BLIND DECONVOLUTION OF VIDEO SEQUENCES

Ferréol Soulez[1,2], Éric Thiébaut[1], Yves Tourneur[3], Alain Gressard[4], Raphaël Dauphin[4]

[1]Centre de Recherche Astrophysique de Lyon, CNRS-UMR 5574, Université Lyon 1, Ecole Normale Supérieure de Lyon, Observatoire de Lyon, 9 avenue Charles André, Saint-Genis Laval cedex, F-69561, France.

[2] Laboratoire Hubert Curien (ex-LTSI), CNRS-UMR 5516, Université Jean Monnet, 18 rue Pr Benoît Lauras, F-42000 Saint-Etienne, France.

[3]Centre Commun de Quantimetrie, Université Lyon 1,8 avenue Rockefeller - 69373 Lyon cedex 08, France

[4] Service de Cardiologie de l'hôpital de la Croix Rousse, Hospices Civiles de Lyon, France

ABSTRACT

We present a new blind deconvolution method for video sequence. It is derived following an inverse problem approach in a Bayesian framework. This method exploits the temporal continuity of both object and PSF. Combined with edge-preserving spatial regularization, a temporal regularization constrains the blind deconvolution problem, improving its effectiveness and its robustness. We demonstrate these improvements by processing various real video sequences obtained by different imaging techniques.

Index Terms— blind deconvolution, denoising, image reconstruction, video signal processing.

1. INTRODUCTION

The problem of reconstructing a signal using observation blurred by unknown process is called blind deconvolution. Nowadays, blind image deconvolution receives increasing attention from the academic world (see [1] for a review). Although, many authors use video sequences in a super-resolution framework (a multi-frame blurred observation of the same scene (see *e.g.* [2]), to our knowledge, few of them have studied the specific problem of blind video sequence deconvolution [3]. In this paper, we propose to use the temporal correlations in video sequence to design an effective blind deconvolution algorithm.

2. MODEL DESCRIPTION

The video sequence $g(s, t)$ observed in a direction s is composed of T successive images $g_t(s)$ taken at different instant t between its beginning t_1 and its end t_T. It is given by:

$$g(s, t) = \iint h(s, t|s', t')x(s', t')ds'dt' + n(s, t), \quad (1)$$

where $x(s', t')$ is the object brightness at instant t', $h(s, t|s', t')$ is the point spread function (PSF), and $n(s, t)$ account for the noise (source and detector). The PSF $h(s, t|s', t')$ is the observed brightness distribution at instant t in the direction s for a point source located at the instant t' in direction s'.

For data sampled on N pixels, Eq. (1) can be written in a matrix form:

$$g = \mathbf{H} \cdot x + n, \quad (2)$$

where $g = (g_{t_1}^{\mathrm{T}}, \ldots, g_{t_T}^{\mathrm{T}})^{\mathrm{T}}$ and $x = (x_{t_1}^{\mathrm{T}}, \ldots, x_{t_T}^{\mathrm{T}})^{\mathrm{T}}$ are vectors of size $N \cdot T$, n is a $N \cdot T$ random vector and $\mathbf{H}$ is a $N \cdot T \times N \cdot T$ matrix.

In this study, we make two major assumptions. First, there is no temporal spread ($g_t = \mathbf{H}_t \cdot x_t + n_t$, $\forall t$) and $\mathbf{H}$ becomes block diagonal. Then, the PSF is shift invariant (isoplanatic) and $\mathbf{H}_t$ can be expressed using spatial convolution product $*$ and its first row h_t:

$$g_t = h_t * x_t + n_t. \quad (3)$$

Under circulant approximation, this can be evaluated rapidly using FFTs.

3. BAYESIAN APPROACH

In blind deconvolution both the object vector and the PSF h (defined by $h = (h_1^{\mathrm{T}}, \ldots, h_T^{\mathrm{T}})^{\mathrm{T}}$), of size $N \cdot T$, must be estimated using $(N \cdot T)$ measurements in g.

Following [4], the *"maximum a posteriori"* solution is:

$$\{x, h\}^{\mathrm{MAP}} = \arg \max_{\{x, h\}} \Pr\{x, h|g\}. \quad (4)$$

From Bayes' theorem, assuming x and h are independent:

$$\Pr\{x, h|g\} = \frac{\Pr\{g|x, h\} \Pr\{x\} \Pr\{h\}}{\Pr\{g\}} \quad (5)$$

and since $\Pr\{g\}$ does not depend on the model, we can write:

$$\{x, h\}^{MAP} = \arg \max_{\{x, h\}} (\Pr\{g|x, h\} \Pr\{x\} \Pr\{h\}). \quad (6)$$

Maximizing $\Pr\{x, h|g\}$ is equivalent to minimizing

$$\varepsilon(x, h) = -\log \Pr\{x, h|g\} = \Phi_{\mathrm{lkl}}(x, h; g) + \Phi_{\mathrm{obj}}(x) + \Phi_{\mathrm{psf}}(h), \quad (7)$$

The penalizing function to minimize defined in Eq. (7) is the sum of three terms : a *likelihood penalty* $\Phi_{\mathrm{lkl}}(x, h; g)$ ensuring the agreement between the model $x * h$ and the data g, and two *regularization penalty* $\Phi_{\mathrm{obj}}(x)$ and $\Phi_{\mathrm{psf}}(h)$ introducing subjective *a priori* knowledge about the object and the PSF respectively.

3.1. The likelihood penalty term

For Gaussian noise, the likelihood penalty reads:

$$\Phi_{\mathrm{lkl}}(x, h; g) = [g - m(x, h)]^{\mathrm{T}} \cdot \mathbf{C}_{\mathrm{noise}}^{-1} \cdot [g - m(x, h)], \quad (8)$$

where $\mathbf{C}_{\mathrm{noise}}$ is the covariance matrix of the noise and $m(x, h)$ the model. Using Eq. (2) the model is defined as:

$$m(x, h) = \mathbf{H} \cdot x, \quad (9)$$

and is evaluated using Eq. (3) for all t. For uncorelated noise, $\mathbf{C}_{\text{noise}}$ is diagonal and Eq. (8) simplifies to:

$$\Phi_{\text{lkl}}(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{g}) = \sum_t \sum_k w_{k,t} \left((\mathbf{H}_t \cdot \boldsymbol{x}_t)_k - g_{t,k} \right)^2, \qquad (10)$$

where $1/w_{r,t}$ is the noise variance for pixel k of frame t. This model can cope with non-stationary noise and can be used to express confidence on measurements on each pixel of the data. Thus it can deal with unmeasured pixels on the sensor ($1/\sigma^2 = 0$ for such pixels).

3.2. *a priori* on the object

As the different dimensions of the object are not homogeneous, we take a separable regularization term:

$$\Phi_{\text{obj}}(\boldsymbol{x}) = \mu^{\text{obj}} \Psi_{\text{obj}}(\boldsymbol{x}) + \sum_{t=1}^{T} \lambda_t^{\text{obj}} \Theta_{\text{obj}}(\boldsymbol{x}_t), \qquad (11)$$

where $\Psi_{\text{obj}}(\boldsymbol{x})$ is a temporal regularization and $\Theta_{\text{obj}}(\boldsymbol{x}_t)$ is a spatial regularization at each instant t.

3.2.1. Spatial regularization on the object

There are many different kinds of spatial regularization used in image reconstruction. As the noise mostly contaminates high frequencies, smoothness is the most effective regularization constraint to avoid amplification of noise. To avoid oversmoothing of sharp features caused by quadratic regularization we choose an edge preserving regularization. This is achieved by taking, at each instant t:

$$\Theta_{\text{obj}}(\boldsymbol{x}_t) = \sum_k \sum_{v \in V_k} \varphi \left(\frac{x_{t,k} - x_v}{d_v} \right), \qquad (12)$$

where x_v is the value of a pixel v in the neighborhood V_k of pixel k (here a $V8$ neighborhood) and d_v its distance to this pixel. In this work, we choose a l_1–l_2 norm. This norm is asymptotically quadratic (resp. linear) for small (resp. large) pixel differences compared to the threshold η. It is defined by:

$$\varphi(u; \eta) = 2\eta^2 \left[|u|/\eta - \log\left(1 + |u|/\eta\right) \right], \qquad (13)$$

The value of the parameter η is not critical for the reconstruction and can be approximately fixed once for all to the value of one quantization level. In that case, this regularisation is close to a pure ℓ_1 regularization that promote sparsity of gradient[5], but it can be minimized faster by our optimization algorithm.

3.2.2. Temporal regularization on the object

If the scene evolution is slow compared to the time sampling, a smoothing temporal regularization, similarly to the spatial regularization but along the time axis, can be defined.

3.3. *a priori* on the PSF

As for the object, our PSF regularization term is split in:

$$\Phi_{\text{psf}}(\boldsymbol{h}) = \mu^{\text{psf}} \Psi_{\text{psf}}(\boldsymbol{h}) + \sum_{t=1}^{T} \lambda_t^{\text{psf}} \Theta_{\text{psf}}(\boldsymbol{h}_t). \qquad (14)$$

3.3.1. Spatial PSF regularization

As there are many different cause of blur (defocus, motion, diffraction, diffusion...), there are a lot of totally different shapes of PSF. In this context we choose a parametric function $p(\theta)$ (Gaussian, generalized Gaussian, Lorentzian...) to set the prior shape of the PSF, then the regularization becomes [6]:

$$\Theta_{\text{psf}}(\boldsymbol{h}) = (\boldsymbol{h} - \boldsymbol{p}(\theta))^{\text{T}} \mathbf{W} (\boldsymbol{h} - \boldsymbol{p}(\theta)), \qquad (15)$$

where $\mathbf{W}$ is a weight matrix. In our work to constraint the PSF to be increasingly close to the prior shape as it moves away from its center, we consider a diagonal $\mathbf{W}$ with a power law on the diagonal. The functions $p(\theta)$, chosen according to the experimental conditions shall have few parameters (width, orientation).

3.3.2. Temporal PSF regularization

As for the object, if the PSF evolution is slow compared to the time sampling, a quadratic smoothing temporal regularization can be used, *e.g.* :

$$\Psi_{\text{psf}}(\boldsymbol{h}) = \sum_{t=2}^{T-1} (2\,\boldsymbol{h}_t - \boldsymbol{h}_{t-1} - \boldsymbol{h}_{t+1})^2, \qquad (16)$$

4. ALGORITHM SUMMARY

In this Bayesian framework, reconstructing the de-blurred video sequence amounts to determine the couple $\{\boldsymbol{x}^+, \boldsymbol{h}^+\}$ that minimizes the criterion defined in Eq. (7), which writes:

$$\begin{aligned} \varepsilon &= \Phi_{\text{lkl}}(\boldsymbol{x}, \boldsymbol{h}; \boldsymbol{y}) + \mu^{\text{obj}} \Psi_{\text{obj}}(\boldsymbol{x}) + \mu^{\text{psf}} \Psi_{\text{psf}}(\boldsymbol{h}) \\ &+ \sum_{t=1}^{T} \left(\lambda_t^{\text{obj}} \Theta_{\text{obj}}(\boldsymbol{x}_t) + \lambda_t^{\text{psf}} \Theta_{\text{psf}}(\boldsymbol{h}_t) \right), \end{aligned} \qquad (17)$$

The optimal reconstructed image $\boldsymbol{x}^+$ and PSF $\boldsymbol{h}^+$ depend on the value of each parameter (four hyper-parameters per frame).

4.1. Hyper-parameters Setting

To simplify the determination of hyper-parameters, we have made several simplification hypothesis. We suppose the noise statistical property, the image dynamic and the PSF shape identical in each frame. As a consequence: (i) the object spatial hyper-parameters in each frame are identical ; (ii) the PSF spatial hyper-parameters in each frame are identical. Thus we only have now four hyper-parameters (λ_{obj}, λ_{psf}, μ_{obj} and μ_{psf}).

Despite this simplification, choosing the optimal values of the hyper-parameters is cumbersome and difficult. Whether methods such as generalized cross-validation (GCV)[7] or the L-curve [8] are suitable for this task deserves an extensive study which is out of the scope of this paper. In the present work, we simply choose hyper-parameter values by visual inspection of the resulting image.

4.2. Minimization Method

An alternating minimization[9, 10] scheme is used to minimize the criterion.

1. initialize the PSF with its *prior* shape $\boldsymbol{h}^{(0)} = \boldsymbol{p}(\boldsymbol{\theta})$,
2. estimate the optimal object $\boldsymbol{x}^{(k+1)}$ given the PSF $\boldsymbol{h}^{(k)}$,
3. estimate the optimal PSF $\boldsymbol{h}^{(k+1)}$ given the object $\boldsymbol{x}^{(k+1)}$,

4. repeat steps 2 and 3 until convergence or after a defined number (k) of iterations.

In order to determine the optimal image x^+ and PSF h^+ in our inverse problem approach, one has to minimize a criterion with respect to a very large number of variables (all the pixel values for every frames). To that end, we used the VMLM-B algorithm [11] which is a limited memory variant of the variable metric method with BFGS updates [12]. This algorithm, which can further accounts for bound constraints on the parameters. We make use of these bound constraints to enforce PSF positivity. This algorithm has proven effectiveness for image reconstruction and only requires the computation of the penalty function to be minimized and its gradient. The memory requirement is a few times the size of the problem.

5. EXPERIMENTAL RESULTS

This algorithm was used with different experimental data sets from both medical and biological fields, and with different dynamical imaging techniques.
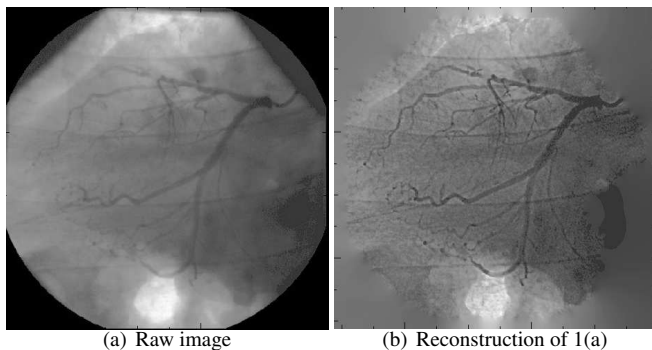
5.1. Coronarography

(a) Raw image   (b) Reconstruction of 1(a)

Fig. 1. One frame of a coronarography sequence

The coronarography is a radiological exam to observe coronary arteries (heart arteries). This gives short (several seconds) video sequences displaying the motion of an impervious to X-ray product in the coronary arteries. These sequences are examined by cardiologist to localized obstacle in the arteries, possibly responsible of an heart attack.

Ten sequences were acquired with the same system, and were then processed with the same assumptions. Several item can be noticed by a simple observation of the raw data (see Fig. 1(a)). (i) The movement of the heart is too important between successive frames to use a temporal constraint on the object. So we set $\mu_{\mathrm{obj}} = 0$. (ii) As the blur seems to be caused by tissue in the axe of projection, and as neither the patient nor the acquisition system was moving during the recording, the PSF is supposed to be constant. As consequence, $\mu_{\mathrm{psf}} = \infty$ or equivalently $h_t = h_1 \,\forall t$. In that case the number of unknown parameters on the PSF is divided by T, increasing the computation speed. (iii) The measured area is not square and is cropped by some parts of the apparatus which remains the same in every frames of every sequences. This is taken in account in the *a priori* weight map $\mathbf{W}$ estimated by a basic thresholding and shown in Fig. 2(b) :

$$w_{k,t} = w_k = \begin{cases} 1 & \text{if } k\text{-th pixel is measured,} \\ 0 & \text{otherwise.} \end{cases} \qquad (18)$$

The prior PSF $p(\theta)$ has a Lorentzian shape with only one parameters: the full width at half maximum θ. Finally only three hyperparameters have to be determined to perform this blind deconvolution: λ_{psf}, λ_{obj} and θ.
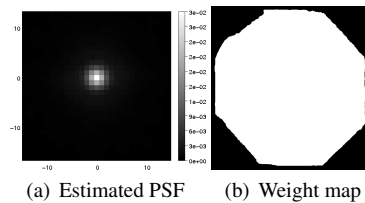

(a) Estimated PSF   (b) Weight map

Fig. 2. Estimated PSF and weight map used for coronarography sequence shown in Fig. 1

These sequences of about 75 frames of 512×512 pixels were processed in about 80 seconds per frames on a Pentium IV CPU at 3.60 GHz. The assumption of an identical PSF in every frames of a same sequence considerably constrains the PSF and improves the convergence of the h^+ determination stage. A further relaxing of this temporal constraint on the PSF, does not show significant improvements and thus the assumption of a constant PSF is verified. A frame of one of the studied sequences is shown Fig. 1(a), along with the coresponding restored frame and the estimated PSF Fig. 2(a). Although the working quantization was very small (at most 10 levels between pixels in the arteries and the background), a visual assessment shows the effectiveness of our technique. Motion perception in the deconvolved video sequences is greatly improved as the separation of the coronaries and the background is enhanced. Let us point out that the cropped area does not disturbed the deconvolution, even for pixels close to the border of this area. These unmeasured parts were just filled with uninformative smooth background in accordance with the smoothing prior. These restored sequences were shown to cardiologists who confirmed the potential utility of the presented technique. If this method proves efficiency with with noisier sequences, they can consider a decrease of the X-ray beam to decrease the irradiation of the patient.
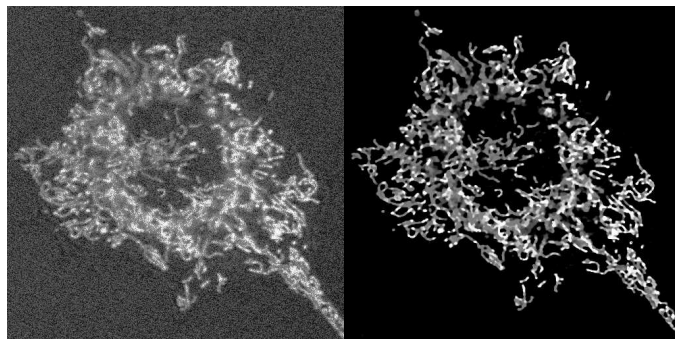
5.2. Confocal Microscopy

Fig. 3. Mitochondria: frames from the raw sequence (right) and the blind deconvolution (left).

Confocal microscopy is an optical technique used for imaging with short depth of field, eliminating out of focus images. In this experiment, a fluorescence product was used to mark mitochondria

in a "non beating HL-1" cell [13]. Temporal evolution of these mitochondria was observed by confocal microscopy. As in the coronarography case and for the same reasons, the assumptions $\mu_{\mathrm{obj}} = 0$ and $\mu_{\mathrm{psf}} = \infty$ are made. Both raw image and its coresponding frame in reconstructed sequence are presented in Fig. 3. The raw sequence is very noisy and the results clearly demonstrates similar ability of regularized blind deconvolution to not only enhance the resolution but also to reduce the noise.

5.3. Conventional Microscopy
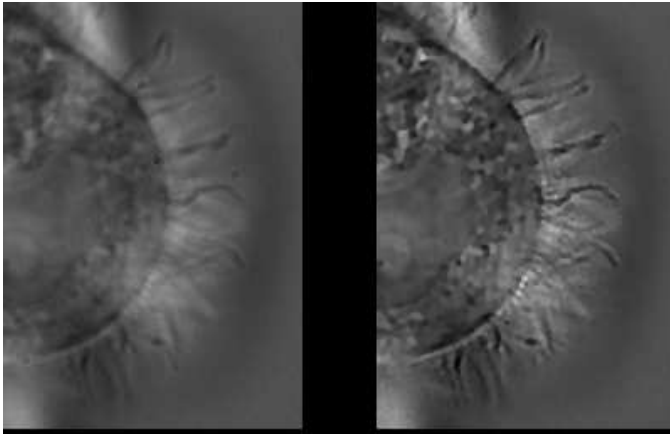
Fig. 4. Epithelial Cell: frames from the raw sequence (right) and the blind deconvolution (left).

A sequence of epithelial hair cells taken with conventional transmission microscope was processed by our method. It is composed of 400 frames of 253×178 square pixels. In this sequences, the depth of field is relatively thick and, as cilia move in the spatial three dimensions, successive focalisations and defocalisations of the same cilium can be observed. The blur on these cilia can be divided in blur due to the apparatus, which is isoplanatic and out of focus blur, which is not isoplanatic. As our method cannot cope with none isoplanatic PSF we tried to remove only the blur due to the apparatus which were considered isoplanatic and identical on every frames ($\mu_{\mathrm{psf}} = \infty$). As only the hairs are moving rapidly, a segmentation based on temporal variation of pixels is made. A temporal constraint μ_{obj} is set for the pixels of the background and the cell, but it is relaxed for pixels near cilia ($\mu_{\mathrm{obj}} = 0$). A frame of both raw and deconvolved sequences are shown in Fig. 4. On several pixels, disturbing diffraction figures due to dust on a glass of the apparatus can be seen. This phenomenon is taken into account by setting the weight these pixels at $w_k^2 = 0$ as in Sec. 5.1. Even in this difficult case, our method achieves to improve resolution of both hairs and inner structures of the cell.

6. CONCLUSION

This paper presents a new method for blind deconvolution of video sequences. We exploit both spatial and temporal continuity to achieve a good characterization of the PSF and then a good video sequence blind deconvolution. This method had proven its capability and robustness with various experimental data.

7. REFERENCES

[1] P. Campizi and K. Egiazarian, Eds., *blind image deconvolution: theory and application*, CRC Press, 2007.

[2] T. J. Schulz, "Multiframe blind deconvolution of astronomical images," *JOSA-A*, vol. 10, no. 5, pp. 1064–1073, may 1993.

[3] D. Li, R.M. Mersereau, D.H. Frakes, and M.J.T. Smith, "New method for suppressing optical turbulence in video," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2005.

[4] E. Thiébaut and J.-M. Conan, "Strict a priori constraints for maximum likelihood blind deconvolution," *JOSA-A*, vol. 12, no. 3, pp. 485–492, March 1995.

[5] David L Donoho and Michael Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l minimization.," *Proc Natl Acad Sci U S A*, vol. 100, no. 5, pp. 2197–2202, Mar 2003.

[6] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial Mathematics, 2005.

[7] Gene H. Golub, Michael Heath, and Grace Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215–223, 1979.

[8] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.

[9] Y.L. You and M. Kaveh, "A regularization approach to joint blur identification and imagerestoration," *Image Processing, IEEE Transactions on*, vol. 5, no. 3, pp. 416–428, 1996.

[10] Tony F. Chan and C.K. Wong, "Convergence of the alternating minimization algorithm for blind deconvolution," *Linear Algebra and its Applications*, vol. 316, pp. 259–285, apr 2000.

[11] E. Thiébaut, "Optimization issues in blind deconvolution algorithms," in *Astronomical Data Analysis II.*, Jean-Luc Starck, Ed., dec 2002, vol. 4847, pp. 174–183.

[12] Jorge Nocedal, "Theory of algorithms for unconstrained optimization," *Acta Numerica*, vol. 1, pp. 199–242, 1992.

[13] S. Pelloux, J. Robillard, R. Ferrera, A. Bilbaut, C. Ojeda, V. Saks, M. Ovize, and Y. Tourneur, "Non-beating hl-1 cells for confocal microscopy: Application to mitochondrial functions during cardiac preconditioning," *Progress in Biophysics and Molecular Biology*, vol. 90, no. 1-3, pp. 270–298, 2006.