Complex wavelet regularization for solving inverse problems in remote sensing

Mikael Carlavan, Pierre Weiss, Laure Blanc-Féraud, Josiane Zerubia

COMPLEX WAVELET REGULARIZATION FOR SOLVING INVERSE PROBLEMS IN REMOTE SENSING

Mikaël Carlavan, Pierre Weiss, Laure Blanc-Féraud, Josiane Zerubia

Projet Ariana - CNRS/INRIA/UNSA
pierre.armand.weiss@gmail.com, {mcarlava, blancf, zerubia}@sophia.inria.fr
INRIA Sophia-Antipolis
2004, route des Lucioles, BP 93
06902 Sophia-Antipolis - France

ABSTRACT

Many problems in remote sensing can be modeled as the minimization of the sum of a data term and a prior term. We propose to use a new complex wavelet based prior and an efficient scheme to solve these problems. We show some results on a problem of image reconstruction with noise, irregular sampling and blur. We also show a comparison between two widely used priors in image processing: sparsity and regularity priors.

1. INTRODUCTION

Some problems in remote sensing consist in retrieving an image $u \in \mathbb{R}^n$ acquired by a satellite, from a damaged observation. This can be modeled as follows:

$$g = Au + n \tag{1}$$

where $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear transform (generally the *Point Spread Function* of the optical system of the satellite), $n \in \mathbb{R}^m$ is a noise (Gaussian noise for example) and $g \in R^m$ is the observed image. The formalism (1) covers a large class of problems: image reconstruction (including deconvolution) [1], zooming [2] or denoising [3]. Usually, finding the original image u from the observation g is an ill-posed problem (A is non invertible or ill-conditioned). Variational approaches have been proposed [3, 4, 5] to solve these problems, using different norms on the data term and the regularizing term. The norm on the data term allows to adapt the restoration model to the noise model. For instance the l^2-norm is adapted to Gaussian noise while the l^1-norm is more robust to impulse noise [4, 6]. A variational approach consists in determining:

$$\arg\min_{u \in \mathbb{R}^n} \left\{ \|Au - g\|_p^p + \lambda J(u) \right\} \tag{2}$$

where $J(u)$ is a regularizing term, $\|.\|_p$ denotes the l^p-norm and λ is a regularizing parameter. Efficient priors are of the form $J(u) = \|Bu\|_1$ where B is a linear transform.

We focus here on our previous work [4] and consider again the problem of the reconstruction of an image sampled on a regular grid from an image sampled on an irregular grid, knowing the position of the irregular samples. We use the same model as in [4] and want to solve:

$$\arg\min_{u \in \mathbb{R}^n} \left\{ \|SHFu - g\|_p^p + \lambda J(u) \right\} \tag{3}$$

where F is the discrete Fast Fourier Transform (FFT), H is the Fourier transform of the PSF of the satellite and S is the transform that creates an irregularly sampled image from its regular samples in the Fourier domain. This last operator can be computed efficiently with the Unequally Spaced Fast Fourier Transform (USFFT) from G. Beylkin [7]. In [4] the authors set $J(u) = \|\nabla u\|_1$ which is the total variation [3]. Total variation is a widely used prior in image processing as it removes noise while preserving the discontinuities of the image. However, this regularization does not allow to recover the textures correctly (this effect is known as "*cartoon*" effect). This is a problem in remote sensing as we want to retrieve thin details. Some errors in the sampling grid may generate huge errors on the intensity result (near edges for example), so the authors of [4] set $p = 1$ in order to be robust against impulse noise. The same problem has been solved by *Almansa et al.* with $p = 2$ [1]. Finally, the model in [4] reduces to:

$$\arg\min_{u \in \mathbb{R}^n} \left\{ \|Au - g\|_1 + \lambda \|\nabla u\|_1 \right\} \tag{4}$$

where $A = SHF$. In this paper, we propose an efficient wavelet based prior allowing to retrieve thin details and a fast algorithm to solve the considered problem. We also show a comparison between two common priors in image processing for wavelet regularized problems.

The authors would like to thank the CS Compagny in Toulouse (France) for partial funding of this research work and the Space French Agency (CNES) for providing the data. We particularly thank Anne Chanié and Rosario Ruiloba from CS for several interesting discussions.

2. COMPLEX WAVELET REGULARIZATION

As previously said, the total variation does not allow to recover the textures correctly. In order to restore all thin details, we set B to be a wavelet transform $W : \mathbb{R}^n \to \mathbb{R}^q$. Real non-redundant wavelets are not translation and rotation invariant, and using them in (2) leads to poor results in practice. We propose to use instead the Dual-Tree Complex Wavelet transform (DTCW) [8]. The choice of the DTCW transform is motivated by the fact that this transform is *quasi*-invariant by translation and rotation with a low redundancy (4 for 2D images). This *quasi*-invariance is a necessary property to be used as a regularizing operator. This wavelet transform is built using two real wavelets transform. One of these wavelet transforms give the real part of the transform while the other provides the complex part. When thresholded, these complex coefficients give less artifacts than usual real wavelets. Moreover, real, non-redundant wavelets suffer from a weakness of directionality that is improved with the Dual-Tree Complex Wavelet transform [8]. Finally the problem under consideration writes:

$$\arg\min_{u\in\mathbb{R}^n} \left\{ \|Au-g\|_1 + \lambda\|Wu\|_1 \right\} \tag{5}$$

where A is a convolution with a blurring operator and an irregular sampling operator and W is the DTCW transform. Due to the l^1-norms and the ill-conditioning of A, this problem is very challenging to solve numerically. In the next section, we present an efficient algorithm to solve it.

3. DUAL PROBLEM AND FAST ALGORITHM

The authors of [4] use a smooth approximation and a gradient descent to solve (4). This method only converges in $O\left(\frac{1}{\sqrt{k}}\right)$, where k is the number of iterations. We propose to use a fast multi-step first order method originally proposed by Y. Nesterov [9] to solve this problem with notable improvements compared to other first order techniques. The idea of Y. Nesterov is that we can improve the convergence rate of classical first order methods, if at each iteration the gradient step is function of the gradient of all the previous iterations and not only the gradient at the current iteration. When applied to convex differentiable functions, this gives an algorithm with a convergence rate in $O\left(\frac{1}{k^2}\right)$, while classical first order methods have a worst case convergence rate in $O\left(\frac{1}{k}\right)$.

The problem (5) is not differentiable, so as in [4] we need to smooth it. But instead of smoothing the primal problem, we smooth the dual problem which offers better results in term of computing time [10]. The dual formulation of (5) writes:

$$\arg\min_{u\in\mathbb{R}^n} \left\{ \max_{y\in Y} \left(< Du-F, y >\right) \right\} \tag{6}$$

with:

$$D = \begin{bmatrix} \lambda W \\ A \end{bmatrix}, \quad F = \begin{bmatrix} 0 \\ g \end{bmatrix} \tag{7}$$

$$Y = \{y = (y_1, y_2) \in \mathbb{R}^q \times \mathbb{R}^m, \|y_1\|_\infty \le 1 \text{ and } \|y_2\|_\infty \le 1\} \tag{8}$$

We smooth the dual problem by adding the term $\frac{\epsilon}{2}\|u-u^0\|_2^2$ (u^0 should be chosen close to the set of minimizer of (5)):

$$\arg\min_{u\in\mathbb{R}^n} \left\{ \max_{y\in Y} \left(< Du-F, y > + \frac{\epsilon}{2}\|u-u^0\|_2^2 \right) \right\} \tag{9}$$

We can now transform the min-max problem (9) in a max-min one. The min problem consists in solving:

$$\arg\min_{u\in\mathbb{R}^n} \left\{ < Du-F, y > + \frac{\epsilon}{2}\|u-u^0\|_2^2 \right\} = -\frac{D^* y}{\epsilon} + u^0 \tag{10}$$

where D^* denotes the complex conjugate of D:

$$D^* = \begin{bmatrix} \lambda W^* & A^* \end{bmatrix} \tag{11}$$

Finally, by introducing the solution of (10) into (9), problem (5) rewrites as follows:

$$-\min_{y\in Y} \Bigg(\underbrace{\frac{1}{2\epsilon}\|D^* y\|_2^2 - < Du^0 - F, y >}_{\Psi_\epsilon(y)} \Bigg) \tag{12}$$

$\Psi_\epsilon(y)$ is a convex and differentiable function with a Lipschitz continuous gradient:

$$\|\nabla\Psi_\epsilon(y_1) - \nabla\Psi_\epsilon(y_2)\|_2 \le L\|y_1 - y_2\|_2 \tag{13}$$

where $L = \frac{1}{\epsilon}\left(\lambda^2\|W\|_2^2 + \|A\|_2^2\right)$. We can apply a slightly modified version [10] of the algorithm of Y. Nesterov on (12) to solve (5). This writes:

Algorithm 1 (Dual)

Choose a number of iterations N.
Set a point u^0.
Set a starting point y^0.
Set ϵ.
Set $\mathcal{A} = 0$, $\eta = 0$, $\bar{u} = 0$ and $y = y^0$.
for $k = 0$ to N **do**
  $a = \frac{1}{L} + \sqrt{\frac{1}{L^2} + \frac{2}{L}\mathcal{A}}$
  $v = \Pi_Y\left(y^0 - \eta\right)$
  $z = \frac{\mathcal{A}y + av}{\mathcal{A} + a}$
  $y = \Pi_Y\left(z - \frac{\nabla\Psi_\epsilon(z)}{L}\right)$
  $\bar{u} = \bar{u} + a\left(-\frac{D^* y}{\epsilon} + u^0\right)$
  $\eta = \eta + a\nabla\Psi_\epsilon(y)$
  $\mathcal{A} = \mathcal{A} + a$
end for
Set $\bar{u}^N = \frac{\bar{u}}{\mathcal{A}}$.

where Π_Y is the projector on the set Y:

$$(\Pi_Y(y))_i = \begin{cases} \frac{y_i}{|y_i|} & \text{if } |y_i| > 1 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

Due to the smoothing of the problem, this algorithm can be shown to converge in $O\left(\frac{1}{k}\right)$ [10] while a classical gradient descent on the smoothed primal problem (as the one used in [4]) converges in $O\left(\frac{1}{\sqrt{k}}\right)$. From a practical point of view, Y. Nesterov's algorithm neatly improves the convergence rate of first order method on all the imaging problems we tested. Moreover, A. Nemirovski showed in [11] that this convergence rate is somehow "optimal". We refer the reader to [11] for a detailed description of its optimality.

Now, an important remark is that we use a "regularizing" prior in (5), while the current trend in signal processing consists in using sparsity priors . With our notations this would consist in solving:

$$\arg\min_{c\in\mathbb{R}^q} \left\{ \|A\tilde{W}c - g\|_1 + \lambda\|c\|_1 \right\} \tag{15}$$

where c are the wavelet coefficients of the image u and $\tilde{W} : \mathbb{R}^q \to \mathbb{R}^n$ is the reconstruction wavelet operator. This prior is largely used in image processing as it is known to improve the sparsity of the model [12, 13, 14, 15]. The idea behind this model is that we can represent more complex signals with a very low number of simple atoms if we increase the size of the dictionary W. When this wavelet transform is a decomposition on a basis ($q = n$), both models are equivalent. However when the wavelet transform is overcomplete, this model does not give good results compared to the model (5), as this sparse representation seems to be more sensitive to the presence of noise [16]. In the next section, we give comparisons of both priors for the following denoising problem:

$$\arg\min_{c\in\mathbb{R}^q} \left\{ \|\tilde{W}c - g\|_2^2 + \lambda\|c\|_1 \right\} \tag{16}$$

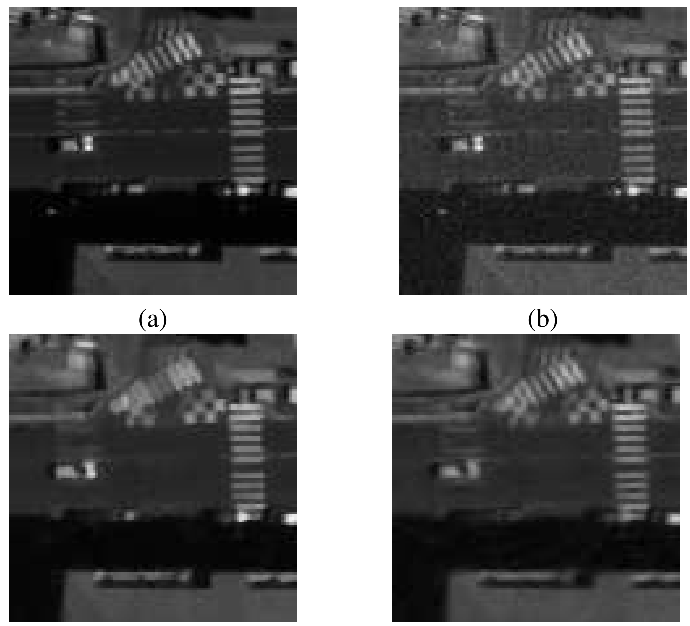
For the same problem, the regularizing prior writes:

$$\arg\min_{u\in\mathbb{R}^n} \left\{ \|u - g\|_2^2 + \lambda\|Wu\|_1 \right\} \tag{17}$$

Both problems can be solved without smoothing. In (16) the proximal operator of the l^1-norm can be computed explicitly and is equal to a soft-thresholding. An iterative thresholding algorithm as [12, 17] can be used to solve it. We use the algorithm 1 to solve (17).

4. RESULTS

Results of the proposed algorithm for the problem of irregular sampling are shown on figure 1. Due to space limitations, the original image in our figures is the regularly sampled image (i.e. the expected result). We can see that the image retrieved with the proposed method allows to retrieve more thin details compared to the one obtained using the TV regularization (look at the diagonal zebra crossing on figures (c) and (d)). For very noisy images, we could check that this regularization gives some artifacts and slightly blurs the image. Small elements may thus lose intensity.
Comparison of the two different priors is shown on the figure 2. As we use the DTCW transform, (16) and (17) are not equivalent. For this denoising problem, we get similar $PSNR$, but we can check that the results are different from a perceptual point of view. We can see that the sparse prior gives more artifacts than the regularizing prior. The authors of [16] consider the same problem and also get better results with the regularizing prior.

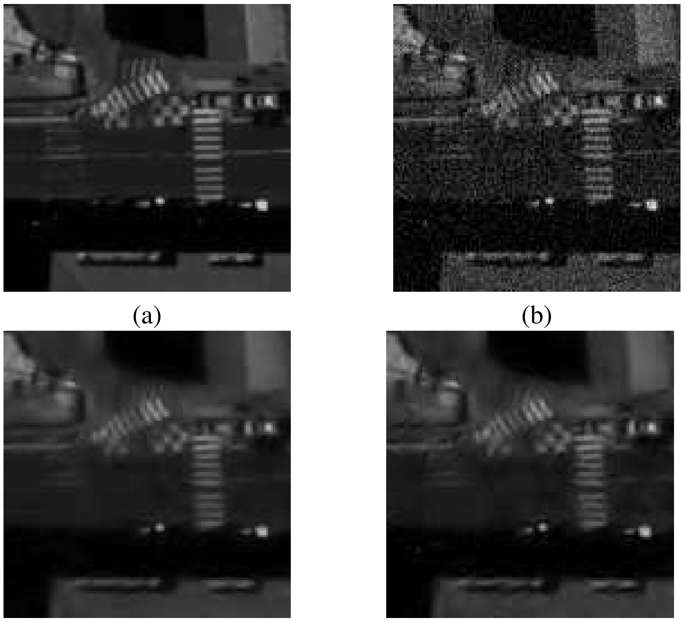


(a) (b) (c) (d)

Fig. 1. Restoration of an irregularly sampled, blurred and noisy image. (a) Original image ©CNES, (b) distorted image (Gaussian noise, $SNR = 15.62\ dB$), (c) result with the TV regularization ($SNR = 24.09\ dB$), and finally (d) result with the DTCW regularization ($SNR = 24.42\ dB$).

5. CONCLUSION

We have proposed a new method for solving restoration problems in image processing using a variational approach. We used the l^1-norm of a complex wavelet transform as a prior. This method has proven to be be really efficient to restore thin details and to remove noise compared to the TV regularization which smooths the oriented textures of the image. To the best of our knowledge, only few results are provided in image deconvolution with wavelet regularization (non orthogonal basis) as the minimization is very time consuming. In this paper, we use a fast algorithm to solve this problem. We also make a comparison of two widely used priors in im-



(a) (b) (c) (d)

Fig. 2. Comparison of the two different priors on a denoising problem. (a) Original image ©CNES, (b) distorted image (Gaussian noise, $PSNR = 22.65\ dB$), (c) result with the regularizing prior ($PSNR = 28.40\ dB$), and finally (d) result with the sparse prior ($PSNR = 28.11\ dB$).

age processing. This simple experiment shows that in some cases the regularizing prior gives better results than the sparsity prior. We will try to analyze the differences between these models from a theoretical point of view.

6. REFERENCES

[1] G. Facciolo, A. Almansa, J. F. Aujol, and V. Caselles, "Irregular to regular sampling, denoising, and deconvolution," *Multiscale Modeling & Simulation*, vol. 7, no. 4, pp. 1574–1608, 2009.

[2] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[3] L.I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proc. IEEE International on Image Processing (ICIP)*, Austin, USA, Nov 1994, vol. 1, pp. 31–35.

[4] E. Bughin, L. Blanc-Féraud, and J. Zerubia, "Satellite image reconstruction from an irregular sampling," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, USA, 2008.

[5] H. Haoying Fu, M. Ng, M. Nikolova, and J. Barlow, "Efficient minimization methods of mixed $l^1 - l^1$ and $l^2 - l^1$ norms for image restoration," *SIAM Journal on Scientific computing*, vol. 27, no. 6, 2006.

[6] M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. Math. Imaging Vis.*, vol. 20, no. 1-2, pp. 99–120, 2004.

[7] G. Beylkin, "On the fast Fourier transform of function with singularities.," *Applied and Computational Harmonic Analysis*, vol. 2, no. 4, pp. 363–381, 1995.

[8] R. G. Baraniuk I. W. Selesnick and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov 2005.

[9] Y. Nesterov, "Gradient methods for minimizing composite objective function," *CORE Discussion Paper 2007/76*, 2007.

[10] P. Weiss, *Algorithmes rapides d'optimisation convexe. Application la reconstruction d'images et la détection de changements*, Ph.D. thesis, Université de Nice Sophia Antipolis, 2008.

[11] A. Nemirovski, "Information-based complexity of linear operator equations," *J. Complex.*, vol. 8, no. 2, pp. 153–175, 1992.

[12] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[13] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Processing*, vol. 7, pp. 319–335, 1996.

[14] J.-L. Starck, M. Elad, and D.L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *Image Processing, IEEE Transactions on*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.

[15] S. S. Chen, D.L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[16] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[17] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V.R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, pp. 1495–1518, 2007.