state-of-the-art PSNR results [6, 7]. In denoising, overlapping patches are computed separately, and then the denoised patches are averaged to reduce boundary artifacts.

However, sparse representation of image patches is not that straightforward to apply in image deconvolution, for the reason that the intensity value of a certain pixel in $\mathbf{Z}$ depends on many different patches in $\mathbf{X}$ due to the convolution operator $\mathbf{H}$. Therefore, the coefficients of different patches are coupled and cannot be inferred independently, which results in a very-large-scale optimization problem. Besides, in denoising the patches are denoised first and then averaged, but to directly assume the image is equal to the average of overlapping patches is questionable since in sparse coding the dictionary is learned from independent patches. This forces us to drop the overlap among patches, which will cause boundary artifacts.

To solve the above problems and take advantage of sparse coding in image deconvolution, we develop a joint model of the representation coefficients of image patches and image gradients, which actually combines analysis-based priors and synthesis-based priors. In the new model, the statistical properties of gradients can be used to regularize the decomposition of image patches and eliminate boundary artifacts. We also propose an efficient iterative algorithm to solve the formulated problem. As we will show, the MAP estimate can be obtained in only a few iterations, and in each iteration the coefficients of patches are decoupled and can be inferred independently. We demonstrate with experimental results that this algorithm provides competitive and even better figures of merit compared with state-of-the-art methods. To the best of our knowledge, this is the first time that sparse coding of patches is used for globally modeling of image priors and applied in image deconvolution. [8] also provides a deconvolution algorithm using sparse coding, but it is still a localized algorithm and can only handle convolution kernels with very small support sizes. The code of this paper is available in [9].

2. JOINT PRIOR MODEL

2.1. Local Model Using Sparse Representation

We define a vectorized image patch as $\mathbf{x} \in \mathbb{R}^{N\times 1}$. According to [4], $\mathbf{x}$ can be represented as a linear combination over an overcomplete dictionary as $\mathbf{x} = \mathbf{\Phi}\mathbf{w}$, where $\mathbf{\Phi} \in \mathbb{R}^{N\times M}$ is the overcomplete dictionary, $\mathbf{w} \in \mathbb{R}^{M\times 1}$ is the representation coefficient which follows some sparse distribution. The most frequently employed distribution is Laplace distribution: $p(\mathbf{w}) = \frac{\lambda}{2}\exp(-\lambda\|\mathbf{w}\|_1)$. The dictionary $\mathbf{\Phi}$ is universal for all natural image patches and can be learned efficiently through some optimization methods [5, 10].

In global modeling of natural images, it is unreasonable to directly assume the image to be the average of overlapping patches that can be sparsely decomposed in the learned dictionary $\mathbf{\Phi}$, since the dictionary is only learned from isolated image patches. Thus

This research was supported by an equipment gift by Intel.

intuitively we can only model the entire image $\mathbf{X} \in \mathbb{R}^{L\times 1}$ as $\sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i$, where $\{\mathbf{w}_i\}$ are representation coefficients of nonoverlapping patches, $\mathbf{R}_i$ is a $L \times N$ matrix that maps the i^{th} patch to its corresponding position in the image. Applying this model in the MAP estimation of image deconvolution, we have

$$\{\hat{\mathbf{w}}_i\} = \arg\min_{\{\mathbf{w}_i\}} \frac{1}{2\sigma^2}\|\mathbf{Z} - \mathbf{H}\left(\sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i\right)\|_2^2 + \lambda\sum_i \|\mathbf{w}_i\|_1 \tag{2}$$

where the right hand side denotes $-\log p(\mathbf{w}|Z)$ and σ^2 is the variance of the Gaussian noise. Coefficients of different patches will affect each other and thus have to be estimated simultaneously, which makes the number of variables up to millions (the number of image pixels times the overcomplete factor). Moreover, lacking smoothness constraints between nonoverlapping patches can easily cause severe artifacts along the boundaries of patches.

2.2. Joint model of image and representation coefficients

To take advantage of learned sparse coding of image patches and settle the mentioned problems above at the same time, we propose a joint model that consists of both the image and its representation coefficients in the learned dictionary. Many learned sparse coding and synthesis-based prior models assume that any image is equivalent to a linear combination of elements in a certain dictionary (possibly plus an estimation error term) [1]. This equivalence restricts the prior model on statistical properties of the representation coefficients only.

In our joint model, we drop this equivalence and introduce some other analysis-based priors of images to further regularize the model. In fact, we can consider the traditional synthesis-based models as a directed graphical model, in which the image is determined by the representation coefficients in some dictionary. Our joint model is an undirected model, so the image and its representation coefficients cannot determine each other. In other words, the representation coefficients are fundamental elements of an image, not just auxiliary variables to model the image. The negative log-prior probability of an image and its representation coefficients in learned sparse coding is defined as

$$-\log p\left(\mathbf{X}, \{\mathbf{w}_i\}\right) = L\left(\mathbf{X}\right) + \lambda\sum_i \|\mathbf{w}_i\|_1 + C\left(\mathbf{X}, \{\mathbf{w}_i\}\right) \tag{3}$$

where the first term is an analysis-based prior of the image, the second term is the sparse prior of the representation coefficients and the third term is the compatibility term that relaxes the equivalence relationship between the image and the linear combination. The analysis-based prior term is introduced to settle the aforementioned problems of learned sparse coding of patches. Considering the simplicity of the model and the fact that the patches do not overlap, we use a sparse gradient prior:

$$L\left(\mathbf{X}\right) = \mu\sum_j |\left(\nabla\mathbf{X}\right)_j|^{2/3}. \tag{4}$$

Here $\nabla\mathbf{X}$ indicates the spatial gradients of the image in both vertical and horizontal directions, μ is a regularization parameter controlling the weight of this term and j is the vector index. This gradient prior assumes a hyper-Laplacian distribution on the gradients [2] and thus enforces smoothness between neighboring pixels, eliminating boundary artifacts. The compatibility term is defined as a Euclidean distance $C\left(\mathbf{X}, \{\mathbf{w}_i\}\right) = \beta\|\mathbf{X} - \sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i\|_2^2$ and thus we obtain the joint prior as

$$-\log p\left(\mathbf{X}, \{\mathbf{w}_i\}\right) = \mu\sum_j |\left(\nabla\mathbf{X}\right)_j|^{2/3} + \lambda\sum_i \|\mathbf{w}_i\|_1 + \beta\|\mathbf{X} - \sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i\|_2^2. \tag{5}$$

The model in (5) accepts the difference between the image and the sparse linear combination, and characterize statistical properties for both of them. The analysis-based gradient prior is widely used alone as an effective prior for image deconvolution, but usually oversmoothes the deconvolution results. The sparse coding of patches, based on the learned overcomplete dictionary, happens to be able to keep textures well in deconvolution.

2.3. Image deconvolution using the joint model

Applying the proposed model (5) in image deconvolution (1) and using MAP estimation, we have

$$\hat{\mathbf{X}}, \{\hat{\mathbf{w}}_i\} = \arg\min_{\mathbf{X},\{\mathbf{w}_i\}} \frac{1}{2\sigma^2}\|\mathbf{Z} - \mathbf{H}\mathbf{X}\|_2^2 + \mu\sum_j |\left(\nabla\mathbf{X}\right)_j|^{2/3} + \lambda\sum_i \|\mathbf{w}\|_1 + \beta\|\mathbf{X} - \sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i\|_2^2. \tag{6}$$

The three weighting parameters μ, β and λ can be fixed for all natural images or tuned carefully to obtain the best results. Compared with (2), although the number of variables becomes even larger, an efficient iterative algorithm will guarantee that the MAP estimation converges fast. Furthermore, as we will show in next section, the image itself serves as an intermediate layer that decouples the representation coefficients of different patches in each iteration.

3. AN ITERATIVE ALGORITHM FOR IMAGE DECONVOLUTION

The formulated problem (6) is not convex due to the compatibility term and the hyper-Laplacian gradient prior. We solve (6) by alternately optimizing with respect to $\mathbf{X}$ or $\{\mathbf{w}_i\}$ while holding the other fixed. Experiments show that a local optimal solution can be reached quickly and serves well as a deconvolved result. Details of the algorithm are derived next.

3.1. w sub-problem

While $\mathbf{X}$ is fixed, we need to solve a problem of the form

$$\min_{\{\mathbf{w}_i\}} \beta\|\mathbf{X} - \sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i\|_2^2 + \lambda\sum_i \|\mathbf{w}_i\|_1 \tag{7}$$

which further decomposes into smaller problems for each patch:

$$\min_{\mathbf{w}_i} \beta\|\mathbf{X}_i - \mathbf{\Phi}\mathbf{w}_i\|_2^2 + \lambda\|\mathbf{w}_i\|_1 \tag{8}$$

where $\mathbf{X}_i$ denotes the i^{th} patch of the image. As shown in (8), for each patch we only need to solve an ℓ_1 regularized square loss minimization problem with a very small scale. Algorithms to solve this kind of problems have been extensively investigated [11].

3.2. X sub-problem

Given a fixed value of $\mathbf{w}$, to find the optimal $\mathbf{X}$ needs solving a more complicated non-convex problem:

$$\min_{\mathbf{X}} \frac{1}{2\sigma^2}\|\mathbf{Z}-\mathbf{H}\mathbf{X}\|_2^2+\beta\|\mathbf{X}-\sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i\|_2^2+\mu\sum_j |\left(\nabla\mathbf{X}\right)_j|^{2/3}. \tag{9}$$

To solve (9) we apply the half-quadratic splitting method which, as demonstrated in [2], can efficiently deconvolve images with hyper-Laplacian gradient priors. We introduce an auxiliary variable $\mathbf{Y}$ that separates the derivative operator ∇ and the $\ell_{2/3}$ pseudo-norm:

$$\begin{aligned}\min_{\mathbf{X},\mathbf{Y}} \frac{1}{2\sigma^2}\|\mathbf{Z}-\mathbf{H}\mathbf{X}\|_2^2+\beta\|\mathbf{X}-\sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i\|_2^2 \\ +\alpha\|\nabla\mathbf{X}-\mathbf{Y}\|_2^2+\mu\sum_j |\left(\mathbf{Y}\right)_j|^{2/3}.\end{aligned} \tag{10}$$

Here, the weighting parameter α will be increased during optimization and the solution to (10) will converge to that of (9) as $\alpha \to \infty$. α can be initialized as some α_{min} and gradually increased by a scale factor k until it is larger than an upperbound α_{max}. For a fixed α, we can minimize (10) by alternating between updating $\mathbf{X}$ and $\mathbf{Y}$ while fixing the other. In practice, it is not necessary to achieve convergence for each fixed α and one iteration of updating will be sufficient.

Given a fixed $\mathbf{Y}$ and α, the optimal $\mathbf{X}$ satisfies

$$\left(\frac{1}{2\sigma^2}\mathbf{H}^T\mathbf{H}+\beta\mathbf{I}+\alpha\nabla^T\nabla\right)\mathbf{X}=\frac{1}{2\sigma^2}\mathbf{H}^T\mathbf{Z}+\beta\mathbf{U}+\alpha\nabla^T\mathbf{Y} \tag{11}$$

where $\mathbf{U}=\sum_i \mathbf{R}_i\mathbf{\Phi}\mathbf{w}_i$ and $\mathbf{I}$ is a unit matrix. There is no need to directly solve the linear equation (11). In fact, since both $\mathbf{H}$ and ∇ are matrix form of 2D convolution operators, they can be diagonalized by 2D Fast Fourier Transform (FFT). If we define $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{u}}$ as the 2D form of vectors $\mathbf{X}, \mathbf{Y}, \mathbf{U}$ and $\mathbf{h}, \mathbf{d}$ as the 2D convolution kernel corresponding to $\mathbf{H}, \nabla$, we have

$$\tilde{\mathbf{x}}=\mathcal{F}^{-1}\left(\frac{\frac{1}{2\sigma^2}\mathcal{F}\left(\mathbf{h}\right)^*\circ\mathcal{F}\left(\tilde{\mathbf{z}}\right)+\beta\mathcal{F}\left(\tilde{\mathbf{u}}\right)+\alpha\mathcal{F}\left(\mathbf{d}\right)^*\circ\mathcal{F}\left(\tilde{\mathbf{y}}\right)}{\frac{1}{2\sigma^2}\mathcal{F}\left(\mathbf{h}\right)^*\circ\mathcal{F}\left(\mathbf{h}\right)+\beta+\alpha\mathcal{F}\left(\mathbf{d}\right)^*\circ\mathcal{F}\left(\mathbf{d}\right)}\right) \tag{12}$$

where $\mathcal{F}\left(\cdot\right)$ is 2D FFT operator, $*$ and $\circ$ denotes complex conjugate and component-wise multiplication. Thus, updating $\mathbf{X}$ only involves 2D FFT and component-wise multiplication (the division in (12) is also component-wise).

When $\mathbf{X}$ and α are fixed, the optimal $\mathbf{Y}$ can be found by solving the following problem for each component:

$$\min_{\left(\mathbf{Y}\right)_j} \alpha\|\left(\nabla\mathbf{X}\right)_j-\left(\mathbf{Y}\right)_j\|_2^2+\mu|\left(\mathbf{Y}\right)_j|^{2/3}. \tag{13}$$

The solution to (13) can be obtained analytically by just solving a quartic function according to [2].

3.3. Summary of algorithm

We describe the proposed deconvolution method in Algorithm 1. To solve the ℓ_1-ℓ_2 optimization in (8) we apply the Least Angle Regression (LARS) method [11]. In practice we set $\alpha_{min}=1$, $k=2\sqrt{2}$ and $\alpha_{max}=2^{20}$. When $\mathbf{X}$ is initialized by a Wiener estimate [3], the algorithm will converge within very few iterations. Our experiments show that the algorithm is very robust to a wide range of possible initializations, including the blurred images themselves.

Algorithm 1 Image deconvolution based on our joint model

INPUT: blurred image $\mathbf{Z}$, updating parameters μ, β, λ
parameters regarding to α: $\alpha_{min}, \alpha_{max}, k$
Initialize $\mathbf{X}$.
repeat
  Update the coefficients for each patch $\mathbf{w}_i$ according to (8).
  $\alpha = \alpha_{min}$.
  while $\alpha < \alpha_{max}$ **do**
    Update $\mathbf{Y}$ according to (13) analytically according to [2].
    Update $\mathbf{X}$ according to (12).
    $\alpha = k\alpha$.
  end while
until convergence
OUTPUT: deconvolved image $\mathbf{X}$

Table 1. Experiment settings with different blur kernels and different values of noise variance σ^2 for pixel values in $[0, 255]$.

Test	Blur kernel h	σ^2
1	Uniform 9×9 kernel	0.308
2	$h_{i,j}=\left(1+i^2+j^2\right)^{-1}$, for $i,j=-7,\cdots,7$	2
3	$h_{i,j}=\left(1+i^2+j^2\right)^{-1}$, for $i,j=-7,\cdots,7$	8
4	Separable kernel with weights $[1,4,6,4,1]/16$ in both directions	49

4. EXPERIMENTAL RESULTS

In the experiments, we work on gray-scale images with intensity values normalized to the range $[0, 1]$. We set the patch size $N = 12\times12$ and tune the parameters as $\lambda = 2.5$, $\mu = 2.5$, $\beta = 200$ empirically for the best average performance. The corresponding dictionary $\mathbf{\Phi}$ is learned from randomly extracted patches from the Berkeley Segmentation database [12] using the online dictionary learning algorithm in [10]. We compare the proposed method with two state-of-the-art algorithms in standard test settings for deconvolution. [13] is based on ℓ_0 sparse optimization in tight frames, while [2] applies only sparse gradient priors. The test settings are shown in Table 1. Table 2 shows the comparison of our test results on two frequently used test images *House* 256×256 and *Cameraman* 256×256 against the other two state-of-the-art methods in terms of ISNR (increment in PSNR) and SSIM (structural similarity) index of the restored image. We also run the three algorithms on 100 images randomly extracted from the PASCAL Visual Object Classes Challenge (VOC) 2007 database to compare the average deconvolution performance. Note that the parameters in the three methods are uniformly tuned for all of the images and fixed during the experiments, and not adapted to any particular test image.

In our experiments, our method outperforms both [2] and [13] in terms of SSIM values in all cases. Our method also outperforms [13] in ISNR for the VOC 2007 database. Compared with [2] that only uses the sparse gradient priors, the proposed algorithm obtains better deconvolution results by combining the sparse gradient priors with learned sparse coding of image patches. In MATLAB simulation, the proposed method takes an average time of 16s for a 256×256 image with an Intel Core2 Duo CPU @ 2.26GHz and 4GB RAM, while [13] and [2] take 15s and 2s respectively on average.

Fig.1 shows a visual comparison between the proposed method and the method in [13] on *House* image in test setting 1. Although our method provides a slightly lower ISNR value, it keeps more brick

Table 2. The ISNR (in dB) and SSIM index of different methods in the 4 standard test settings as described in Table 1

256×256 *House*				
		[2]	[13]	Proposed
Test 1	ISNR	10.06	**10.74**	10.57
	SSIM	0.8894	0.8918	**0.8934**
Test 2	ISNR	7.82	8.40	**8.49**
	SSIM	0.8701	0.8737	**0.8799**
Test 3	ISNR	6.77	**7.12**	7.11
	SSIM	0.8582	0.8566	**0.8606**
Test 4	ISNR	4.34	4.55	**4.57**
	SSIM	0.8619	0.8588	**0.8620**
256×256 *Cameraman*				
		[2]	[13]	Proposed
Test 1	ISNR	8.65	**9.10**	8.98
	SSIM	0.8812	0.8797	**0.8871**
Test 2	ISNR	7.55	7.70	**7.76**
	SSIM	0.8772	0.8724	**0.8851**
Test 3	ISNR	5.37	**5.55**	5.43
	SSIM	0.8438	0.8371	**0.8464**
Test 4	ISNR	2.59	**2.94**	2.71
	SSIM	0.8454	0.8408	**0.8484**
the VOC 2007 database				
		[2]	[13]	Proposed
Test 1	ISNR	7.50	7.72	**7.79**
	SSIM	0.8614	0.8634	**0.8657**
Test 2	ISNR	6.69	6.80	**7.12**
	SSIM	0.8799	0.8837	**0.8860**
Test 3	ISNR	5.11	5.12	**5.23**
	SSIM	0.8343	0.8321	**0.8353**
Test 4	ISNR	2.67	2.58	**2.77**
	SSIM	0.8425	0.8440	**0.8441**

textures on the wall of the house, by benefiting from the learned overcomplete dictionary for image patches (the difference might only be observed in electronic copy when zoomed in).

5. CONCLUSIONS

We have proposed a method to globally model image priors based on the sparse representation of small patches over a learned overcomplete dictionary and the sparse distribution of spatial gradients. The sparse representation of image patches provides well matching with various textures, while the sparse gradient prior guarantees smoothness of the image. We also developed an iterative algorithm for image deconvolution with the proposed joint model, with convergence in only a few iterations. In our experiments, our method generally outperforms the best existing deconvolution algorithms. The proposed model and algorithm can be extended to more general image restoration problems like denoising and super-resolution.

6. REFERENCES

[1] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, 2007.

[2] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," *Advances in Neural Information Processing Systems*, vol. 22, pp. 1–9, 2009.

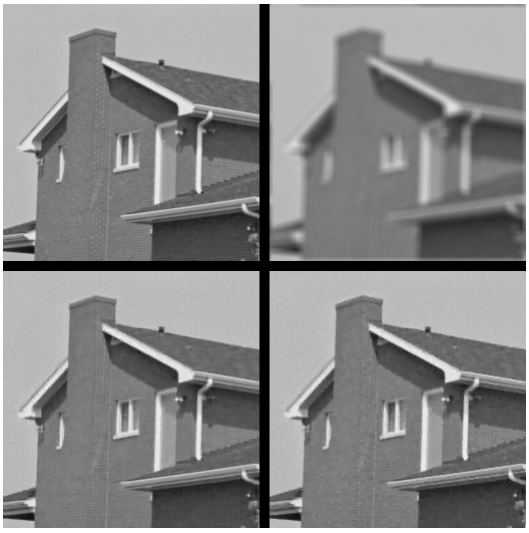

Fig. 1. Visual comparison of *House* image in test setting 1. From left to right and from top to bottom: original image, blurred image, image deconvolved by L0-AbS [13] and image deconvolved by the proposed method.

[3] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.

[4] B.A. Olshausen and D.J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[5] M. Aharon, M. Elad, A. Bruckstein, and Y. Katz, "K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, pp. 4311–4322, 2006.

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. on Computer Vision*, 2009.

[7] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[8] Y. Lou, A. Bertozzi, and S. Soatto, "Direct sparse deblurring," Tech. Rep. 09-15, CAM-UCLA, 2009.

[9] C. Jia and B.L. Evans, "MATLAB code for image deconvolution using joint sparsity," in *http://users.ece.utexas.edu/˜bevans/papers/2011/sparsity/*.

[10] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. on Machine Learning*, 2009.

[11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–451, Apr. 2004.

[12] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int. Conf. on Computer Vision*, July 2001.

[13] J. Portilla, "Image restoration through L0 analysis-based sparse optimization in tight frames," in *Proc. IEEE Int. Conf. on Image Processing*, 2009, pp. 3909–3912.