A GENERALIZATION OF NEGATIVE NORM MODELS IN THE DISCRETE SETTING

Application to Stripe Denoising

Jérôme Fehrenbach[1], Pierre Weiss[1] and Corinne Lorenzo[2]
[1]*Institut de Mathématiques de Toulouse, Toulouse University, Toulouse, France*
[2]*ITAV, Toulouse canceropole, Toulouse, France*

Keywords: Denoising, Cartoon+texture decomposition, Primal-dual algorithm, Stationary noise, Fluorescence microscopy.

Abstract: Starting with a book of Y.Meyer in 2001, negative norm models attracted the attention of the imaging community in the last decade. Despite numerous works, these norms seem to have provided only luckwarm results in practical applications. In this work, we propose a framework and an algorithm to remove stationary noise from images. This algorithm has numerous practical applications and we show it on 3D data from a newborn microscope called SPIM. We also show that this model generalizes Meyer's model and its successors in the discrete setting and allows to interpret them in a Bayesian framework. It sheds a new light on these models and allows to pick them according to some a priori knowledge on the texture statistics. Further results are available on our webpage at http://www.math.univ-toulouse.fr/∼weiss/PagePublications.html.

1 INTRODUCTION

The purpose of this article is to provide variational models and algorithms in order to remove *stationary noise* from images. The models that are proposed here turn out to be a generalization of the discretized negative norm models. This allows to analyse them in a Bayesian framework. By *stationary noise*, we mean that the noise is generated by convolving white noise with a given kernel. The noise thus appears as "structured" in the sense that some pattern might be visible, see Figure 3(b),(c),(d).

This work was primarily motivated by the recent development of a microscope called Selective Plane Illumination Microscope (SPIM). The SPIM is a fluorescence microscope which allows to perform optical sectioning of a specimen, see (Huisken et al., 2004). One difference with conventional microscopy is that the fluorescence light is detected at an angle of 90 degrees with the illumination axis. This procedure tends to degrade the images with stripes aligned with the illumination axis, see Figure 5(a). This kind of noise is well described by a stationary process. The first contribution of this paper is to provide effective denoising algorithms dedicated to this imaging modality.

It appears that our models generalize the negative norms models proposed by Y. Meyer (Meyer, 2001). This work initiated numerous research in the domain of texture+cartoon decomposition methods (Vese and Osher, 2003; Osher et al., 2003; Aujol et al., 2006; Garnett et al., 2007). Meyer's idea is to decompose an image into a piecewise smooth component and an oscillatory component. The use of a negative norm $\|\cdot\|_N$ to capture oscillating patterns is motivated by the fact that if (v_n) converges weakly to 0 then $\|v_n\|_N \to 0$. This interpretation is however not really informative on what kind of textures are well captured by negative norms. The second contribution of this paper is to propose a Bayesian interpretation of these models in the discrete setting. This allows a better understanding of the decomposition models:

- We can associate a probability density functions (p.d.f.) to the negative norms. This allows to choose a model depending on some a priori knowledge on the texture.
- We can synthetize textures which are adapted to these negative norms.
- The Bayesian interpretation suggests a new broader and more versatile class of translation invariant models used e.g. for SPIM imaging.

Connection to Previous Works. This work shares flavors with some previous works. In (Aujol et al.,

Fehrenbach J., Weiss P. and Lorenzo C. (2012).
A GENERALIZATION OF NEGATIVE NORM MODELS IN THE DISCRETE SETTING - Application to Stripe Denoising.
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, pages 337-342
DOI: 10.5220/0003742603370342

2006) the authors present algorithms and results using similar approaches. However, they do not propose a Bayesian interpretation and consider a narrower class of models. An alternative way of decomposing images was proposed in (Starck et al., 2005). The idea is to seek components that are sparse in given dictionaries. Different choices for the elementary atoms composing the dictionary will allow to recover different kind of textures. See (Fadili et al., 2010) for a review of these methods and a generalization to the decomposition into an arbitrary number of components.

The main novelties of the present work are:

1. We do not restrict to sparse components, but allow for a more general class of random processes.
2. Similarly to (Fadili et al., 2010), the texture is described through a dictionary. In the present work each dictionary is composed of a single pattern shifted in space, ensuring translation invariance.
3. A Bayesian approach takes into account the statistical nature of textures more precisely.
4. The decomposition problem is recast into a convex optimization problem that is solved with a recent algorithm (Chambolle and Pock, 2011) allowing to obtain results in an interactive time.
5. Codes are provided on our webpage http://www.math.univ-toulouse.fr/∼weiss/PageCodes.html.

Notation: Let u be a gray-scale image. It is composed of $n = n_x \times n_y$ pixels, and $u(\mathbf{x})$ denotes the intensity at pixel $\mathbf{x}$. The convolution product between u and v is $u * v$. The discrete gradient operator is denoted ∇. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex closed function (see (Rockafellar, 1970)). $\partial\varphi$ denotes its sub-differential. The Fenchel conjugate of φ is denoted $\varphi^\star$, and its resolvent is defined by:

$$(Id + \partial\varphi)^{-1}(u) = \arg\min_{v \in \mathbb{R}^n} \varphi(v) + \frac{1}{2}\|v - u\|_2^2,$$

2 NOISE MODEL

Our objective can be formulated as follows: we want to recover an original image u, given an observed image $u_0 = u + b$, where b is a sample of some random process.

The most standard denoising techniques explicitly or implicitly assume that the noise is the realization of a random process that is pixelwise independent and identically distributed (i.e. a white noise). Under this assumption, the maximum a posteriori (MAP) approach leads to optimization problems of kind:

$$\text{Find } u \in \arg\min_{u \in \mathbb{R}^n} J(u) + \sum_{\mathbf{x}} \phi(u(\mathbf{x}) - u_0(\mathbf{x})),$$

where

1. $\exp(-\phi)$ is proportional to the p.d.f. of the noise at each pixel,
2. $J(u)$ is an image prior.

The assumption that the noise is i.i.d. appears too restrictive in some situations, and is not adapted to structured noise (see Figures 2 and 3).

The general model of noise considered in this work is the following:

$$b = \sum_{i=1}^{m} \lambda_i * \psi_i, \tag{1}$$

where $\{\psi_i\}_{i=1}^m$ are filters that describe patterns of noise, and $\{\lambda_i\}_{i=1}^m$ are samples of white noise processes $\{\Lambda_i\}_{i=1}^m$. Each process Λ_i is a set of n i.i.d. random variables with a p.d.f. $\exp(-\phi_i)$.

In short, the convolution that appears in the right-hand side of (1) states that the noise b is composed of a certain number of patterns $\psi_1, \ldots, \psi_m$ that are replicated in space. The noise b in (1) is a wide sense stationary noise (Shiryaev, 1996). Examples of noises that can be generated using this model are shown in Figure 3.

1. example (b) is a Gaussian white noise. It is the convolution of a Gaussian white noise with a Dirac delta function.
2. example (c) is a sine function in the x direction. It is a sample of a uniform white noise in $[-1, 1]$ convolved with the filter that is constant equal to $1/n_y$ in the first column and zero otherwise.
3. example (d) is composed of a single pattern that is located at random places. It is the convolution of a sample of a Bernoulli process with the elementary pattern.

3 RESTORATION ALGORITHM

The Bayesian approach requires a p.d.f. on the space of images. We assume that the probability of an image u reads $\mathbf{p}(u) \propto \exp(-J(u))$. In this work we will consider priors of the form:

$$J(u) = \alpha\|\nabla u\|_{1,\varepsilon},$$

where $\alpha > 0$ is a fixed parameter and if $q = (q_1, q_2) \in \mathbb{R}^{n\times 2}$, $\|q\|_{1,\varepsilon} = \sum_{\mathbf{x}} f_\varepsilon\left(\sqrt{q_1(\mathbf{x})^2 + q_2(\mathbf{x})^2}\right)$, with

$$f_\varepsilon(t) = \begin{cases} |t| & \text{if } |t| \geq \varepsilon \\ |t|^2/2\varepsilon + \varepsilon/2 & \text{otherwise.} \end{cases}$$

Note that $\lim_{\varepsilon\to 0}\|\nabla u\|_{1,\varepsilon} = TV(u)$ is the discrete total variation of u, and that $\lim_{\varepsilon\to+\infty}\varepsilon\|\nabla u\|_{1,\varepsilon} = \frac{1}{2}\|\nabla u\|_2^2$ up to an additive constant. This model thus includes TV and H^1 regularizations as limit cases.

The maximum a posteriori approach in a Bayesian framework leads to retrieve the image u and the weights $\{\lambda_i\}_{i=1}^m$ that maximize the conditional probability

$$\mathbf{p}(u,\lambda_1,\dots,\lambda_m|u_0) = \frac{\mathbf{p}(u_0|u,\lambda_1,\dots,\lambda_m)\mathbf{p}(u,\lambda_1,\dots,\lambda_m)}{\mathbf{p}(u_0)}.$$

By assuming that the image u and the noise components λ_i are samples of independent processes, standard arguments show that maximizing $\mathbf{p}(u,\lambda_1,\dots,\lambda_m|u_0)$ amounts to solving the following minimization problem:

$$\text{Find } \{\lambda_i^\star\}_{i=1}^m \in \arg\min_{\{\lambda_i\}_{i=1}^m} \sum_{i=1}^m \phi_i(\lambda_i) + F(\nabla(\sum_{i=1}^m \lambda_i * \psi_i)), \tag{2}$$

where

$$F(q) = \alpha\|\nabla u_0 - q\|_{1,\varepsilon}.$$

The denoised image $u^\star$ is then $u^\star = u_0 - \sum_{i=1}^m \lambda_i^\star * \psi_i$.

We propose in this work to solve problem (2) with a primal-dual algorithm developed in (Chambolle and Pock, 2011). Let A be the following linear operator:

$$A: \begin{array}{ccc} \mathbb{R}^{m\times n} & \to & \mathbb{R}^{n\times 2} \\ \lambda & \mapsto & \nabla\left(\sum_{i=1}^m \lambda_i * \psi_i\right). \end{array}$$

By denoting

$$G(\lambda) = \sum_{i=1}^m \phi_i(\lambda_i) \tag{3}$$

problem (2) can be recast as the following convex-concave saddle-point problem:

$$\min_{\lambda\in\mathbb{R}^{n\times m}} \max_{\|q\|_\infty\le 1} \langle A\lambda, q\rangle - F^*(q) + G(\lambda). \tag{4}$$

We denote $\Delta(\lambda,q)$ the duality gap of this problem (Rockafellar, 1970). This problem is solved using the following algorithm (Chambolle and Pock, 2011): In practice, for a correct choice of inner products and parameters σ and τ, this algorithm requires around 50 low-cost iterations for $\varepsilon = 10^{-3}$. More details will be provided in a forthcoming research report.

Algorithm 1: Primal-Dual algorithm.

Input:
ε: the desired precision;
(λ_0, q_0): a starting point;
Output:
λ_ε: an approximate solution to problem (4).
begin
  $n = 0$; $\bar{\lambda}_0 = \lambda_0$;
  while $\Delta(\lambda_n, q_n) > \varepsilon\Delta(\lambda_0, q_0)$ **do**
    $q_{n+1} = (\mathrm{Id} + \sigma\partial F^*)^{-1}(q_n + \sigma A\bar{\lambda}_n)$;
    $\lambda_{n+1} = (\mathrm{Id} + \tau\partial G)^{-1}(\lambda_n - \tau A^* q_{n+1})$;
    $\bar{\lambda}_{n+1} = \lambda_{n+1} + \theta(\lambda_{n+1} - \lambda_n)$;
    $n = n+1$;
  end
end

4 BAYESIAN INTERPRETATION OF THE DISCRETIZED NEGATIVE NORM MODELS

In the last decade, the texture+cartoon decomposition models based on negative norms attracted the attention of the scientific community. These models often take the following form:

$$\inf_{u\in BV(\Omega), v\in V, u+v=u_0} TV(u) + \|v\|_N \tag{5}$$

where:

- u_0 is an image to decompose as the sum of a texture v in V and a structure u in the space of bounded variation functions $BV(\Omega)$,
- V is a Sobolev space of negative index,
- $\|\cdot\|_N$ is an associated semi-norm.

Y. Meyer's seminal model consists in taking $V = W^{-1,\infty}$ and the following norm:

$$\|v\|_N = \|v\|_{-1,\infty} = \inf_{g\in L^\infty(\Omega)^2, div(g)=v} \|g\|_\infty.$$

In the discrete setting the negative norm models read:

$$\text{Find } (u,g) \in \begin{array}{cc} \arg\min & \|g\|_p + \alpha\|\nabla u\|_1 \\ \text{subject to} & u_0 = u + v \\ & v = \nabla^T g \end{array} \tag{6}$$

where u_0, u and v are in $\mathbb{R}^n$, $g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \in \mathbb{R}^{n\times 2}$ and $\nabla^T g = \partial_1^T g_1 + \partial_2^T g_2$. The operators ∂_1 and ∂_2 denote the discrete derivatives with respect to both space directions. If $p = \infty$, we get the discrete Meyer model. From an experimental point of view, the choices $p = 2$ and $p = 1$ seem to provide better practical results (Vese and Osher, 2003).

In order to show the equivalence of these models with the ones proposed in Equation (2), we express the differential operators as convolution products. As the discrete derivative operators are usually translation invariant, this reads:

$$\nabla^T g = h_1 * g_1 + h_2 * g_2.$$

where $*$ denotes the convolution product and h_1 and h_2 are derivative filters (typically $h_1 = (1,-1)$ and $h_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$).

This simple remark leads to an interesting interpretation of g: *it represents the coefficients of an image v in a dictionary composed of the vectors h_1 and h_2 translated in space*.

The negative norms models can thus be interpreted as decomposition models in a very simple texture dictionary. Next, let us show that problem (5) can be interpreted in a MAP formalism.

Let us define a probability density function:

Definition 1 (Negative Norm p.d.f.)**.** *Let Γ be a random vector in $\mathbb{R}^n$ and Θ be a random vector in $[0,2\pi]^n$. Let us assume that $\mathbf{p}(\Gamma) \propto \exp(-\|\Gamma\|_p)$ and that Θ has a uniform distribution. These two random vectors allow to define a third one:*

$$G = \begin{pmatrix} \Gamma\cos(\Theta) \\ \Gamma\sin(\Theta) \end{pmatrix}.$$

Now let us show that problem (6) actually corresponds to a MAP decomposition. Let us assume that:

$$u_0 = u + v$$

with u and v realization of independant random vector such that $\mathbf{p}(u) \propto \exp(-\alpha\|\nabla u\|_1)$ and $v = \nabla^T g$ with g a realization of G. Then the classical Bayes reasoning leads to the following equations:

$$\begin{aligned}
&\arg\max_{u\in\mathbb{R}^n, v\in\mathbb{R}^n} \mathbf{p}(u,v|u_0) \\
&= \arg\max_{u\in\mathbb{R}^n, v\in\mathbb{R}^n} \frac{\mathbf{p}(u_0|u,v)\cdot\mathbf{p}(u,v)}{\mathbf{p}(u_0)} \\
&= \arg\max_{u+v=u_0, u\in\mathbb{R}^n, v\in\mathbb{R}^n} \frac{\mathbf{p}(u,v)}{\mathbf{p}(u_0)} \\
&= \arg\min_{u+v=u_0, u\in\mathbb{R}^n, v\in\mathbb{R}^n} -\log(\mathbf{p}(v)) - \log(\mathbf{p}(u)) \\
&= \arg\min_{u+v=u_0, u\in\mathbb{R}^n, v\in\mathbb{R}^n} -\log(\mathbf{p}(v)) + \alpha\|\nabla u\|_1 \\
&= \arg\min_{u+v=u_0, u\in\mathbb{R}^n, v=\nabla^T g} \|g\|_p + \alpha\|\nabla u\|_1
\end{aligned}$$

which is exactly problem (6). Also note that the model above is equivalent to a slight variant of the model defined in Equation (2) in the case $m = 2$:

$$\begin{aligned}
&\arg\min_{u+v=u_0, u\in\mathbb{R}^n, v=\nabla^T g} \|g\|_p + \alpha\|\nabla u\|_1 \\
&= \arg\min_{g=(g_1,g_2)\in\mathbb{R}^{2n}} \|g\|_p + \alpha\|\nabla(u_0 - \nabla^T g)\|_1 \\
&= \arg\min_{(g_1,g_2)\in\mathbb{R}^{2n}} G(g_1,g_2) + F(\nabla(u_0 - h_1 * g_1 - h_2 * g_2))
\end{aligned}$$

where

- $G(g_1,g_2) = \left(\sum_{\mathbf{x}} (g_1(\mathbf{x})^2 + g_2(\mathbf{x})^2)^{p/2}\right)^{1/p}$ is a mixed-norm variant of the function G defined in Equation (3) (Kowalski, 2009),
- $F(q) = \alpha\|q\|_1$,
- the filters h_1 and h_2 are the discrete derivative filters defined above.

The same reasoning holds for most negative norms models proposed lately (Meyer, 2001; Aujol et al., 2006; Vese and Osher, 2003; Osher et al., 2003; Garnett et al., 2007), and problem (2) actually generalizes all these models. To our knowledge, the Chambolle-Pock implementation (Chambolle and Pock, 2011) proposed here or the ADMM method (Ng et al., 2010) (for strongly monotone problems) are the most efficient numerical approaches.

5 NEGATIVE NORM TEXTURE SYNTHESIS

The MAP approach to negative norm models described above also sheds a new light on the kind of texture appreciated by the negative norms. In order to synthetize a texture with p.d.f. (1), it suffices to run the following algorithm:

1. Generate a sample of a uniform random vector $\theta \in [0,2\pi]^n$.
2. Generate a sample of a random vector γ with p.d.f. proportional to $\exp(-\|\gamma\|_p)$.
3. Generate two vectors $g_1 = \gamma\cos(\theta)$ and $g_2 = \gamma\sin(\theta)$.
4. Generate the texture $v = \nabla^T \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$.

The results of this simple algorithm are presented in Figure 1.

6 RESULTS OF THE DENOISING ALGORITHM

6.1 Synthetic Image

The method was validated on a synthetic example, where a ground truth is available. A synthetic image was created by adding to a cartoon image (a disk) the sum of 3 different stationary noises. The resulting synthetic image is shown in Figure 2. The cartoon image and the 3 noise components are presented in Figure 3(a,b,c,d). The first noise component is a sample of a Gaussian white noise. The second component is a sine function in the horizontal direction. The third component is the sum of elementary patterns, this is a sample of a Bernoulli law with probability 5.10^{-4} convolved with an elementary filter.

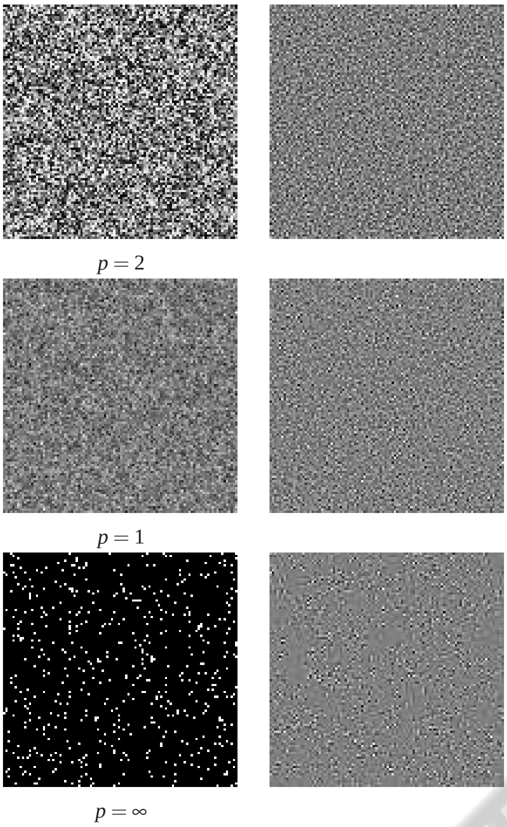

Figure 1: Left: standard noises. Right: different textures synthetized with the negative norm p.d.f. Note: we synthetize the "Laplace" noise by approximating it with a Bernoulli process.

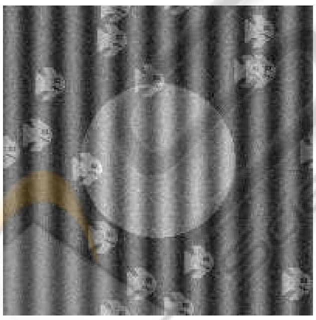

Figure 2: Synthetic image used for the toy example.

The results of Algorithm 1 are presented in Figure 3(e,f,g,h). The decomposition is almost perfect. This example is a good proof of concept.

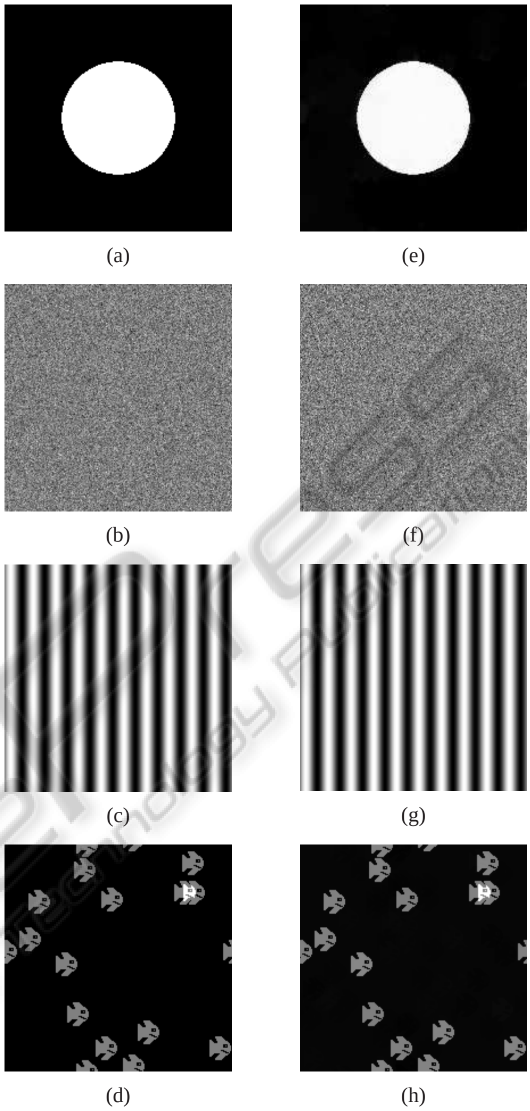

Figure 3: Toy example. Left column: real components; right column: estimated components using our algorithm. (a,e): cartoon component - (b,f): Gaussian noise, std 0.2 - (c,g): Stripes component (sine)- (d,h): 'Poisson' noise component (*poisson* means *fish* in French).

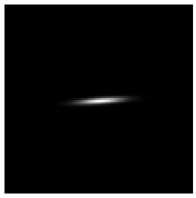

Figure 4: A detailed view of filter ψ_2.

6.2 Real SPIM Image

Algorithm 1 was applied to a zebrafish embryo image obtained using the SPIM microscope. Two filters ψ_1 and ψ_2 were used to denoise this image. The first filter ψ_1 is a Dirac (which allows the recovery of Gaussian white noise), and the second filter ψ_2 is an anisotropic Gabor filter with principal axis directed by the stripes (this orientation was obtained by user). The filter ψ_2 is shown in Figure 4.

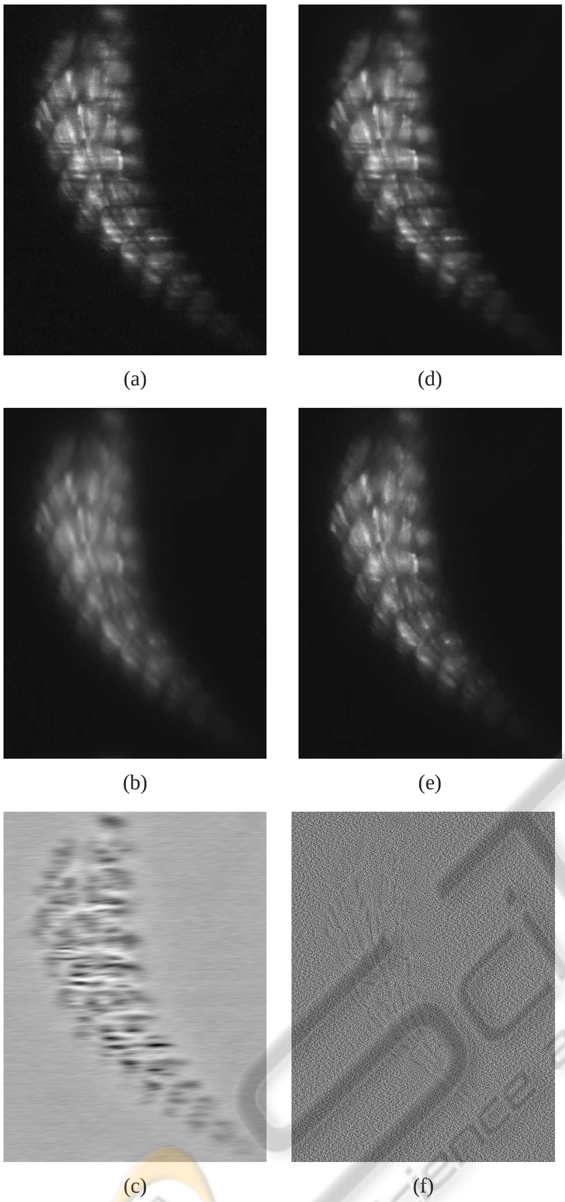

Figure 5: Top-Left: original image zebrafish embryo Tg.SMYH1:GFP Slow myosin Chain I specific fibers - Top-Right: TV-L2 denoising - Mid-Left: H^1-Gabor restoration - Mid-Right: TV-Gabor restoration - Bottom-Left: stripes identified by our algorithm - Bottom-Right: white noise.

The original image is presented in Figure 5(a), and the result of Algorithm 1 is presented in Figure 5(e). We also present a comparison with two other algorithms in Figures 5(d,b):

- a standard TV-L^2 denoising algorithm. The algorithm is unable to remove the stripes as the prior is unadapted to the noise.
- an "H^1-Gabor" algorithm which consists in setting $F(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ in equation (2). The image prior thus promotes smooth solutions and provides blurry results.

ACKNOWLEDGEMENTS

The authors wish to thank Julie Batut for providing the images of the zebrafish. They also thank Valérie Lobjois, Bernard Ducommun, Raphael Jorand and François De Vieilleville for their support during this work.

REFERENCES

Aujol, J.-F., Gilboa, G., Chan, T., and Osher, S. (2006). Structure-texture image decomposition - modeling, algorithms, and parameter selection. *Int. J. Comput. Vision, vol. 67(1), pp. 111-136*.

Chambolle, A. and Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis. 40(1): 120-145*.

Fadili, M., Starck, J.-L., Bobin, J., and Moudden, Y. (2010). Image decomposition and separation using sparse representations: an overview. *Proc. of the IEEE, Special Issue: Applications of Sparse Representation, vol. 98(6), pp. 983-994,*.

Garnett, J., Le, T., Meyer, Y., and Vese, L. (2007). Image decompositions using bounded variation and generalized homogeneous besov spaces. *Appl. Comput. Harmon. Anal., 23, pp. 25-56*.

Huisken, J., Swoger, J., Bene, F. D., Wittbrodt, J., and Stelzer, E. (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science, vol. 305, 5686, p.1007*.

Kowalski, M. (2009). Sparse regression using mixed norms. *Appl. Comput. Harmon. A. 27, 3, 303–324*.

Meyer, Y. (2001). *Oscillating patterns in image processing and in some nonlinear evolution equations, in 15th Dean Jacqueline B. Lewis Memorial Lectures*. AMS.

Ng, M., Weiss, P., and Yuan, X.-M. (2010). Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. *SIAM Journal on Scientific Computing*, 32.

Osher, S., Sole, A., and Vese, L. (2003). Image decomposition and restoration using total variation minimization and the h^{-1} norm. *SIAM Multiscale Model. Sim. 1(3), pp. 339-370*.

Rockafellar, T. (1970). *Convex Analysis*. Princeton University Press.

Shiryaev, A. (1996). *Probability, Graduate Texts in Mathematics 95*. Springer.

Starck, J., Elad, M., and Donoho, D. (2005). Image decomposition via the combination of sparse representations and a variational approach. *IEEE Trans. Im. Proc., vol. 14(10)*.

Vese, L. and Osher, S. (2003). Modeling textures with total variation minimization and oscillating patterns in image processing. *J. Sci. Comput., 19(1-3), pp. 553-572*.