Biological video reconstruction using linear or non-linear Fourier measurements

Yoann Le Montagner[a,b], Elsa Angelini[b], and Jean-Christophe Olivo-Marin[a]

[a]Institut Pasteur, Unité d'Analyse d'Images Quantitative, CNRS URA 2582, F-75015 Paris
[b]Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, F-75013 Paris

August 2013

ABSTRACT

In this paper, we investigate how the CS framework can be adapted to biological video microscopy acquisition problems. We first consider a frame-by-frame linear acquisition model in the Fourier domain of the signal, and discuss the relevance of several sparsity models that can be used to drive the reconstruction of the whole video sequence. Then, we switch to a non-linear acquisition model – therefore beyond the "pure" CS framework – in which only the modulus of the Fourier transform of the signal is acquired: by exploiting sparsity properties similar to the one used in the linear acquisition case, we demonstrate the feasibility of a phase retrieval reconstruction procedure applied to video signals.

Keywords: V ideo reconstruction, compressed sensing, phase retrieval, total variation, Fourier measurements.

1. INTRODUCTION

1.1 Video acquisition through Fourier measurements

Compared to 2D images, processing 2D+T video signals leads to particular problems related to the large size of this type of data. However, the counter part of this large size is that natural video signals are in general highly redundant, which allows them to undergone important compression ratio without noticeable degradations. Formally, this property can be exploited to represent the 2D+T video signals in an highly sparse or compressible manner, making these type of signals good candidates for being acquired as advocated by the compressed sensing theory.

In this paper, we assume that this sequence is acquired through a measurement device that works as follows:

1. for each frame of the video, take the 2D Fourier transform of the frame,

2. return a subset of the Fourier coefficients of each frame.

The reason why we focus on this type of measurement operator is that the Fourier transform can be implemented using optical devices, upstream from the actual photo-electric sensors, allowing a simplification of the whole device with a CS-like acquisition strategy.

We focus in particular on two situations, that involve different optical set-ups, as well as different reconstruction algorithms:

- First, the two components of the Fourier coefficients – modulus and phase – are actually measured. In this situation, the measurement vector y can be modeled as a linear transformation of the signal of interest x, up to some noise:

$$y = \Phi x + \text{noise} \tag{1}$$

- Second situation, only the modulus of the Fourier coefficients is measured:

$$y = |\Phi x| + \text{noise} \tag{2}$$

The first situation is simpler in terms of mathematical formulation and reconstruction algorithm, as – up to the definition of a regularization energy promoting sparsity for the class of studied video sequence – it is a direct application of the general theory of compressed sensing (see sec. 2.1). However, measuring the phase of an optical wave involves quite sophisticated set-ups (such as holography set-ups), which limits the usability of this acquisition scheme.

On the other hand, the lack of phase information in the second situation is more challenging in terms of reconstruction procedure. In particular, two video sequences equal up to a spatial translation would be indistinguishable through this acquisition scheme, which means that this ambiguity somehow has to be removed by injecting some prior knowledge during the reconstruction procedure.

1.2 Sparse representations adapted to biological video sequences

In the context of microscopy video, the signals of interest often have similar characteristics, that can be used for reconstruction. In particular, similarly to what was proposed in,[11, 12] we assume that the following properties hold for a video sequence x composed of several frames x_t $(1 \leq t \leq T)$:

1. each frame x_t has a sparse 2D gradient map ∇x_t (intra-frame sparsity),

2. the difference map $x_t - x_{t-1}$ between two consecutive frames is sparse (inter-frame sparsity),

3. the non-zeros coefficients in the spatial gradient maps ∇x_t, which mainly correspond to edges in the underlying frames x_t, are located close to the non-zero coefficients in the corresponding difference maps $x_t - x_{t-1}$.

The first hypothesis accounts for the fact that biological images are well approximated by piecewise constant signals, as they contain few textured areas. The second hypothesis relies on the assumption that the sampling period between two frames is small enough with respect to the time constant that characterizes the displacements in the image medium. Finally, the last hypothesis can be interpreted as a coupling between sparsity in the spatial domain (intra-frame sparsity) and in the temporal domain (inter-frame sparsity): in video sequences of moving and deforming objects and structures, the non-zeros coefficients in the difference maps $x_t - x_{t-1}$ are located ahead of and behind the displacement fronts of the object boundaries, which correspond also – by definition – to areas where the spatial gradient is non-zero. This last hypothesis dramatically reduces the set of *a priori* acceptable sequences, which improves video reconstructions as shown by the results presented below.

2. RECONSTRUCTION FROM LINEAR MEASUREMENTS: CS WITH 3D TOTAL VARIATION

In the case of the linear acquisition scheme (1), the reconstruction problem fits right into the general theory of compressed sensing. We recall here some of the main aspects of this theory, before presenting how it applies to our particular reconstruction problem.

2.1 Compressed sensing background

The inverse problem tackled by CS can be formulated as follows: given a signal of interest $x \in \mathbb{R}^N$ measured through a *random* linear operator Φ that outputs a vector $y \in \mathbb{R}^M$ of observations with $M \ll N$, can x be recovered from y? The randomness of the measurement operator Φ should not be understood in strict meaning, but rather as the fact that Φ should spread the information contained in x over the whole vector y. Examples of such operators include random Gaussian or Bernouilli matrices,[5] randomly subsampled Fourier or Hadamard transforms,[3] or dedicated unitary matrices.[6]

Previous results (see[2, 4, 7]) establish that x can be recovered from y if it has a sparse representation in some known dictionary Ψ, i.e. there exists a sparse vector $\gamma \in \mathbb{R}^D$ such that $x = \Psi\gamma$, and if Φ behaves *like an isometry* for sparse linear combinations of columns of Ψ; this idea is quantified using the notion of *restricted isometry*

property (see[2] for more details). Up to some technical hypothesis, an estimator $\hat{x}$ of x can be defined as a solution of the following convex problem:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \|\Psi^* x\|_1 \text{ s.t. } \|\Phi x - y\|_2 \leq \epsilon \tag{3}$$

where ϵ is a parameter tuned according to the level of noise that corrupts the observations y. Candes *et al.*[2] gives an upper bound on the estimation error $\|\hat{x} - x\|_2$.

In the following, we denote the entities involved in our problem as follows: the 2D+T signal of interest composed of T successive 2D frames $x_t \in \mathbb{R}^N$ ($1 \leq t \leq T$) is $X \in \mathbb{R}^{NT}$. This signal is measured through a linear memoryless operator Φ, resulting in a vector $Y \in \mathbb{R}^M$ of observations. Formally:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} \phi_1 & & & \\ & \phi_2 & & \\ & & \ddots & \\ & & & \phi_T \end{bmatrix}}_{\Phi} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}}_{X} \tag{4}$$

2.2 Sparsity prior based on 3D total variation

To enforce the sparsity properties of microscopy video sequences presented in Sec. 1.2, we propose to use the three dimensional total variation as a regularization term in the reconstruction problem (3). 3D total variation (3D-TV) is defined as:

$$\|X\|_{\text{TV-3D}} = \sum_{t=1}^{T-1} \sum_{P} \sqrt{|(D_h x_t)[P]|^2 + |(D_v x_t)[P]|^2 + |(x_{t+1} - x_t)[P]|^2} \tag{5}$$

where P visits every pixel of the frames, and D_h and D_v stand for the discrete derivative operators respectively in the horizontal and vertical directions.

The reason explaining why minimizing $\|X\|_{\text{TV-3D}}$ enforces the first two sparsity properties mentioned in Sec. 1.2 (i.e. intra-frame and inter-frame sparsity) stems from the following inequalities, that can be easily derived from the definition of the 3D total variation (5):

$$\max\left(\sum_t \|x_t\|_{\text{TV}}, \sum_t \|x_{t+1} - x_t\|_1\right) \leq \|X\|_{\text{TV-3D}} \leq \left(\sum_t \|x_t\|_{\text{TV}}\right) + \left(\sum_t \|x_{t+1} - x_t\|_1\right) \tag{6}$$

where $\|\cdot\|_{\text{TV}}$ is the usual 2D total variation (see for instance[15]). Indeed, minimizing $\|X\|_{\text{TV-3D}}$ leads to small values of both the cumulated 2D TV of all the frames of the sequence $\sum_t \|x_t\|_{\text{TV}}$ and the cumulated l_1-norm of all the frame to frame differences $\sum_t \|x_{t+1} - x_t\|_1$, and reciprocally. Moreover, from the concavity property of the square root function, it can be shown that:

$$\left(\sum_t \|x_t\|_{\text{TV}}\right) + \left(\sum_t \|x_{t+1} - x_t\|_1\right) \leq \sqrt{2} \cdot \|X\|_{\text{TV-3D}} \tag{7}$$

and that this inequality is tight if and only if, for all $t \in [1, T-1]$ and all pixels P, the following holds:

$$\sqrt{|(D_h x_t)[P]|^2 + |(D_v x_t)[P]|^2} = |(x_{t+1} - x_t)[P]| \tag{8}$$

In other words, for given values of $\sum_t \|x_t\|_{\text{TV}}$ and $\sum_t \|x_{t+1} - x_t\|_1$ – which can be thought as measures of, respectively, the intra-frame and inter-frame sparsity – the 3D total variation is minimal when, at each pixel P and time point t, the amplitude of the local spatial gradient $\sqrt{|(D_h x_t)[P]|^2 + |(D_v x_t)[P]|^2}$ is equal to amplitude of the local frame-to-frame difference $|(x_{t+1} - x_t)[P]|$: this explain the relation between the minimization of $\|X\|_{\text{TV-3D}}$ and the third sparsity property enforced on 2D+T video sequences mentioned in Sec. 1.2. This relation can also be explained by interpreting $\|X\|_{\text{TV-3D}}$ as a particular mixed $l_{1,2}$-norm (see for example the work of Bach[1] and references therein) operating on a linear transform of X that would stack its discrete derivatives in the horizontal, vertical and diagonal directions.

3. RECONSTRUCTION FROM NON-LINEAR MEASUREMENTS: PHASE RETRIEVAL WITH TOTAL VARIATION CONSTRAINTS

3.1 Phase retrieval background

The problem of recovering a signal from the modulus of its Fourier transform, known as the *phase retrieval* problem, has been studied for a long time: this reconstruction technique is used for instance for X-ray microscopy applications in crystallography (see[9, 13]). To recover a signal $x \in \mathbb{R}^N$ from a measurement vector y defined as (2), the algorithm proposed in[9] defines two subsets of $\mathbb{R}^N$:

- the *data set* $\mathcal{D}_{y,\epsilon}$, that contains all the signals x that correspond to the measured samples, with a certain tolerance ϵ that depends on the noise that affects these measurements:

$$\mathcal{D}_{y,\epsilon} = \left\{ x \in \mathbb{R}^N \text{ s.t. } \|y - |\Phi x|\|_2 \leq \epsilon \right\} \tag{9}$$

- a *regularization set* $\mathcal{R}$ that corresponds to all the signals that meet certain prior conditions which are known to be true for the actual solution. For crystallography applications, $\mathcal{R}$ typically consists in all the 2D images that are supported on a given subset of pixels.

Then, an estimator $\hat{x}$ of the solution is obtained as the limit of alternated projections over the two sets $\mathcal{D}_{y,\epsilon}$ and $\mathcal{R}$:

$$\hat{x} = \left(\Pi_{\mathcal{R}} \circ \Pi_{\mathcal{D}_{y,\epsilon}} \circ \Pi_{\mathcal{R}} \circ \cdots \circ \Pi_{\mathcal{D}_{y,\epsilon}} \right)(x_0) \tag{10}$$

where $\circ$ is the composition operator, x_0 is an initial guess of the solution, and $\Pi_{\mathcal{D}_{y,\epsilon}}$ and $\Pi_{\mathcal{R}}$ stand respectively for the projection operators over $\mathcal{D}_{y,\epsilon}$ and $\mathcal{R}$:

$$\Pi_{\mathcal{D}_{y,\epsilon}}(x) = \arg\min_{z \in \mathcal{D}_{y,\epsilon}} \|z - x\|_2 \tag{11}$$

and similarly for $\Pi_{\mathcal{R}}$. It was shown in[9] that the sequence of estimators (10) converges toward the intersection of $\mathcal{D}_{y,\epsilon}$ and $\mathcal{R}$.

In our case, the prior information is quite different from the one available in crystallography applications (i.e. support constraints). We therefore introduce in the next sections a general reconstruction approach suited for microscopic video reconstruction which exploits a similar alternated projection iteration scheme as the one used in phase retrieval, but allows different *a priori* information to be used.

3.2 Problem formulation

We consider here a reconstruction scheme that operates iteratively on consecutive frames. Starting from an initial key-frame assumed to be available, we reconstruct the following frame using its partial Fourier modulus data, and propagate the reconstruction process to the next step. This step-by-step procedure differs from the reconstruction scheme presented in Sec. 2, in which all the frames of the video sequence are reconstructed in a joint manner.

With this formulation, it is important to note that, although the set $\mathcal{D}_{y,\epsilon}$ is not convex, (11) can be solved explicitly. In the case of $\epsilon = 0$, $\Pi_{\mathcal{D}_{y,\epsilon}}(x)$ is computed as follows: 1st) take the Fourier transform $\tilde{x}$ of x; 2nd) for all the spatial frequencies k for which a measure $y(k)$ is available, replace the obtained modulus $|\tilde{x}(k)|$ with $y(k)$; 3rd) finally, inverse the Fourier transform. For $\epsilon > 0$, the second step is slightly more complex, but can still be run in $\mathcal{O}(N)$ operations. The overall procedure is then dominated by the first and third steps, whose complexity $\mathcal{O}(N \log N)$ corresponds to the evaluation of the Fourier transform.

3.3 Definition of a regularization set using an hybrid total variation

To enforce the sparsity properties mentioned in Sec. 1.2, we introduce a hybrid total variation energy over the space $\mathbb{R}^N$ of 2D images, defined as:

$$\|x\|_{\mathrm{hTV},a} = \sum_P \sqrt{|(D_h x)\,[P]|^2 + |(D_v x)\,[P]|^2 + |(x-a)\,[P]|^2} \tag{12}$$

where P visits every pixel, and where the discrete derivation operators D_h and D_v are defined as in (5). The hybrid TV of each frame x_t and the 3D total variation (5) of the whole sequence X are related through the following relation:

$$\|X\|_{\mathrm{TV\text{-}3D}} = \sum_t \|x_t\|_{\mathrm{hTV},x_{t+1}} \tag{13}$$

As a consequence, the hybrid total variation inherits the properties of the 3D total variation in terms of sparsity enforcement (see Sec. 2.2), and can be used to reconstruct a video sequence that presents the characteristics stated in Sec. 1.2.

The regularization set $\mathcal{R}$ involved in the alternated projection scheme (10) is defined as a level set of the hybrid TV (12): $\mathcal{R}_{a,\tau} = \left\{ x \in \mathbb{R}^N \text{ s.t. } \|x\|_{\mathrm{hTV},a} \leq \tau \right\}$. $\tau > 0$ is a parameter that can be either set *a priori* or adaptively estimated during the reconstruction process, as proposed in.[12]

3.4 Evaluation of the projection operator $\Pi_{\mathcal{R}_{a,\tau}}$

The alternated projection scheme (10) involves several projections on the set $\mathcal{R}_{a,\tau}$, which implies to solve the following problem for several values of the argument x:

$$\Pi_{\mathcal{R}_{a,\tau}}(x) = \arg\min_z \|z - x\|_2 \text{ s.t. } \|z\|_{\mathrm{hTV},a} \leq \tau \tag{14}$$

This convex optimization problem can be solved by using an algorithm derived from the total variation projection method presented by Fadili and Peyré,[8] where the constrained problem (14) is recast into the following unconstrained one:

$$\arg\min_u \tau \|u\|_\infty + \left\langle u^{(1)}, a \right\rangle + \frac{1}{2} \left\| x - u^{(1)} + \mathrm{Div}\, u^{(2,3)} \right\|_2^2 \tag{15}$$

where u represents a 3D vector field (i.e. a 3-channel image) over the same domain than x, $u^{(1)}$ and $u^{(2,3)}$ denote respectively the first and the concatenation of the second and third components of u, Div is the adjoint operator of $-\nabla$, and $\|u\|_\infty = \max_P \|u\,(P)\|_2$ where P visits every pixel. It can be shown (see[10]) that (14) and (15) are equivalent in that the corresponding optimal variables $z^\star$ and $u^\star$ are related through the equation $z^\star = x - u^{\star(1)} + \mathrm{Div}\, u^{\star(2,3)}$. Problem (15) can be solved with a first-order gradient method, with a Nesterov accelerated scheme (see[14] and algorithm 4.2 in[16]). The projection operator $\Pi_{\mathcal{R}_{a,\tau}}$ is then evaluated iteratively, with each iteration having a $\mathcal{O}\,(N \log N)$ complexity, and an overall quadratic convergence rate thanks to the Nesterov scheme. It is however much slower than the computation of $\Pi_{\mathcal{D}_{y,\epsilon}}$, although a careful initialization of the gradient descent proved to provide a significant speed up of the computation.

4. RESULTS

4.1 Simulation data

We present in the following a comparison of the two reconstructions schemes on an test sequence showing *amoeba* cells, that comprises 80 frames of 256×256 pixels. The reconstructions were performed according to the following conditions:

- *CS*-TV-3D *with linear measurements*: 10% of the Fourier samples (amplitude and phase) selected in a uniform random manner were used. The whole sequence is reconstructed in a joint manner.

- *Phase retrieval with non-linear measurements*: we used 25% of Fourier samples (amplitude only), also selected in a uniform random manner. In this case, the first frame of the sequence is used as a key frame to initialize the iterative reconstruction process.
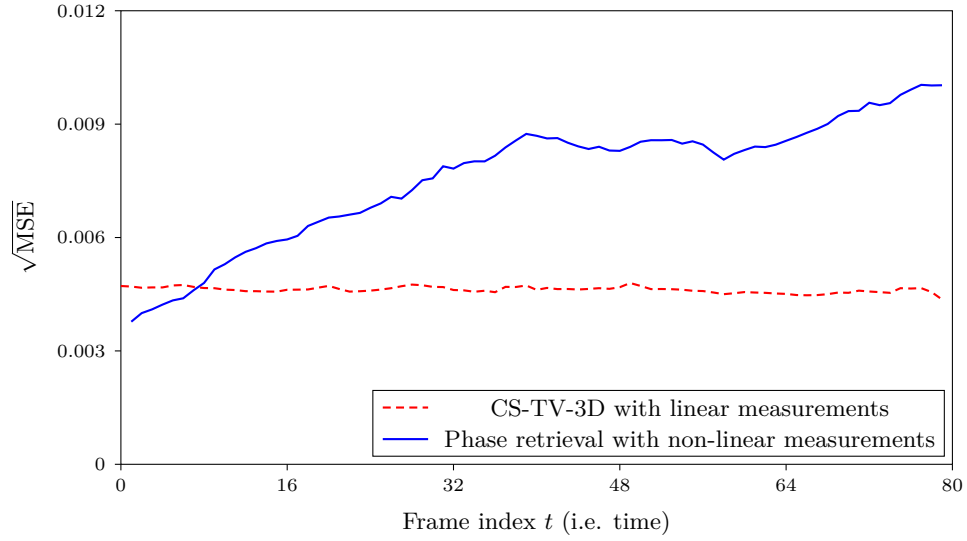
Figure 1. Frame-wise RMSE between the original *Amoeba cells* video sequences and the reconstructions obtained with the two tested methods.
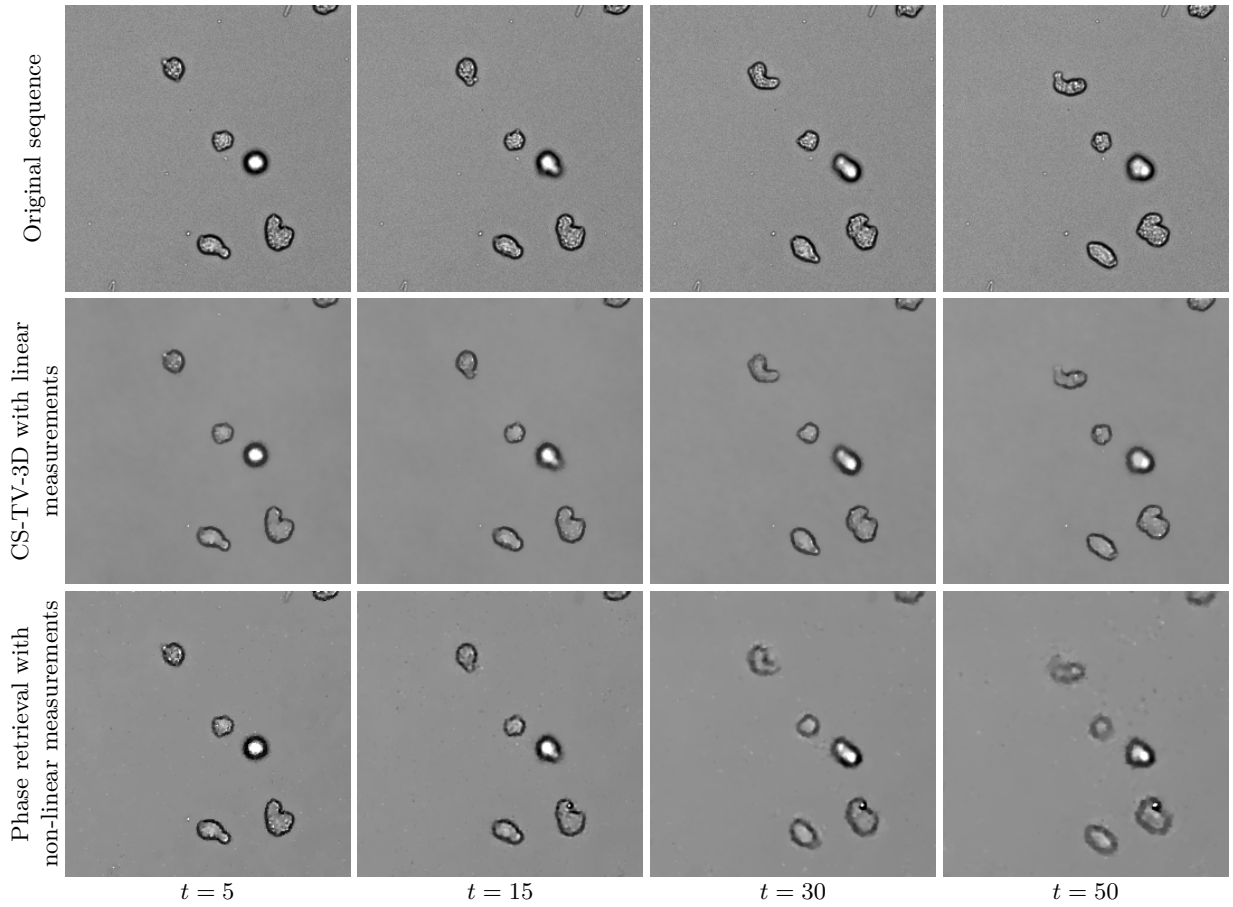

Figure 2. Original test sequence showing *amoeba* cells (top row), and reconstruction results obtained with the two tested methods. The parameter t represents the frame index in the sequence.

4.2 Quantitative and qualitative results

The reconstruction results are presented in Figs. 1 and 2. Figure 1 plots the reconstructed error (root mean squared error) computed frame-by-frame for the two schemes, and Figure 2 shows examples of intermediate reconstructed frames. It can be seen that when reconstructing the sequence from linear measurements using CS-TV-3D, the visual quality of the reconstructed sequence is very good: in particular, we observe sharp edges, flat background, and well-contrasted objects. These characteristics are stable all along the sequence, as confirmed by the almost constant RMSE error ($\approx 5 \times 10^{-3}$) in Fig. 1.

This is in contrast with the results obtained for the reconstruction from non-linear measurements using phase retrieval: in this case, the quality of the reconstructed sequence is quite satisfactory at the beginning of the sequence (i.e. in the vicinity of the key frame), but degrades rapidly. This degradation is clearly visible on the RMSE curve, and appears as an increasing blurring effect. The fact that the degradation increases with time can be explained by error accumulations, as in the iterative phase retrieval scheme, each frame is reconstructed from the previous ones. This effect could be limited by using more frequent key-frames, or by slightly modifying the expression of the hybrid total variation to include an appropriate motion prediction heuristic (see[12]).

5. CONCLUSION

In this paper, we presented two sparsity-based video microscopy reconstruction methods, using random projections in the Fourier domain, either linear (amplitude and phase) or non-linear (amplitude only). In the former case, the reconstruction scheme exploits general CS reconstruction results with a 3D-TV based reconstruction functional, where all the frames of the sequence are reconstructed in a joint manner. In the latter case the reconstruction is inspired by previous variational phase retrieval techniques, and introduces a regularization set exploiting an hybrid TV regularization energy within an iterative scheme. The results demonstrate that video reconstruction can be performed from partial Fourier measurements, but that the quality of the resulting images is much higher if phase information is available.

REFERENCES

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Structured sparsity through convex optimization. *Statistical Science*, 27(4):450–468, Nov. 2012.

[2] E. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2010.

[3] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, June 2007.

[4] E. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, Dec. 2005.

[5] E. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, Dec. 2006.

[6] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran. Fast and Efficient Compressive Sensing Using Structurally Random Matrices. *IEEE Transactions on Signal Processing*, 60(1):139–154, 2012.

[7] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, Apr. 2006.

[8] J. M. Fadili and G. Peyré. Total variation projection with first order schemes. *IEEE Transactions on Image Processing*, 20(3):657–669, Mar. 2011.

[9] J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, Aug. 1982.

[10] Y. Le Montagner. *Algorithmic solutions toward applications of compressed sensing for optical imaging.* PhD thesis, Télécom ParisTech, December 2013.

[11] Y. Le Montagner, E. Angelini, and J.-C. Olivo-Marin. Video reconstruction using compressed sensing measurements and 3D total variation regularization for bio-imaging applications. In *International Conference on Image Processing*, pages 917–920, Orlando, 2012. IEEE.

[12] Y. Le Montagner, E. Angelini, and J.-C. Olivo-Marin. Phase retrieval with sparsity priors and application to microscopy video reconstruction. In *International Symposium on Biomedical Imaging*, pages 600–603, San Francisco, 2013. IEEE.

[13] J. Miao, P. Charalambous, and J. Kirz. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(July):342–344, 1999.

[14] Y. Nesterov. Gradient methods for minimizing composite objective function. Technical report, Université Catholique de Louvain, 2007.

[15] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Journal Physica D*, 60:259–268, 1992.

[16] P. Weiss. *Fast convex optimization algorithms: application to image restoration and change detection.* PhD thesis, Université de Nice - Sophia Antipolis, 2008. document in French.