QUASI-MAXIMUM LIKELIHOOD BLIND DECONVOLUTION OF IMAGES ACQUIRED THROUGH SCATTERING MEDIA

Alexander M. Bronstein, Michael M. Bronstein, Michael Zibulevsky, and Yehoshua Y. Zeevi

Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel.

ABSTRACT

We address the problem of restoration of images obtained through a scattering medium. We present an efficient quasi-maximum likelihood blind deconvolution approach based on the fast relative Newton algorithm and optimal distribution-shaping approach (sparsification), which allows to use simple and convenient sparsity prior for a wide class of images. Simulation results prove the efficiency of the proposed method.

1. INTRODUCTION

Imaging through scattering media plays an important role in optical tomography [1] and other medical applications. Multiple scattering results in image blurring, which can be modelled as a result of convolution with and unknown kernel. Many techniques are therefore based on *blind deconvolution* are commonly used [1, 2]. The advantage of such approaches is that they do not require expensive hardware such as nanosecond gating devices.

According to the convolution model, the observed sensor image X is created from the *source image* S passing through a linear shift-invariant system described by the impulse response W,

$$X = (W * S). \tag{1}$$

We assume that the action of W is invertible (at least approximately), i.e. there exists some other kernel H such that $(W * H)_{mn} \approx \delta_{mn}$. This assumption holds well since blurring kernels resulting from scattering are usually Lorenzian-shaped and their inverse can be approximated by small FIR kernels. The aim of blind deconvolution is to find such *deconvolution* (*restoration*) kernel H that produces an estimate $\tilde{S}$ of S up to integer shift and scaling factor:

$$\tilde{S}_{mn} = (H * X)_{mn} \approx c \cdot S_{m-\Delta_M, n-\Delta_N}, \tag{2}$$

or equivalently, the *global system response* should be

$$G_{mn} = (W * H)_{mn} \approx c \cdot \delta_{m-\Delta_M, n-\Delta_N}. \tag{3}$$

This research has been supported by the HASSIP Research Network Program HPRN-CT-2002-00285, sponsored by the European Commission.

Unlike approaches estimating the image and the blurring kernel, we estimate the restoration kernel only, which results in a much lower dimensionality of the problem. In this work, we present a quasi-maximum likelihood blind deconvolution algorithm, which generalizes the fast relative Newton algorithm previously proposed for blind source separation [3]. We also propose optimal distribution-shaping approach (sparsification), which allows to use simple and convenient sparsity prior for a wide class of images. For technical details see [4, 5].

2. QUASI-ML BLIND DECONVOLUTION

Denote by $Y = H * X$ the source estimate and let us assume that S is zero-mean i.i.d. In the zero-noise case, the normalized minus log likelihood function of the observed signal X, given the restoration kernel H, is

$$L(H; X) = -\frac{M_X N_X}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log \left| \hat{H}(\xi, \eta) \right| d\xi d\eta + \sum_{m,n} \varphi(Y_{mn}), \tag{4}$$

where $\varphi(\cdot) = -\log p_S(\cdot)$, $p_S(\cdot)$ stands for the source probability density function (PDF), $M_X N_X$ is the observation sample size, and $\hat{H}(\xi, \eta)$ denotes the Fourier transform of H_{mn}. We will henceforth assume that H is a FIR, supported on $[-M, ..., M] \times [-N, ..., N]$ and denote $K_M = 2M + 1$, $K_N = 2N + 1$. Cost functions similar to (4) were also obtained in the 1D case using negative joint entropy and information maximization considerations [6].

2.1. The choice of $\varphi(\cdot)$

Source images arising in most applications usually have non-log-concave, multi-modal distributions. These are difficult to model and are not suitable for optimization. However, consistent estimator of S can be obtained by minimizing $L(H; X)$ even when $\varphi(\cdot)$ is not exactly equal to $-\log p_S(\cdot)$. Such *quasi-ML estimation* has been shown to be practical in instantaneous blind source separation [7, 3, 8] and blind deconvolution of time signals [4]. For example, when the source is super-Gaussian (sparse), a smooth approximation of the absolute value function is a good choice for $\varphi(\cdot)$ [9, 10]. Although natural images are usually far from being sparse, they can be transformed into a space of a sparse representation. We will therefore focus our attention on modelling super-Gaussian distributions using a family of convex smooth functions

$$\varphi_\lambda(t) = |t| - \lambda \log\left(1 + \frac{|t|}{\lambda}\right) \quad (5)$$

with λ a positive smoothing parameter; $\varphi_\lambda(t) \to |t|$ as $\lambda \to 0^+$.

2.2. Approximation of the log-likelihood function

In practice it is difficult to evaluate the first term of $L(H;X)$ containing the integral; however, it can be approximated with any desired accuracy by [4]

$$\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \log\left|\hat{H}(\xi,\eta)\right| d\xi d\eta \approx \frac{1}{M_F N_F}\sum_{k=0}^{M_F}\sum_{l=0}^{N_F} \log\left|\hat{H}_{kl}\right|, \quad (6)$$

where $\hat{H}_{kl} = \hat{H}\left(\frac{2\pi k}{M_F}, \frac{2\pi l}{N_F}\right)$ are the 2D DFT coefficients of H_{mn} [4], zero-padded to $M_F \times N_F$. The approximation error vanishes as M_F, N_F grow to infinity. M_F and N_F should be chosen as integer powers of 2, which allows the use of FFT. For convenience, we will denote

$$f_1(H) = \sum_{k=0}^{M_F}\sum_{l=0}^{N_F} \log\left|\hat{H}_{kl}\right|^2 \quad (7)$$

$$f_2(Y) = \sum_{m,n} \varphi(Y_{mn}), \quad (8)$$

and, accordingly, the approximate normalized minus-log likelihood function becomes

$$L(H;X) = -\frac{f_1(H)}{2M_F N_F} + \frac{f_2(Y)}{M_X N_X}. \quad (9)$$

2.3. Gradient and Hessian of $L(H;X)$

The optimization algorithm discussed in Section 2.4 requires knowledge of the gradient and the Hessian of $L(H;X)$. For convenience, we will parse the variables column-wise into a $K_M K_N \times 1$ vector $h = \mathrm{vec}(H)$, and define the gradient and the Hessian of $L(H;X)$ as a $K_M K_N \times 1$ vector and a $K_M K_N \times K_M K_N$ matrix, respectively. The gradient of f_1 is

$$\nabla f_1 = \mathrm{vec}\left(Q'_{kl} + Q'^{*}_{kl}\right), \quad (10)$$

and the i-th row of the Hessian of f_1 is given by [5]

$$\left(\nabla^2 f_1\right)_i = \mathrm{vec}\left(Q''_{k+k',l+l'} + Q''^{*}_{k+k',l+l'}\right), \quad (11)$$

where

$$\begin{aligned} Q'_{kl} &= \mathcal{F}_{M_F,N_F}\left\{\hat{H}^{-1}_{mn}\right\}_{kl} \\ Q''_{k+k',l+l'} &= -\mathcal{F}_{M_F,N_F}\left\{\hat{H}^{-2}_{mn}\right\}_{k+k',l+l'}, \end{aligned} \quad (12)$$

and $k' = (i-1) \mod K_M - M$ and $l' = \lfloor\frac{i-1}{K_M}\rfloor - N$. The gradient of f_2 is given by

$$\nabla f_2 = \mathrm{vec}\left((\Phi' * \mathcal{J}X)_{kl}\right) \quad (13)$$

and the i-th row of the Hessian of f_2 is given by [5]

$$\left(\nabla^2 f_2\right)_i = \mathrm{vec}\left(\left(A^{k'l'} * \mathcal{J}X\right)_{kl}\right), \quad (14)$$

where $\Phi'_{mn} = \varphi'(Y_{mn})$, $A^{k'l'}_{mn} = \varphi''(Y_{mn}) \cdot X_{m-k',n-l'}$, $(\mathcal{J}X)_{mn} = X_{M_X-m,N_X-n}$, $k' = (i-1) \mod (K_M) - M$, and $l' = \lfloor\frac{i-1}{K_M}\rfloor - N$. Computational complexity of f_1, ∇f_1 and $\nabla^2 f_1$ is $\mathcal{O}(M_F N_F \log M_F N_F)$; complexity of f_2 and ∇f_2 is $\mathcal{O}(M_X N_X \log M_X N_X)$, and complexity of $\nabla^2 f_2$ is $\mathcal{O}(M N M_X N_X \log M_X N_X)$.

2.4. Minimization of $L(H;X)$

For minimization of $L(H;X)$, we use the Newton method, which often provides very fast (quadratic) rate of convergence. In the standard Newton approach [11], the direction d at each iteration is given by solution of the linear system

$$\nabla^2 L \cdot d = -\nabla L. \quad (15)$$

Since the objective function is non-convex, in order to guarantee descent direction, positive definiteness of the Hessian is forced by using modified Cholesky factorization, which requires about $\frac{1}{6}K_M^3 K_N^3 + K_M^2 K_N^2$ operations [11]. We find a new iterate by performing backtracking linesearch in the direction d, which guarantees monotonic decrease of the objective function at every iteration. We restrict the search to a subspace of all kernels H that possess a stable inverse [4].

It is also possible to use the fast relative Newton method, based on sparse approximation of the Hessian, which results in a very efficient algorithm with fast convergence and computational complexity per iteration compared to that of gradient methods [3, 5].

3. OPTIMAL SPARSE REPRESENTATIONS

The sparsity prior used in the quasi-ML function (9) is valid for sparse sources and not valid for natural images in their native space. On the other hand, it is especially convenient for the underlying optimization problem due to its convexity; moreover, deconvolution of sparse sources is especially accurate. While it is difficult to model actual distributions of natural images, it is much easier to transform an image in such a way that it fits the sparsity prior. This idea was previously successfully exploited in blind source separation [9, 10, 12].

Let us assume that there exist a *sparsifying transformation* $\mathcal{T}_S$, which makes the source S sparse, such that our algorithm is likely to produce a good source estimate of the restoration kernel H. The problem is that in the blind deconvolution setting, S is not available, and we can apply $\mathcal{T}_S$ to the observation X only. Hence, it is necessary that the sparsifying transformation commute with the convolution operation, i.e.

$$(\mathcal{T}_S S) * W = \mathcal{T}_S(S * W) = \mathcal{T}_S X, \tag{16}$$

such that applying $\mathcal{T}_S$ to X is equivalent to applying it to S. It is obvious that $\mathcal{T}_S$ must be a shift-invariant (SI) transformation. We will use X', S' to denote $\mathcal{T}_S X$ and $\mathcal{T}_S S$, respectively; the subindex "S" in $\mathcal{T}_S$ will be omitted for brevity. For simplicity, we limit our attention to linear shift-invariant (LSI) transformations, i.e. $\mathcal{T}$ that can be represented by convolution with a *sparsifying kernel* $\mathcal{T}S = T * S$.

Thus, we obtain a general blind deconvolution algorithm, which is not limited to sparse sources. We first sparsify the observation data X by convolving it with T, and then apply the sparse blind deconvolution algorithm on X'. The obtained restoration kernel H is then applied to Y to produce the source estimate.

3.1. Optimal sparsifying kernels

Assume that the source S is given. It is desired that the unity restoration kernel δ_{mn} (up to a scaling factor) be a local minimizer of the quasi-maximum likelihood given the transformed source $S * T$ as an observation, i.e.:

$$\nabla L(\delta_{mn}; S * T) = 0. \tag{17}$$

Informally, this means that $S * T$ optimally fits the sparsity prior (at least in local sense). Due to equivariance [6, 4], (17) is equivalent to $\nabla L(T; S) = 0$. In other words, we can define the following optimization problem:

$$\min_T L(T; S) \tag{18}$$

whose solution is the "most sparsifying kernel" for S. This problem is equivalent to the deconvolution problem itself, with the exception of the stability condition, which is not needed here since T is not necessarily invertible.

Unfortunately, since the source image S itself is not available, computation of the sparsifying kernel T_S is possible only theoretically. However, for images belonging to the same class, the sparsifying kernels are likely to be sufficiently similar. Let $\mathcal{C}_1$ denote a class of images, and assume that the unknown source S belongs to $\mathcal{C}_1$. We can find find

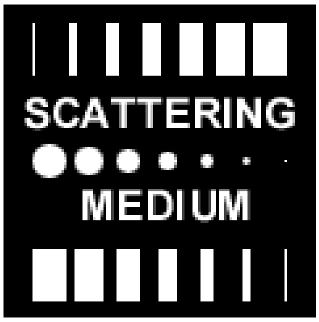


Fig. 1. Source image.

a *training set* of images $S^{(1)}, S^{(2)}, ..., S^{(N_T)} \in \mathcal{C}_1$ and use them to find the optimal sparsifying kernel of S. Optimization problem (18) becomes in this case

$$\min_T \left\{ \frac{-f_1(T)}{2M_F N_F} + \frac{1}{M_X N_X N_T} \sum_{n=1}^{N_T} f_2(S^{(i)} * T) \right\}, \tag{19}$$

i.e. T is required to be the optimum sparsifying kernel for all $S^{(1)}, S^{(2)}, ..., S^{(N_T)}$ simultaneously. Given that the images in the training set are "sufficiently similar" to S, the optimum sparsifying kernel obtained from (19) will be similar enough to T_S.

4. SIMULATION RESULTS

The point spread function of the scattering medium was obtained in a Monte-Carlo simulation of radiative transfer, performed according to the optical model of biological tissues presented in [13]. An normally incident laser beam was used to illuminate a 20 mm$\times$20 mm$\times$10 mm scattering medium, in which the mean free path was set to 0.5 mm. 10^7 photons were generated, of which 4×10^6 were collected by a detector with 31×31 bins, and the rest was absorbed by the medium. The obtained PSF had a Lorenzian shape (Figure 4a), characteristic to PSFs of scattering media [13, 1]. The PSF obtained from simulation was convolved with a 100×100 phantom image (Figure 1) in zero-noise conditions. Figure 2 depicts the observed image.

Blind deconvolution was performed with a 3×3 FIR kernel (Figure 4b). A 2×2 corner detector was used as the sparsifying kernel (Figure 4d). Fast relative Newton method was used to minimize $L(H; X)$, in which the smoothing parameter was set to $\lambda = 10^{-2}$. Optimization was terminated when $\|\nabla L\|$ fell below 10^{-10}. Generally, convergence was obtained in 10-20 iterations, requiring about 0.1 sec per iteration on a PC workstation. Restoration results are depicted in Figure 3.

Restoration quality of SIR = 20.57 dB and SIR$_\infty$ = 35.06 dB was achieved. SIR refers to the interference energy, whereas SIR$_\infty$ to the maximum interference.

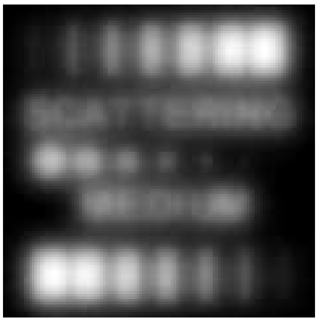


Fig. 2. Observed image.

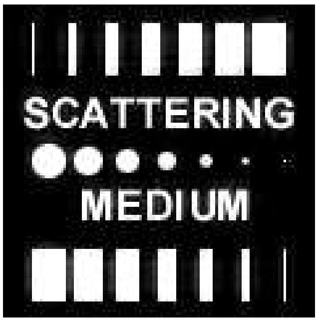
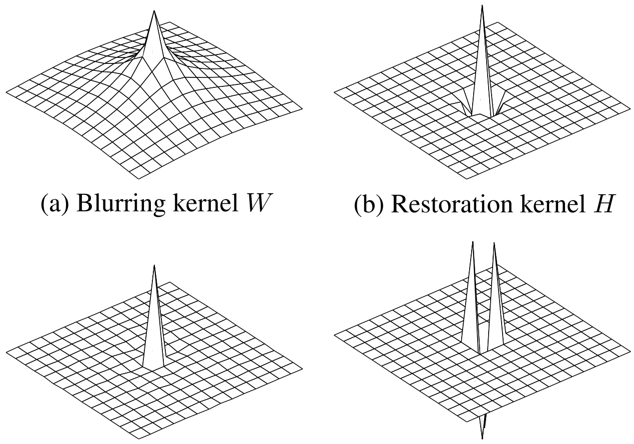


Fig. 3. Restored image.

5. CONCLUSIONS

We have presented a quasi-ML blind deconvolution algorithm for restoration of images obtained through a scattering medium. The source sparsity prior was assumed. We have also shown that the method is applicable for a wider class of images, which can be represented as sparse ones by a shift-invariant transformation, and presented a way of finding such transformations by training. Good performance was achieved on simulated data in moderate noise conditions. Possible applications are microscopy, optical tomography, *in vivo* optical imaging, etc.



(a) Blurring kernel W (b) Restoration kernel H

(c) Global system response G (d) Sparsifying kernel T

Fig. 4. (a) Simulated scattering medium PSF; (b) Restoration kernel estimated in the zero-noise case; (c) Global system response in the zero-noise case. (d) Sparsifying kernel.

6. REFERENCES

[1] S. M. Jefferies, K. J. Schulze, C. L. Matson, K. Stoltenberg, and E. K. Hege, "Blind deconvolution in optical diffusion tomography," *Optics Express*, vol. 10, no. 1, pp. 46–53, 2002.

[2] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Sig. Proc. Magazine*, pp. 43–64, May 1996.

[3] M. Zibulevsky, "Sparse source separation with relative Newton method," in *Proc. ICA2003*, April 2003, pp. 897–902.

[4] A. M. Bronstein, M.M. Bronstein, and M. Zibulevsky, "Blind deconvolution with relative Newton method," Tech. Rep. 444, Technion, Israel, October 2003.

[5] A. M. Bronstein, M.M. Bronstein, Y. Y. Zeevi, and M. Zibulevsky, "Quasi-maximum likelihood blind deconvolution of images using sparse representations," Tech. Rep., Technion, Israel, 2003.

[6] S.-I. Amari, A. Cichocki, and H. H. Yang, "Novel online adaptive learning algorithms for blind deconvolution using the natural gradient approach," in *Proc. SYSID*, July 1997, pp. 1057–1062.

[7] D. Pham and P. Garrat, "Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Sig. Proc.*, vol. 45, pp. 1712–1725, 1997.

[8] P. Kisilev, M. Zibulevsky, and Y.Y. Zeevi, "Multiscale framework for blind source separation," *JMLR*, 2003 (in press).

[9] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition," *Neural Computation*, vol. 13, no. 4, 2001.

[10] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi, and B. A. Pearlmutter, "Blind source separation via multinode sparse representation," in *Proc. NIPS*, 2002.

[11] D. P. Bertsekas, *Nonlinear Programming (2nd edition)*, Athena Scientific, 1999.

[12] A. M. Bronstein, M.M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, "Separation of reflections via sparse ICA," in *Proc. IEEE ICIP*, 2003.

[13] M. Moscoso, J. B. Keller, and G. Papanicolaou, "Depolarization and blurring of optical images by biological tissue," *J. Opt. Soc. Am. A*, vol. 18, no. 4, pp. 948–960, 2001.