Towards a Better Removal of Subsurface Fluoresence Artifacts in Block-Face Imaging

Andrey Tikhonov*

Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia

Pavel Voronin†

Kurchatov NBIC Centre, National Research Centre "Kurchatov Institute", Moscow, Russia
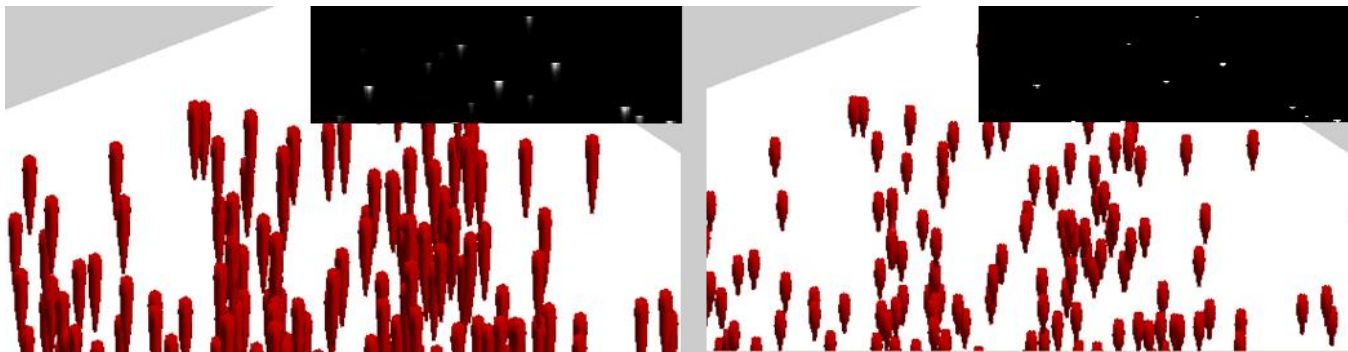
Figure 1: 3D expression map of *c-Fos* gene in an adult mouse brain, before and after subsurface fluorescence removal (fragment, iso-surface rendering); inset: 2D sections.

Abstract

Acquisition, transfer, registration and processing of 3D images of biological objects are all complex procedures that tend to introduce various artifacts, reducing the resolution of the imaging techniques. One such technique is block-face imaging, which suffers greatly from subsurface fluorescence. Removal of this type of artifacts can be reformulated as a deconvolution problem. Although several classic and specialized algorithms have been succesfully applied to this problem, they only work for relatively low amount of blur, sparsely distributed objects and predetermined convolution kernel. Following recent advances in deblurring of 2D natural images, we propose a new Expectation-Minimization algorithm approximating the maximum a posteriori probability solution. The algorithm is not limited to the case of fixed kernel, but can also work with partially known (parametric) or totally unknown kernels. We test the algorithm on model and real data.

Keywords: image processing, 3D deblurring, blind deconvolution, block-face imaging, EM-algorithm

1 INTRODUCTION

The last few years saw rapid development of 3D cryogenic block-face imaging methods. New methods of scanning combined with modern capabilities of visualization techniques allowed researchers to measure and analyze cell-level activity on the scale of the whole brain or even the whole organism [Roy et al. 2010]. These methods can be used in many applications including stem cell therapy, metastatic cancer treatment, gene expression mapping and phenotyping of laboratory animals [Krishnamurthi et al. 2010].

---

*e-mail:AndreyAlexTikhonov@yandex.ru
†pavel.voronin@gmail.com

While it became easier to obtain high-resolution 3D volumetric data, certain imaging artifacts can significantly reduce the resolving power of the technique. For block-face imaging of fluorescent proteins, the biggest problem is subsurface fluorescence and scattering of light. Effectively a type of blur, it is generally modeled as convolution with a point spread function (PSF):

$$\mathbf{I} = \mathbf{L} \otimes \mathbf{f} + \mathbf{n}. \tag{1}$$

Here we used the following notation:

- $\mathbf{I}$ — observed blurred image;

- $\mathbf{L}$ — clear latent image we want to find;

- $\mathbf{f}$ — kernel (PSF), could be known (non-blind deconvolution) or not (blind deconvolution);

- $\mathbf{n}$ — additive noise, some of its properties are usually known.

Point spread function describes the response of an imaging system to a point source or point object. PSF is usually assumed to be shift-invariant (every voxel is convolved with the same kernel). Such assumption is widely used in practice because it leads to more stable and efficient algorithms. We will follow this assumption, but for many practical cases it is not sufficiently adequate and should be utilized with care.

Note that equation 1 is a Fredholm integral equation of the first kind. It is an ill-posed problem, and some prior knowledge is needed to regularize it.

2 METHODS OVERVIEW

3D deconvolution necessarily involves processing huge amounts of data, which places strong constraints on the

methods' computational demands. This has led to a wide usage of simple (and thus computationally cheap) methods.

So far, the research has been focused on non-blind deconvolution. Same as in 2D, the most commonly used methods are Richardson-Lucy algorithm and Wiener filter [Krishnamurthi et al. 2010]. Additionaly, a new cryo-imaging oriented Next-image algorithm was proposed [Steyer et al. 2009].

Wiener filtering is a well-known approach to deconvolution of noisy images coming from signal processing theory [Gonzalez and Woods 2006]. It works in the frequency domain of equation 1, attempting to minimize the impact of deconvoluted noise at frequencies which have a poor signal-to-noise ratio. It is a non-iterative algorithm, implementing an explicit formula.

The Richardson-Lucy algorithm (also known as Lucy-Richardson deconvolution) is an iterative procedure recovering original image under the assumption that the noise is Poisson-distributed [Richardson 1972], [Lucy 1974]. It has been shown that if this algorithm converges, it converges to the maximum likelihood solution.

Next-image algorithm avoids solving deconvolution problem altogether, turning instead to modeling light propagation in tissue directly [Steyer et al. 2009]. Using a crude approximation and a set of physically measured or semi-automatically estimated parameters, the derived scheme is very simple. It essentially boils down to going through the data, blurring each layer to estimate its contribution to subsurface fluorescence, and subtracting it from the consequtive layers.

While data volume for 3D problems seems to limit the complexity of applicable algorithms, we argue that it can still be handled by more general approaches. For most applications, one can use much more rigid kernel priors than in 2D; furthemore, PSF is often known up to a low-dimensional set of parameters. This can be utilized to make the problem tractable by the more complex mordern MAP-algorithms.

3 PROPOSED ALGORITHM

3D deconvolution is a challenging problem and thus recieved modest development. Meanwhile, methods for 2D image deblurring were developing rapidly during the last decade. Exponential growth of the available computational power seems to have led to the current capabilities of computers being high enough so that we can apply more efficient deblurring techniques from this closely related field of research.

We will combine and adapt several modern approaches to 2D deconvolution to construct a 3D blind deconvolution algorithm.

3.1 Blind deconvolution method

Many modern 2D blind deconvolution algorithms are based on approximating maximum aposteriori probability (MAP) solution:

$$(\mathbf{L}, \mathbf{f}) = \arg\max \log p(\mathbf{L}, \mathbf{f}|\mathbf{I}). \tag{2}$$

While applying MAP framework, one must construct probabilistic model for the process of image convolution:

$$p(\mathbf{L}, \mathbf{f}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{L}, \mathbf{f})p(\mathbf{L})p(\mathbf{f}). \tag{3}$$

Each factor in the right part of the model should be chosen with great care, and many variations have been proposed. Very little is known about about appropriate priors for 3D images, so we based our model on some of the more robust 2D priors.

Once the model is fixed, the problem of blind deconvolution is solved by an iterative EM-like procedure. Before the first iteration, PSF is initialized with some realistic values. Then, we alternate between these two steps till convergence:

- Fix kernel $\mathbf{f}$ and solve non-blind deconvolution problem:

$$(\mathbf{L^{new}}|\mathbf{f}) = \arg\max \log p(\mathbf{L}|\mathbf{f}, \mathbf{I}). \tag{4}$$

- Fix latent image estimate $\mathbf{L}$ and update kernel $\mathbf{f}$:

$$(\mathbf{f^{new}}|\mathbf{L}) = \arg\max \log p(\mathbf{f}|\mathbf{L}, \mathbf{I}). \tag{5}$$

3.2 Estimation of the latent image

To calculate optimal clear image $\mathbf{L^{new}}$ according to suggested model, we must choose each factor in the probability model. Basically, we use the approach of [Shan et al. 2008] with the exception of the kernel prior. To simplify the optimization process, following [Cho and Lee 2009], we opted for L_2 regularization instead of L_1.

By taking the negative logarithm of the aposteriori probability $p(\mathbf{L}, \mathbf{f}|\mathbf{I})$, we restate the probability maximization problem as an energy minimization problem, defining energy $E(\mathbf{L}, \mathbf{f}) = -\log p(\mathbf{L}, \mathbf{f}|\mathbf{I})$. Thus, the appropriate target function can be formulated as follows:

$$E(\mathbf{L}, \mathbf{f}) \propto \left(\sum_{\partial^* \in \Theta} w_{\kappa(\partial^*)} \|\partial^* \mathbf{L} \otimes \mathbf{f} - \partial^* \mathbf{I}\|_2^2 \right) + \\ + \sum_{\nu \in \{x,y,z\}} \lambda_1 \|\Phi(\partial_\nu \mathbf{L})\|_1 + \\ + \lambda_2 \left(\|\partial_\nu \mathbf{L} - \partial_\nu \mathbf{I}\|_2^2 \circ \mathbf{M} \right) + \|\mathbf{f}\|_2. \tag{6}$$

Here $M = \{m_i\}$ is an estimate of the local smoothness map with $\circ$ denoting masking (element-wise multiplication); Θ is the set of partial derivative operators: $\Theta = \{\partial_0, \partial_x, \partial_y, \partial_z, \partial_{xx}, \partial_{xy}, \partial_{xz}, \partial_{yy}, \partial_{yz}, \partial_{zz}, \}$, where ∂_0 returns pixel intensities of an image: $\partial_0 \mathbf{I} = \mathbf{I}$. Θ is used to regularize the mismatch between the observed blurred image ($\mathbf{I}$) and estimated latent image convolved with PSF ($\mathbf{L} \otimes \mathbf{f}$). Using second order derivatives was proposed by [Shan et al. 2008] as a means to improve the robustness of the algorithm. Φ is a function specifically designed to penalize unrealistic distribution of gradients in the estimated image.

Even after significant simplification, target function of equation 6 is still highly non-convex and has thousands to millions of variables. In order to optimize it efficiently, authors of [Shan et al. 2008] propose a variable substitution scheme as well as an iterative parameter re-weighting technique. The basic idea is to separate the complex convolutions from other terms so that they can be computed using Fourier transform. Transferring the approach to 3D, we get the following two-step scheme for the inner iteration process of computing the new estimate $\mathbf{L}^{new}$:

- Fix $\mathbf{L}^{new}$ (initialize it with $\mathbf{I}$ before the first iteration) and find the pseudogradients $\Psi = \{\psi_i\}$ of $\mathbf{L}^{new}$ that minimize energy:

$$E'_{\psi_{i,\nu}} = \lambda_1 |\Phi(\psi_{i,\nu})| + \lambda_2 m_i(\psi_{i,\nu}) - \\ - \partial_\nu \mathbf{I}_i^2 + \gamma(\psi_{i,\nu} - \partial_\nu \mathbf{L}_i^2). \tag{7}$$

This energy can be optimized very fast, independently for each pixel i and for each partial derivative $\partial\nu \in \partial\{x, y, z\}$.

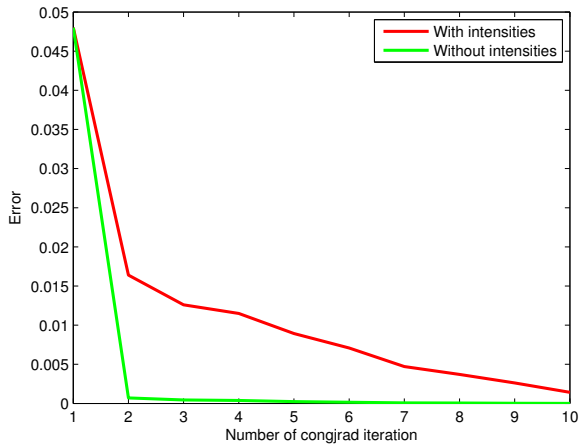- Fix the gradient estimation Ψ and update the latent image $\mathbf{L}^{new}$:

Figure 2: Excluding intensities from the energy improves the optimization convergence.

$$\mathbf{L^{new}} =$$
$$= \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\mathbf{f})}\circ\mathcal{F}(\mathbf{I})\circ\Delta+\gamma\displaystyle\sum_{\nu\in\{x,y,z\}}\overline{\mathcal{F}(\partial_\nu)}\circ\mathcal{F}(\Psi_\nu)}{\overline{\mathcal{F}(\mathbf{f})}\circ\mathcal{F}(\mathbf{f})\circ\Delta+\gamma\displaystyle\sum_{\nu\in\{x,y,z\}}\overline{\mathcal{F}(\partial_\nu)}\circ\mathcal{F}(\partial_\nu)}\right). \qquad (8)$$

Here $\mathcal{F}, \mathcal{F}^{-1}$ denote direct and inverse Fourier transform, respectively. $\Delta = \sum_{\partial^*\in\Theta} w_{\kappa(\partial^*)}\overline{\mathcal{F}(\partial^*)}\circ\mathcal{F}(\partial^*)$, and $\overline{(\cdot)}$ is the complex conjugate operator. Division is performed element-wise.

3.3 Estimation of the kernel

We consider two types of convolution kernels — PSF with a sparsity prior only (commonly used in 2D deblurring), and a parametric set of kernels — cheaper and more robust approach.

3.3.1 Unconstrained PSF

Method used in [Shan et al. 2008] for kernel estimation (based on the interior point method) is extremely slow, and even for 2D images it suffers from memory overhead. Consequently, we decided to use the approach of [Cho and Lee 2009], that is much more efficient.

When $\mathbf{L}$ is fixed, energy in eq. 6 is simplified:

$$E(\mathbf{f}) \propto \left(\sum_{\partial^*\in\Theta} w_{\kappa(\partial^*)}\|\partial^*\mathbf{L}\otimes\mathbf{f}-\partial^*\mathbf{I}\|_2^2\right)+\|\mathbf{f}\|_2; \qquad (9)$$

Optimization of this energy is a quadratic programming problem:

$$\left(\sum_{\partial^*\in\Theta} w_{\kappa(\partial^*)}\|\partial^*\mathbf{L}\otimes\mathbf{f}-\partial^*\mathbf{I}\|_2\right)+\|\mathbf{f}\|_2 \to \max_{\mathbf{f}}$$
$$f_i \geq 0 \qquad (10)$$
$$\sum_i f_i = 1$$

If we relax the problem, ignoring constrains on $\mathbf{f}$, we can find solution analytically, using Fourier transforms. Doing so would require inverting the corresponding matrix, which in this case is a computationally expensive and unrobust operation. Instead, following [Cho and Lee 2009], we employ

conjugate gradients (CG) method to optimize 9. Recent analysis in [Levin et al. 2009; Levin et al. 2011] also indicates the energy in eq. 9 should be further simplified by excluding pixel intensities, $\Theta \to \Theta' = \Theta \setminus \partial_0$. It leads to a significant improvement of conjugate gradients convergence speed and, at the same time, produces a more robust solution (for what is a typical example for our experiments, see fig. 2).

As convolution for two signals ($p\times p$ and $q\times q$) is equivalent to multiplication of two matrices ($p^2 \times q^2$ and $q^2 \times 1$), the problem can be rewritten in matrix form. Denoting matrix corresponding to the sum in 10 by A, we get the following functional to be minimized at each iteration of CG:

$$E_k(\mathbf{f}) = \|Af-b\|_2 + \beta\|f\|_2 =$$
$$= (Af-b)^T(Af-b) + \beta f^T f; \qquad (11)$$

To use the CG method, we also need to compute the gradient of $E_k(\mathbf{f})$:

$$\frac{dE_k(\mathbf{f})}{d\mathbf{f}} = 2A^T A\mathbf{f} + 2\beta\mathbf{f} - 2A^T b; \qquad (12)$$

In general, our optimization procedure produces an unnormalized solution with negative values. So, after the optimization process is finished, we set the small values in the kernel to zero (to be precise, those smaller than $\frac{1}{20}$ of the biggest one); the remaining elements are normalized so that their sum is equal to one.

As CG can not guarantee convergence to the global optimum, the optimization is done in a pyramidal coarse-to-fine manner [Cho and Lee 2009].

3.3.2 Parametric PSF

Now, let's assume the PSF can be parameterized by a small set of variables, $\mathbf{f} = \mathbf{f}(\theta)$. While the energy can still be computed using eq. 11, the equation for the gradient changes somewhat:

$$\frac{dE_k(\mathbf{f})}{d\theta} = 2A^T A\mathbf{f}\frac{d\mathbf{f}}{d\theta} + 2\beta\mathbf{f}\frac{d\mathbf{f}}{d\theta} - 2A^T b\frac{d\mathbf{f}}{d\theta}; \qquad (13)$$

Effectively, this constitutes a more robust procedure, since the dimensionality of the problem is tipically much lower and, even more importantly, the kernel is guaranteed to be normalized and have no negative elements.

3.4 Implementation details

We have implemented the algorithms described in previous section for both 2D and 3D using MATLAB. The implementation was evaluated on model data and real images. The first part of the proposed method is non-blind deconvolution, so we used the standard MATLAB algorithms for evaluation.

Also, we tried two different configurations of coarse-to-fine pyramid: 1) to downsample both the kernel and the images, and 2) to downsample the kernel only (while applying conjugate gradients). The second method is simpler, but the first one often produces a better solution, especially for the coarsest approximations.

Note that working with the downsampled images demands an appropriate scaling of the non-blind algorithm's parameters (otherwise, significant artifacts in early levels of the pyramid will result in a grossly suboptimal solution).

4 EXPERIMENTAL RESULTS

We have tested our algorithm on model and real data.

The model 3D scene consists of numerous tightly packed small spheres. Each sphere is a dense intensity source, with luminocity peaking in the center and falling to about half

of that towards the edges. Positions and densities of the spheres are sampled from a normal distribution. Fig. 3 (upper left) shows an example of our model scene.

In our experiments, we used a simple kernel parametrization, a gaussian multiplied by a hyperbola. This PSF that can be scaled in each dimension (and it fits well with our real data):

$$f_{i,j,k}(h, w, d) = \frac{h - k}{h} * \mathcal{N}([w, d]^T, \frac{w}{5}); \qquad (14)$$

The only parameters are $\{h, w, d\}$ — height, width and depth of the kernel, accordingly. The PSF has a paraboloid-like shape, with most of the density concentrated around the axis. Example of our PSF can be seen on fig. 3 (upper right).

After applying the convolution, many of the spheres stick together and cease to be discrete objects. This replicates the effect subsurface fluorescence has on block-face imaging: individual cells can no longer be easily discerned as such, thus decrising the resolution (resolving power) of the technique. Fig.3 (lower left) shows the model scene after applying convolution with the kernel.

Fig.3 (lower right) shows the output of the proposed non-blind deconvolution method. Even the more closely seeded spheres have been been succesfully separated. The estimation of the densities is also pretty close to the original.

Fig.4 compares the result of non-blind and fully blind (no kernel parameterization information) deconvolution schemes. The source image and the kernel used for blur are the same in both cases. Note that although the reconstructed shapes are very similar, non-blind deconvolution did a much better job at estimating densities.

When parameters of the kernel are taken into account, blind deconvolution algorithm finds a good estimate for the PSF in just a few steps of CG. For the model data, the results are visually indistinguishable from those of non-blind deconvoluton (because of the space limitations, we omit them here).

Fig.1 shows an iso-surface rendering of a fragment of a 3D map of c-Fos expression in an adult mouse brain, before and after subsurface fluorescence removal procedure; inset are 2D section of the same data. Kernel parameterization is the same as above. The proposed method clearly results in a significant improval of the quality of the scan.

References

CHO, S., AND LEE, S. 2009. Fast motion deblurring. *ACM Transactions on Graphics (SIGGRAPH ASIA 2009) 28*, 5, article no. 145.

GONZALEZ, R. C., AND WOODS, R. E. 2006. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

KRISHNAMURTHI, G., WANG, C. Y., STEYER, G., AND WILSON, D. L. 2010. Removal of subsurface fluorescence in cryo-imaging using deconvolution. *Opt. Express 18*, 21 (Oct), 22324–22338.

LEVIN, A., WEISS, Y., DURAND, F., AND FREEMAN, W. T. 2009. Understanding and evaluating blind deconvolution algorithms. *IEEE CVPR* (June).

LEVIN, A., WEISS, Y., DURAND, F., AND FREEMAN, W. T. 2011. Efficient marginal likelihood optimization in blind deconvolution. *IEEE CVPR*.
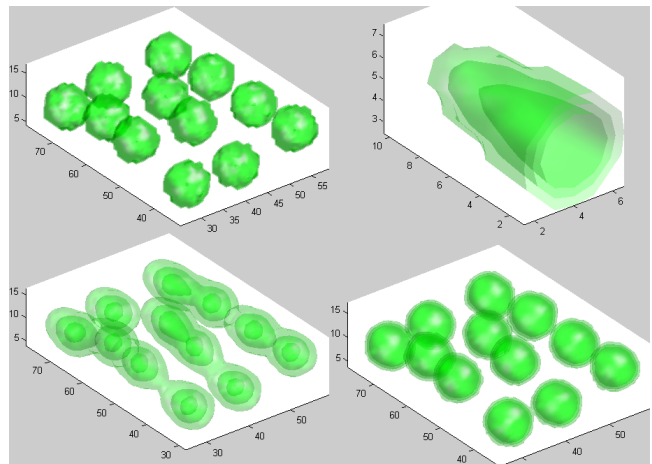
Figure 3: Non-blind deconvolution result. From left to right, from top to bottom: sharp image, convolution kernel, blurred image, restored image
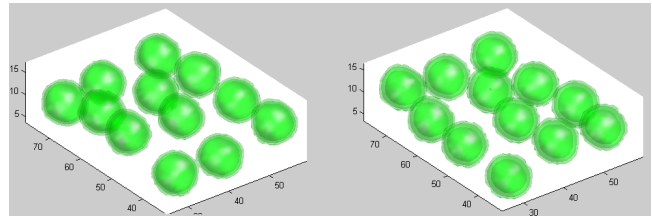

Figure 4: Non-blind (left) versus blind deconvolution.

LUCY, L. B. 1974. An iterative technique for the rectification of observed distributions.

RICHARDSON, H. W. 1972. Bayesian-Based Iterative Method of Image Restoration. *Journal of the Optical Society of America 62*, 1 (Jan.), 55–59.

ROY, D., GARGESHA, M., STEYER, G. J., HAKIMI, P., HANSON, R. W., AND WILSON, D. L. 2010. Multiscale characterization of the pepck-cmus mouse through 3d cryo-imaging. *Journal of Biomedical Imaging 2010* (January), 5:1–5:11.

SHAN, Q., JIA, J., AND AGARWALA, A. 2008. High-quality motion deblurring from a single image. *ACM Transactions on Graphics (SIGGRAPH)*.

STEYER, G. J., ROY, D., SALVADO, O., STONE, M. E., AND WILSON, D. L. 2009. Removal of out-of-plane fluorescence for single cell visualization and quantification in cryo-imaging. *Annals of Biomedical Engineering 37*, 8, 1613–1628.

AUTHORS

- Andrey Tikhonov is a fifth year student of the faculty of Computational Mathematics and Cybernetics in Lomonosov Moscow State University.

- Pavel Voronin is a researcher in Kurchatov NBIC Centre, National Research Centre "Kurchatov Institute", Moscow.