

BLIND IMAGE DECONVOLUTION USING STUDENT'S-T PRIOR WITH OVERLAPPING GROUP SPARSITY

In S. Jeon Deokyoung Kang Suk I. Yoo

Dept. of Computer Science and Engineering, Seoul National University, Seoul, South Korea

ABSTRACT

In this paper, we solve blind image deconvolution problem that is to remove blurs from a signal degraded image without any knowledge of the blur kernel. Since the problem is ill-posed, an image prior plays a significant role in accurate blind deconvolution. Traditional image prior assumes coefficients in filtered domains are sparse. However, it is assumed here that there exist additional structures over the sparse coefficients. Accordingly, we propose new problem formulation for the blind image deconvolution, which utilize the structural information by coupling Student's-t image prior with overlapping group sparsity. The proposed method resulted in an effective blind deconvolution algorithm that outperforms other state-of-the-art algorithms.

Index Terms— image restoration, blind deconvolution, Bayesian, Student's-t prior, group sparsity, MM algorithm

1. INTRODUCTION

Image deconvolution is the problem of restoring an image x from its blurred and noisy version y . Generally, the image y is modeled as

$$y = h \otimes x + \varepsilon, \quad (1)$$

where \otimes denotes two-dimensional convolution, h is blur kernel, and ε is noise term. Since there exists infinitely many solution for x , (1) is an ill-posed problem [1, 2]. Hence, a regularization reflecting our prior knowledge is necessary to be imposed on the image x to obtain a meaningful solution. The regularization is generally embedded by assigning prior distribution $p(x)$ in the Bayesian formulation of the problem.

The choice of the image prior is varying, but the most popular one is the sparsity-enforcing prior. It is well-known that when high-pass filters are applied to natural images, the resulting coefficients are sparse [3, 4]; i.e., most of the coefficients are zero or very small while only a small number of them are large (e.g., at the edges). This important characteristic has been utilized in many image deconvolution algorithms. Fergus et al. [5] introduced a mixture-of-Gaussians (MoG) prior with a filtered image representation and showed the proposed approach was practical. After the success of his work, many researchers have subsequently suggested other

kinds of the sparsity-enforcing image prior such as total variation [6, 7], hyper-Laplacian [8], and Student's-t [9].

When the blur kernel h in (1) is unknown in addition to the unknown image x , it becomes a “blind” image deconvolution problem. This is much difficult to solve than the non-blind image deconvolution problem since there exist infinitely many possible combinations of x and h making it severely ill-posed. Levin et al. [10] reveals that the conventional sparsity-enforcing priors [5, 6, 7, 8] confront a limitation in the case of blind image deconvolution; they indicate that these priors favor the blurred image over the correct one because the blurred image often has more zero coefficients (i.e. more sparse) than the clear image in high-pass filtered domains.

To avoid this “no-blur” solution, Levin et al. [11] employ a marginal likelihood optimization. Krishnan et al. [12] suggest a normalized sparsity measure to mitigate this problem. Moreover, Wipf and Zhang [13] emphasize the power of image prior that can discriminate a good sharp image from blurred images is critical to recovering the unknown image. Both Babacan [14] and Perrone [15] are good examples of this approach. They [13, 14, 15] all utilized non-convex image priors to strongly promote the sparsity of the signal, regardless of the natural statistics, and achieved state-of-the-art blind image deconvolution result.

To overcome the limitation of the traditional image priors, there is another growing interest for structured sparsity [16, 17, 18, 19]. In fact, all the image priors presented in [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] assumes the coefficient in the filtered domains is sparse; however, large values of the coefficients generally do not occur in isolation. Hence authors in [16, 17, 18, 19] assumes the coefficients exhibits a simple form of structure that is called structured sparsity. Liu et al. [20] adapt the group sparsity regularizer to recover a noise corrupted image, and it is proven to be very effective in alleviating staircase effects. Shi et al. [21] also shows a hyper-Laplacian constrained with overlapping group sparsity leads to a good image deconvolution result.

In this paper, we utilize the structured sparsity [16, 17, 18, 19] to solve the blind image deconvolution problem. We presented new problem formulation for the blind image deconvolution, which incorporates conventional Student's-t image prior [9] with overlapping group sparsity regularizer [19]. The proposed problem formulation resulted in an

effective blind image deconvolution algorithm that outperforms recently introduced state-of-the-art algorithms such as [11, 14, 15]. We verified this result in the experimental section.

The rest of this paper is organized as follows. Section 2 describes our modeling of the blind deconvolution problem. The inference algorithm is outlined in section 3. In Section 4 we present the experimental results. Section 5 concludes the paper.

2. PROBLEM FORMULATION

Since we are in the discrete domain, convolution in (1) is equivalent to vector-matrix multiplication. Then we can rewrite the model as

$$y = Hx + \varepsilon = Xh + \varepsilon. \quad (2)$$

y , x , and ε are lexicographically arranged n^2 dimensional vector. H and X are the two-dimensional convolution matrices obtained from the kernel $h \in \mathbb{R}^{k^2}$ and image $x \in \mathbb{R}^{n^2}$ respectively. We further assume the noise ε is i.i.d Gaussian with variance σ^2 , then we obtain following observation model:

$$p(y|x, h) = \mathcal{N}(y|Hx, \sigma^2 I). \quad (3)$$

From Bayesian perspective, the goal of blind image deconvolution is to infer the unknown (latent) variables x and h . Estimating these variables is usually done by maximizing a posterior distribution: $p(x, h|y) \propto p(y|x, h)p(x)p(h)$ where $p(y|x, h)$ is likelihood defined as (3). $p(x)$ and $p(h)$ are the prior distributions for the unknown image and the unknown blur kernel respectively. This approach is commonly referred to as the maximum a posterior (MAP) estimation.

2.1. Kernel Prior

For the kernel prior, we choose the flat prior: $p(h) \propto 1$. Considering the second part (i.e. $y = Xh + \varepsilon$) in (2), we have n^2 observations and aim at estimating k^2 coefficients while solving the kernel h . Since the image size is usually much larger than the kernel size (i.e. $k^2 \ll n^2$), n^2 observations should be sufficient to obtain a good kernel estimate. Based on the fact, many authors [13, 14, 15] also used the flat prior on the kernel, enforcing only its non-negativity and normalization constraints.

2.2. Student's-t Image Prior

The image prior $p(x)$ is based on m filtered versions of the image: $g_m = F_m x$, where F_m are two dimensional convolution matrix obtained from high-pass filters: $\{f_m\}_{m=1}^M$. Specifically, we use the first-order differences between 4 local neighbors.

Assuming that each pixel $g_{m,i}$ at index i follows Gaussian distribution with distinct precision $\gamma_{m,i}$, and the precision is

Gamma random variable with the shape parameter α and the scale parameter β , we can define a hierarchical joint distribution for each $(g_{m,i}, \gamma_{m,i})$ as follows:

$$p(g_{m,i}, \gamma_{m,i}) \propto p_{m,i}(g_{m,i}|\gamma_{m,i})p(\gamma_{m,i}) \\ \propto \mathcal{N}(g_{m,i}|0, \gamma_{m,i}^{-1})\text{Gamma}(\gamma_{m,i}|\alpha, \beta). \quad (4)$$

The marginalization of (4) with respect to $r_{m,i}$ is equivalent to Student's-t distribution. Hence, this hierarchical prior closely resembles the Student's-t prior enforcing the sparseness of the image pixels in the filtered domains [9]. Also, notice that new auxiliary variable $\gamma_{m,i}$ is introduced, which will be estimated jointly.

By multiplying (4) across all the spatial indexes i and the filter indexes m , we can obtain Student's t image prior as follows:

$$p(x, \gamma) \propto p(x|\gamma)p(\gamma) \\ \propto \prod_{m=1}^M \mathcal{N}(x|0, (F_m^T \text{diag}\{\gamma_m\}F_m)^{-1}) \\ \times \prod_{m=1}^M \prod_{i=1}^N \text{Gamma}(\gamma_{m,i}|\alpha, \beta) \quad (5)$$

with $\gamma_m = (\gamma_{m,1}, \dots, \gamma_{m,N})^T$.

MAP estimation is equivalent to minimizing the negative log posterior. Accordingly, with the equation (3), (5), and $p(h) \propto 1$, we get the following optimization problem for the blind image deconvolution:

$$\min_{x, \gamma, h} -\log p(x, \gamma, h|y) \\ = \min_{x, \gamma, h} -\log(p(y|x, h)) - \log(p(x, \gamma)) - \log(p(h)) \quad (6) \\ = \min_{x, \gamma, h} ||Hx - y||^2 + \lambda_1 \psi(x, \gamma),$$

where $\lambda_1 = 1/\sigma^2$ is a regularization parameter, and

$$\psi(x, \gamma) = \sum_{m=1}^M x^T F_m^T \text{diag}\{\gamma_m\} F_m x \\ + 2 \sum_{m=1}^M \sum_{i=1}^N ((1 - \alpha) \log \gamma_{m,i} + \beta \gamma_{m,i}). \quad (7)$$

$\psi(x, \gamma)$ is a regularization term obtained from Student's-t image prior promoting the sparsity of coefficients in the filtered domains. However, it does not take account the structural information among the coefficients.

2.3. Overlapping Group Sparsity

To capture the structural information among the coefficients, we define two-dimensional $W \times W$ points group in the two-

dimensional signal s as follows:

$$\tilde{s}_{(i,j),W} = \begin{bmatrix} s_{(i-m_1,j-m_1)} & \cdots & s_{(i-m_1,j+m_2)} \\ s_{(i-m_1+1,j-m_1)} & \cdots & s_{(i-m_1+1,j+m_2)} \\ \vdots & \ddots & \vdots \\ s_{(i+m_2,j-m_1)} & \cdots & s_{(i+m_2,j+m_2)} \end{bmatrix} \in \mathbb{R}^{W \times W}, \quad (8)$$

with $m_1 = \lfloor \frac{W-1}{2} \rfloor$, $m_2 = \lfloor \frac{W}{2} \rfloor$, where $\lfloor x \rfloor$ denotes the floor function, and W is the window size. Hence, $\tilde{s}_{(i,j),W}$ is a group of $W \times W$ contiguous samples centered at (i, j) .

By stacking the columns of $\tilde{s}_{(i,j),W}$, a vector $s_{(i,j),W}$ is obtained: $s_{(i,j),W} = \tilde{s}_{(i,j),W}(:, :)$. Then, the overlapping group sparsity (OGS) functional [20, 21] is

$$\varphi_{\text{OGS}}(s) = \sum_{i,j=1}^n \|s_{(i,j),W}\|_2. \quad (9)$$

φ_{OGS} takes account all the overlapping groups of pixels on the spatial domain of the signal s . With (9), we define OGS regularization term for the blind deconvolution problem:

$$\phi(x) = \sum_{m=1}^M \varphi_{\text{OGS}}(g_m), \quad (10)$$

where $g_m = F_m x$. If $W = 1$, ϕ is the commonly used anisotropic TV prior. In this sense, ϕ is also referred as generalized total variation [22].

Therefore, we present a novel problem formulation, which incorporates Student's-t image prior and the overlapping group sparsity, to solve the blind image deconvolution problem as follows:

$$\min_{x,\gamma,h} (R(x) = \|Hx - y\|^2 + \lambda_1 \psi(x, \gamma) + \lambda_2 \phi(x)) \quad (11)$$

where the functional $\psi(x, \gamma)$ and $\phi(x)$ is given by (7) and (10) respectively. λ_1 and λ_2 are the regularization parameters.

3. INFERENCE ALGORITHM

We use majorization-minimization (MM) as in [20, 19, 22] to derive a computationally efficient algorithm to solve the problem (11). To find a majorizer of $R(x)$ in (11), we first find a majorizer of $\varphi_{\text{OGS}}(v)$ in (9). To this end, note that

$$\frac{1}{2\|u\|_2} \|v\|_2^2 + \frac{1}{2} \|u\|_2 \geq \|v\|_2 \quad (12)$$

for all v and $u \neq 0$ with equality when $u = v$. Substituting each group of $\varphi_{\text{OGS}}(v)$ into (12) and summing them, we get a majorizer of $\varphi_{\text{OGS}}(v)$

$$P(v, u) = \frac{1}{2} \sum_{i,j=1}^n \left(\frac{1}{\|u_{(i,j),W}\|_2} \|v_{(i,j),W}\|_2^2 + \|u_{(i,j),W}\|_2 \right) \quad (13)$$

with

$$P(v, u) \geq \varphi_{\text{OGS}}(v), \quad P(u, u) = \varphi_{\text{OGS}}(u) \quad (14)$$

provided $\|u_{(i,j),W}\|_2 \neq 0$ for all i, j . With a simple calculation, $P(v, u)$ can be rewritten as

$$P(v, u) = \frac{1}{2} v^T \Lambda(u) v + C, \quad (15)$$

where C is constant that dose not depend on v , and $\Lambda(u)$ is a diagonal matrix with each diagonal component

$$[\Lambda(u)]_{l,l} = \sum_{i,j=-m_1}^{m_2} \left[\sum_{w_1, w_2=-m_1}^{m_2} |u_{(r-i+k_1, t-j+k_2)}|^2 \right]^{-\frac{1}{2}} \quad (16)$$

with $l = (t-1)n + r$, and $r, t = 1, 2, \dots, n$.

Substituting each $\varphi_{\text{OGS}}(g_m)$ in (10) into (14), a majorizer of $R(x)$ in (11) can be obtained by

$$\begin{aligned} G(x, x') &= \|Hx - y\|^2 + \lambda_1 \psi(x, \gamma) + \lambda_2 \phi'(x, x') \\ &\geq R(x) = \|Hx - y\|^2 + \lambda_1 \psi(x, \gamma) + \lambda_2 \phi(x), \end{aligned} \quad (17)$$

where

$$\begin{aligned} \phi'(x, x') &= \sum_{m=1}^M P(g_m, g'_m) = \sum_{m=1}^M x^T F_m^T \Lambda(F_m x') F_m x, \end{aligned} \quad (18)$$

where x' is the estimation of x at the previous iteration.

MM algorithm solve the problem (11) by iteratively minimizing $G(x, x')$ in (17). Since the first term in $G(x, x')$ is the simple quadratic, and the second and the third term are also differentiable, we can summarize an optimization algorithm for solving the minimization problem (11) as follows:

Algorithm 1 for solving the minimization problem (11)

Inputs: $y, \{F_m\}_{m=1}^M, \alpha, \beta, \lambda_1, \lambda_2, W$, and max-iter L .

Initialization: $x^{(0)} = y, h^{(0)} = h_0, \gamma_{m,i}^{(0)} = 1, l = 0$.

Iteration:

1. Obtain convolution matrix H from $h^{(l)}$.
2. $g_m^{(l)} = F_m x^{(l)}$.
3. Compute $\Lambda(g_m^{(l)})$ according to (16).
4. $\gamma_{m,i}^{(l+1)} = (\alpha + 1/2)/(\beta + (1/2)(g_{m,i}^{(l)})^2)$.
5. $x^{(l+1)} = (H^T H + \sum_{m=1}^M F_m^T (\lambda_1 * \text{diag}\{\gamma_m^{(l)}\} + \lambda_2 * \Lambda(g_m^{(l)})) F_m)^{-1} H^T y$.
6. Obtain convolution matrix X from $x^{(l)}$.
7. $h^{(l+1)} = (X^T X)^{-1} X^T y$.
8. $l = l + 1$.

Until $l < L$.

Output: $x^{(L)}, h^{(L)}$.

	ker01		ker02		ker03		ker04		ker05		ker06		ker07		ker08	
Img01	42.49	22.85	43.95	24.87	31.50	16.65	105.58	51.12	26.62	17.12	28.29	17.82	49.07	20.87	58.55	26.14
	26.07	25.45	33.21	31.08	17.07	17.59	60.31	46.52	17.37	14.39	29.91	17.95	32.49	30.54	41.06	37.99
Img02	55.61	49.42	60.51	54.6	49.98	36.06	102.26	74.03	29.35	35.42	29.44	20.14	57.25	38.23	69.12	55.23
	42.29	34.91	33.83	35.25	31.66	33.47	45.18	54.54	38.33	23.64	76.32	31.36	32.28	33.82	34.66	37.55
Img03	35.26	28.66	42.45	43.96	17.99	15.36	67.39	70.82	17.18	13.94	21.16	26.75	27.18	24.35	37.74	27.83
	19.43	19.25	19.76	23.58	12.90	15.26	25.17	28.26	11.86	12.37	9.65	13.13	11.98	16.59	25.52	27.63
Img04	79.76	75.56	136.58	201.12	52.11	24.01	97.35	261.48	38.20	23.05	76.19	65.30	101.42	120.18	118.15	424.02
	45.68	42.27	75.88	71.99	19.51	22.43	126.19	52.47	17.39	20.27	39.69	30.62	54.55	56.94	58.35	57.95

Table 1. SSD error of 32 test images, achieved by (Top Left) Levin et al. [11], (Top Right) Babacan et al. [14], (Bottom Left) Perrone et al. [15], and (Bottom Right) the proposed method with the same non-blind deblurring algorithm [23].

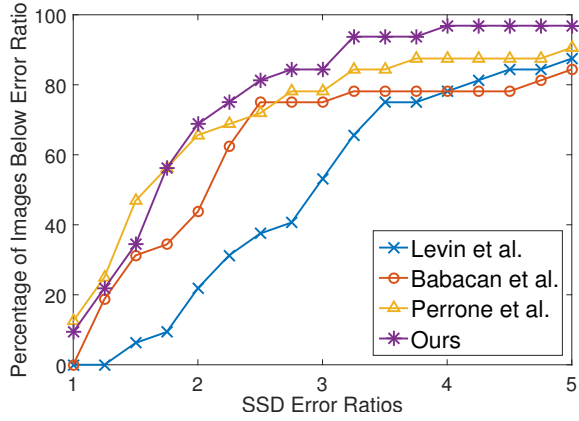


Fig. 1. Cumulative histogram of SSD error ratio

4. EXPERIMENT

We evaluated the proposed algorithms and the other state-of-the-art algorithms [11, 14, 15] on the dataset from Levin et al. [10]. The dataset is made of 4 images of size 255×255 pixels blurred with 8 different blur kernels, and it is provided with ground truth sharp images and blur kernels.

In practice, we employed a multiscale approach to deal with the large blur support problem [5]. The input image and the blur are down sampled at each level by $\sqrt{2}$, and the parameter, λ_1 and λ_2 , are divided by the number 2. Then the algorithm 1 was applied at each scale. The number of levels of the pyramid is computed such that at the top level the blur kernel has a support of 3 pixels. We used the fixed parameter values for all the tests, $\lambda_1 = 4.5e-5$, $\lambda_2 = 5e-6$, $\alpha = 1e-18$, $\beta = 1/1700$, $W = 3$, and 4500 iterations for each pyramid level. The parameter values have been found experimentally. For the other algorithms, we used the parameters provided by the authors.

First, we measured the sum of squared distance (SSD) between the recovered images and the ground truth images in Table. 1. To measure the effectiveness of estimated blur kernels, the SSD ratio proposed in [10] was computed. The SSD ratio is defined by $\sum_{i=1}^N (x_i^L - x_i^G)^2 / \sum_{i=1}^N (x_i^H - x_i^G)^2$, where x^G is the ground truth image, x^L is the image obtained



Fig. 2. Blind deconvolution results on Img04 with ker04.

by solving a non-blind deconvolution with the estimated blur, and x^H is the image obtained by solving a non-blind deconvolution with the ground truth blur. For all the tests, we used the non-blind deconvolution algorithm from Levin et. al [23] with $\lambda = 10^{-3}$.

In Fig.1, we plot the cumulative histogram of SSD ratios (e.g., bin=3 counts the percentage of test examples achieving SSD ratio below 3). Our algorithm performs SSD ratio equal to 2 for more than 60% of the images, clearly outperforming the method from Levin et al. [11] and Babacan et al. [14]. Our method is on par with high performing blind image deconvolution algorithm (logMM) from Perrone et al. [15]. Lastly, Fig. 2 presents some of the estimated images and blur kernels from the experiment.

5. CONCLUSION

In this paper, we presented a blind image deconvolution algorithm employing structural information among the sparse coefficients. Specifically, a novel problem formulation combining Student's-t image prior and overlapping group sparsity is proposed. Its effectiveness has been demonstrated by the experiment. Future work may include faster approximation of the algorithm using ADMM [24] and extensive analysis on the effect of the group size in OGS term.

6. REFERENCES

- [1] N. B. Karayiannis and A. N. Venetsanopoulos, "Regularization theory in image restoration-the stabilizing functional approach," *IEEE Trans. Acoustic Speech Signal Processing.*, vol. 38, pp. 1155–1179, 1990.
- [2] A. Tikhonov and V. Arsenin, *Solution of ill-posed problems*, Winston, Washington DC, 1977.
- [3] B.A. Olshausen and D.J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–608, 1996.
- [4] S.G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Analysis and Machine Intelligence.*, vol. 11, pp. 674–694, 1989.
- [5] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graphics.*, vol. 25, pp. 787–794, 2006.
- [6] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graphics.*, 2008.
- [7] S. cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graphics.*, vol. 28, 2009.
- [8] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," in *NIPS*, 2009.
- [9] D. G. Tzikas, A. C. Kikas, and N. P. Galatsanos, "Variational bayesian kernel-based blind image deconvolution with student's-t priors," *IEEE Trans. Image Processing*, vol. 18, pp. 753–764, 2009.
- [10] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *CVPR*, 2009.
- [11] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *CVPR*, 2011.
- [12] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *CVPR*, 2011.
- [13] D. Wipf and H. Zhang, "Revisiting bayesian blind deconvolution," *Journal of Machine Learning Research*, vol. 15, pp. 3595–3634, 2014.
- [14] S. D. Babacan, R. Molina, and M. N. Do, "Bayesian blind deconvolution with general sparse image priors," in *ECCV*, 2012.
- [15] D. Perrone and P. Favaro, "A logarithmic image prior for blind deconvolution," *Int. J. Comput Vis.*, 2015.
- [16] G. Peyre and J. Fadili, "Group sparsity with overlapping partition functions," in *In Proc. European Sig. Image Proc. Conf. (EUSIPCO)*, 2011.
- [17] M. Figueiredo and J. Bioucas-Dias, "An alternating direction algorithm for (overlapping) group regularization," in *In Signal Process. Adaptive Sparse Structured Representation*, 2011.
- [18] F. Bach, R. Jenatton, J. Mairl, and G. Obozinski, "Structured sparsity through convex optimization," *Stat. Sci.*, vol. 27, pp. 450–468, 2012.
- [19] I.W. Selesnick and P.Y. Chen, "Total variation denoising with overlapping group sparsity," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013.
- [20] J. Liu, T. Huang, I.W. Selesnick, X.Lv, and P. Chen, "Image restoration using total variation with overlapping group sparsity," *Inf. Sci.*, vol. 295, pp. 232–246, 2015.
- [21] M. Shi, T. Han, and S. Liu, "Total variation image restoration using hyper-laplacian prior with overlapping group sparsity," *Signal Processing*, vol. 126, pp. 65–76, 2016.
- [22] Ivan W. Selesnick, "Generalized total variation," in *Signal Processing Letters. IEEE*, 2015, vol. 22.
- [23] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *Trans. Graph. ACM*, 2007, vol. 26(3).
- [24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Foundations and Trends in Machine Learning, Now Publishers Inc. Hanover, MA, USA, 1st edition, 2011.