

Extraction de textures localement parallèles par un espace de Hilbert adapté

Pierre MAUREL¹, Jean-François AUJOL¹, Gabriel PEYRÉ²

¹CMLA, ENS Cachan, CNRS, UniverSud, 61 avenue du Président Wilson, 94235 Cachan Cedex, France

²Ceremade, Université Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France
{maurel, aujol}@cmla.ens-cachan.fr, gabriel.peyre@ceremade.dauphine.fr

Résumé – Ce papier³ décrit un nouvel algorithme de séparation d’une image en la somme d’une composante géométrique et d’une composante texturée. Les textures présentes dans l’image sont modélisées à l’aide d’un espace de Hilbert à poids. Les poids sont calculés sur une décomposition de Fourier locale, et permettent de favoriser certaines orientations et fréquences présentes localement dans les motifs texturés. Un algorithme itératif adapte la norme Hilbertienne à l’image à analyser, de façon à optimiser la séparation de la texture par rapport aux contours. L’adaptation de la norme de texture permet d’améliorer l’état de l’art en séparation. Cette nouvelle méthode permet également de restaurer une image bruitée grâce à l’extraction des motifs texturés.

Abstract – This article presents a new algorithm to decompose an image into a cartoon layer and a locally parallel texture layer. Locally parallel oscillating patterns are modeled with a weighted Hilbert space defined over local Fourier coefficients. The weights on the local Fourier atoms are optimized to match the local orientation and frequency of the texture. An iterative algorithm adapts the hilbertian norm to the input image and optimize the separation of the texture and the contours. Numerical results show that our method improves state of the art algorithms for directional and complex textures.

1 Introduction

Une méthode de décomposition variationnelle cherche une décomposition $f = u + v$ d’une image f où u contient les contours géométriques des objets et v les textures présentes dans f . Cette décomposition est la solution du problème de minimisation

$$(u, v) = \operatorname{argmin}_{\tilde{u} + \tilde{v} = f} \frac{1}{2} \|\tilde{v}\|_{\mathbb{T}}^2 + \lambda \|\tilde{u}\|_{\text{TV}}, \quad (1)$$

où $\|\tilde{u}\|_{\text{TV}} = \int |\nabla \tilde{u}|$ est la variation totale de \tilde{u} et $\|\tilde{v}\|_{\mathbb{T}}$ est une norme dite «de texture» qui est petite pour les textures que nous souhaitons modéliser. Le paramètre d’échelle λ contrôle l’influence de chacun des termes, de sorte qu’augmenter λ permet d’extraire plus de textures. Rudin, Osher et Fatemi [7] ont proposé de capturer le bruit présent dans une image en utilisant la norme L^2 usuelle : $\|v\|_{\mathbb{T}}^2 = \int |v|^2$. Meyer [4] a proposé une norme plus sophistiquée pour capturer les motifs fortement oscillants d’une image. Cette idée a inspiré un certain nombre d’auteurs, voir par exemple [1, 6, 5]. Aujol et Gilboa [2] utilisent une norme Hilbertienne à poids qui favorise certaines fréquences globales dans l’image. Starck, Elad et Donoho [8] utilisent une norme qui favorise la parcimonie de la décomposition de la texture dans une transformée en cosinus locaux.

3. Ce travail a été effectué avec le soutien de l’Agence Nationale de la Recherche (ANR), projet NATIMAGES (ANR-08-EMER-009), "Adaptivité pour les images naturelles et les textures".

Cet article modélise les textures oscillantes et localement parallèles afin d’extraire des motifs oscillants avec une variabilité spatiale et fréquentielle. On commence par définir la norme de texture $\|\cdot\|_{\mathbb{T}} = \|\cdot\|_{\xi}$ en fonction d’un paramètre $\xi(x)$ qui est un champ de fréquences instantanées. Pour un point x de l’image, $\xi(x)$ indique la fréquence $\|\xi(x)\|$ et l’orientation $\xi(x)/\|\xi(x)\|$ locales de la texture autour de x . La norme $\|\cdot\|_{\xi}$ est faible pour un motif texturé présent autour du point x si sa fréquence principale est voisine de $\xi(x)$. Ce paramètre $\xi(x)$ sera alors optimisé en même temps que la décomposition de f .

Dans la suite, nous considérons une image discrète $f \in \mathbb{R}^N$ de $N = n \times n$ pixels.

2 Décomposition d’images par un modèle TV-Hilbert localement adaptatif

Décomposition TV-Hilbert : Aujol et Gilboa [2] proposent d’utiliser une norme de Hilbert $\|v\|_{\mathbb{T}}^2 = \langle Kv, v \rangle$ définie par une noyau K positif et symétrique. Cette norme peut être calculée dans une trame ("frame") $\{\psi_{\ell}\}_{1 \leq \ell \leq P}$ composée de $P \geq N$ vecteurs de \mathbb{R}^N . L’opérateur de trame $\Psi : \mathbb{R}^N \rightarrow \mathbb{R}^P$ calcule la décomposition de f dans cette famille $\Psi f = \{\langle f, \psi_{\ell} \rangle\}_{\ell=1}^P \in \mathbb{R}^P$. Étant donné un ensemble de poids positifs $\gamma_{\ell} \geq 0$, une norme pondérée est définie comme

$$\|f\|_{\mathbb{T}}^2 = \sum_{\ell} \gamma_{\ell}^2 |\langle f, \psi_{\ell} \rangle|^2 = \|\Gamma \Psi f\|_{L^2}^2, \quad (2)$$

où $\Gamma = \text{diag}_\ell(\gamma_\ell) \in \mathbb{R}^{P \times P}$. Cette norme correspond à un espace de Hilbert associé au noyau $K = \Psi^* \Gamma^2 \Psi$, où $\Psi^* \in \mathbb{R}^{N \times P}$ est l'adjoint de $\Psi \in \mathbb{R}^{P \times N}$ et la minimisation (1) se réécrit :

$$u = \underset{\tilde{u} \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2} \|\Gamma \Psi \tilde{u} - y\|_{L^2}^2 + \lambda \|\tilde{u}\|_{\text{TV}} \quad \text{avec} \quad y = \Gamma \Psi f. \quad (3)$$

Décomposer une image par une minimisation TV-Hilbert requiert donc la définition d'une transformée Ψ conçue pour capturer les textures oscillantes, et le calcul de poids Γ adaptés à la texture que l'on souhaite extraire de l'image f . La construction originale de [2] utilise la base de Fourier discrète. Nous utilisons à la place une trame redondante de Fourier locale afin de capturer les textures localement parallèles dont l'orientation et la fréquence peuvent varier dans l'espace.

Trame de Fourier locale : Un atome de Fourier localisé autour du point $x_p = p\Delta_x$ et avec une fréquence $\xi_k = k\Delta_\xi = k(n/q)$ est défini par $\forall k \in \{0, \dots, q-1\}^2$, $p \in \{0, \dots, n/\Delta_x\}^2$,

$$\psi_{p,k}[y] = q^{-1} g[y - p\Delta_x] e^{\frac{2i\pi}{q}(y_1 k_1 + y_2 k_2)},$$

où g est une fenêtre lisse, centrée en 0, dont le support est de taille $q \times q$ pixels avec $q \geq \Delta_x$. La trame de Fourier locale $\{\psi_{p,k}\}_{p,k}$ est une famille redondante de $P = (q/\Delta_x)^2 N$ vecteurs de \mathbb{R}^N . La décomposition $\Psi f = \{\langle f, \psi_{p,k} \rangle\}_{p,k} \in \mathbb{R}^P$ d'une image f dans cette trame se calcule à l'aide de N/Δ_x^2 transformées de Fourier rapides de taille $q \times q$.

L'opérateur dual Ψ^* reconstruit une image $\Psi^* c \in \mathbb{R}^N$ à partir d'un ensemble de coefficients c . Il se calcule avec N/Δ_x^2 transformées de Fourier rapides inverses. On peut montrer que $\Psi^* \Psi = \text{diag}_x(\sum_y g[\Delta_x y - x]^2)$. Dans la suite, on normalise la fenêtre g de manière à avoir $\sum_y g[\Delta_x y - x]^2 = 1$. Ceci implique que $\Psi^* \Psi = \text{Id}_N$ et donc Ψ^* est le pseudo inverse de Ψ . La famille correspondante $\{\psi_{p,k}\}_{k,p}$ est une trame ajustée ("tight frame") de \mathbb{R}^N et on a à la fois la formule d'analyse-synthèse $f = \sum_{p,k} \langle f, \psi_{p,k} \rangle \psi_{p,k}$ et la conservation de l'énergie $\|f\|^2 = \sum_{p,k} \langle f, \psi_{p,k} \rangle^2$. De plus, l'opérateur de décomposition Ψ est de norme égale à 1 : $\|\Psi\| = 1$.

Choix des poids et définition de la norme de texture : Nous définissons une norme de Hilbert $\|\cdot\|_{\text{T}}$ adaptée aux textures oscillantes comme une norme pondérée sur les coefficients de Fourier locaux. La formule générale (2) appliquée à la famille $\psi_\ell = \psi_{p,k}$ s'écrit $\|f\|_{\text{T}}^2 = \sum_{p,k} \gamma_{p,k}^2 |\langle f, \psi_{p,k} \rangle|^2$, où les $\gamma_{p,k} \geq 0$ pondèrent l'influence de chacun des atomes de Fourier locaux dans le modèle de texture. Un poids $\gamma_{p,k} \geq 0$ doit être petit si l'image f contient une texture de fréquence proche de ξ_k autour du point x_p . Nous considérons un modèle de texture localement oscillant, dans lequel les textures typiques sont donc localement bien approximées par un atome de Fourier unique. C'est à dire que, pour un champ de fréquences instantanées $\xi(\cdot)$ ($\xi(x)$ donnant la fréquence $\|\xi(x)\|$ et l'orientation $\xi(x)/\|\xi(x)\|$ locales de la texture autour de x), le poids $\gamma_{p,k}$ est petit si ξ_k est proche de $\xi(x_p)$ ou de $-\xi(x_p)$. Par convention, $\xi(x)$ est choisi

égal à $(0, 0)$ s'il n'y a pas de motifs oscillants autour de x dans l'image.

Les poids sont donc définis en fonction de ξ de la manière suivante : quand il n'y a pas de textures oscillantes autour de x_p ($\xi(x_p) = (0, 0)$), nous choisissons $\gamma_{p,k} = 1$ pour tout k , afin de ne pas privilégier une fréquence arbitraire au cours de l'extraction, dans le cas contraire, on pose

$$\gamma_{p,k}(\xi) = \left(1 - G_\sigma(\|\xi_k + \xi(x_p)\|)\right) \left(1 - G_\sigma(\|\xi_k - \xi(x_p)\|)\right) \quad (4)$$

où $G_\sigma(x) = \exp(-(x/\sigma)^2/2)$ et σ est un paramètre d'échelle qui représente la déviation qu'on s'attend à trouver dans le spectre local de la texture par rapport à la fréquence instantanée (dans nos expériences numériques, nous avons fixé $\sigma = 1$). La norme de texture est donc finalement donnée par :

$$\|v\|_{\text{T}}^2 = \|v\|_{\xi}^2 = \sum_{p,k} \gamma_{p,k}(\xi)^2 |\langle v, \psi_{p,k} \rangle|^2 = \|\Gamma(\xi) \Psi v\|_{L^2}^2. \quad (5)$$

où $\Gamma(\xi) = \text{diag}_{\ell=(p,k)}(\gamma_{p,k}(\xi))$. Il s'agit en fait d'une semi-norme puisque $\|v\|_{\text{T}} = 0$ n'implique pas $v = 0$ mais, par abus de langage, nous utiliserons le terme de norme dans la suite.

Décomposition d'une image bruitée : On cherche une décomposition $f = u + v + w$ d'une image f où u contient les contours géométriques des objets, v les textures présentes dans f et w un éventuel bruit. En modélisant les textures par la norme définie par (5), cette décomposition peut s'écrire comme la solution du problème de minimisation suivant :

$$(u, v, \xi) = \underset{\tilde{u}, \tilde{v}, \xi \in \mathcal{C}}{\text{argmin}} \mu \|\tilde{v}\|_{\xi}^2 + \lambda \|\tilde{u}\|_{\text{TV}} + \frac{1}{2} \|f - \tilde{u} - \tilde{v}\|_{L^2}^2 \quad (6)$$

$$w = f - u - v.$$

\mathcal{C} est un ensemble de contraintes sur ξ . Tout d'abord, pour ne pas extraire de basses fréquences dans la composante de texture la fréquence locale $|\xi|$ doit être suffisamment grande : on impose $|\xi(x)| > \tau$ pour un certain $\tau > 0$. De plus, un motif oscillant de fréquence $\xi(x_p)$ est supposé présent dans l'image f autour du point x_p seulement si $|\langle f, \psi_{p,k} \rangle| > \eta_p$ où $k = \xi(x_p)/\Delta_\xi$ et $\eta_p > 0$ est un paramètre réel positif. En effet, on veut éviter de sélectionner une fréquence arbitraire pour une région de l'image qui ne contient pas de motif oscillant. Pour nos expériences numériques nous avons choisi $\tau = 2/q$, où q est la taille des fenêtres de Fourier locales, et $\eta_p = 2|\overline{\Psi f_p}|$ où $|\overline{\Psi f_p}|$ est la valeur moyenne de $|\langle f, \psi_{p,k'} \rangle|$ pour $k' \in \{-q/2, \dots, q/2 - 1\}^2$. En résumé, \mathcal{C} contient donc les champs $\xi : \mathbb{R}^{N/\Delta_x} \mapsto \mathbb{R}^2$ vérifiant les deux propriétés suivantes :

$$\begin{aligned} (i) \quad & \forall p, \quad |\xi(x_p)| > \tau \\ (ii) \quad & \forall p, \quad (\forall k, |\langle f, \psi_{p,k} \rangle| \leq \eta_p) \Rightarrow \xi(x_p) = (0, 0) \end{aligned} \quad (7)$$

Algorithme : On cherche donc à minimiser (6) avec la norme de texture définie par (5). Cette minimisation est faite successivement sur chacune des variables u, v et ξ .

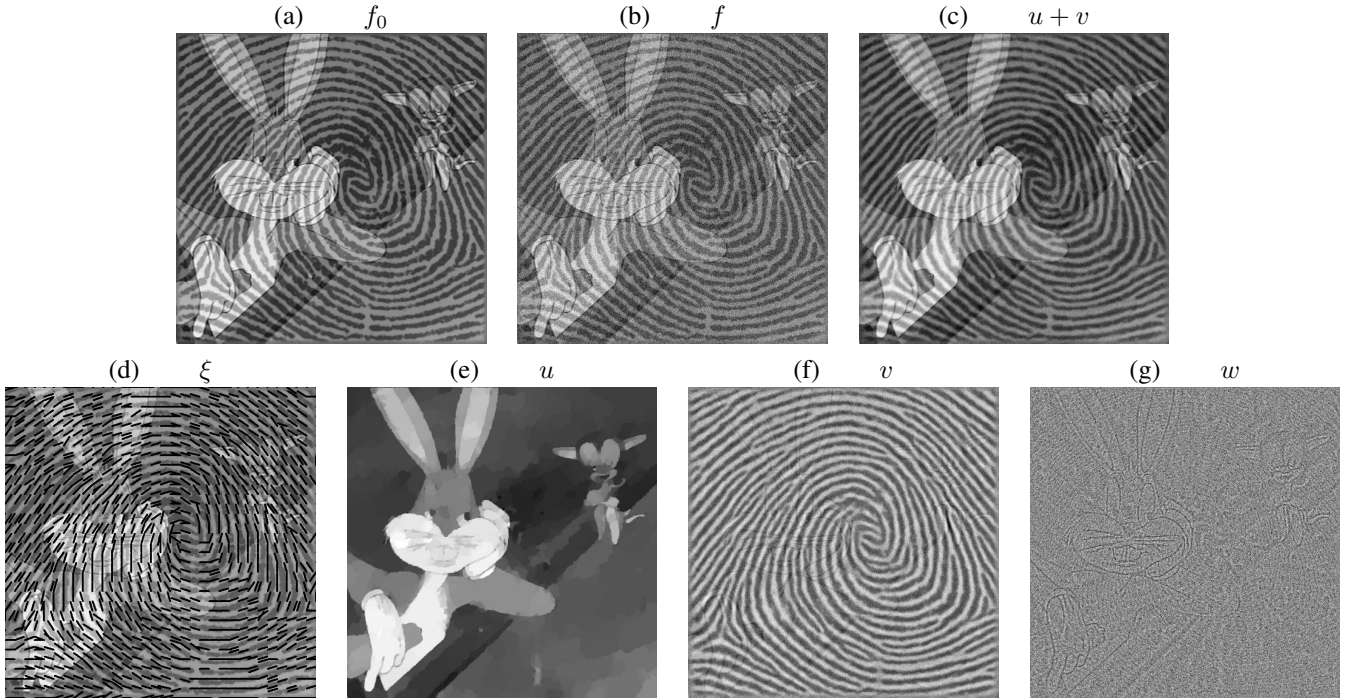


FIGURE 1 – Première ligne : (a) l’image initiale non bruitée, (b) l’image bruitée f et (c) l’image restaurée $u + v$. Deuxième ligne : le résultat de la minimisation, (d) l’estimation des orientations ξ des motifs oscillants et la décomposition en trois composantes, (e) la géométrie u , (f) la texture v et (g) le bruit w , $f = u + v + w$.

- *minimisation sur ξ (u et v sont fixés)* : on cherche le champ de fréquence $\xi^{(i+1)}$ vérifiant :

$$\xi^{(i+1)} = \operatorname{argmin}_{\xi} \|v^{(i)}\|_{\xi}^2.$$

Ceci requiert, pour chaque p , le calcul de $\xi(x_p)^{(i+1)}$ minimisant $\sum_k \gamma_{p,k} (\tilde{\xi}(x_p))^2 |\langle v^{(i)}, \psi_{p,k} \rangle|^2$. Si le σ qui intervient dans la définition des poids est suffisamment petit, cette minimisation revient à calculer $\max_k |\langle v^{(i)}, \psi_{p,k} \rangle|$, ce qui permet d’accélérer les calculs en définissant

$$\xi(x_p)^{(i+1)} = \frac{n}{q} \operatorname{argmax}_{k \in [-q/2, q/2]^2} |\Psi v^{(i)}[p, k]|$$

La figure 2 illustre le principe de cette estimation d’orientations : pour un point x_p donné, une fréquence $\xi(x_p)$ unique est sélectionnée et les poids correspondants $\gamma_{p,k}(\xi)$ sont construits selon (4).

- *minimisation sur u (v et ξ sont fixés)* : on doit minimiser

$$u^{(i+1)} = \operatorname{argmin}_{\tilde{u}} \lambda \|\tilde{u}\|_{\text{TV}} + \frac{1}{2} \|(f - v^{(i)}) - \tilde{u}\|_{L^2}^2. \quad (8)$$

La solution de cette minimisation peut être calculée à l’aide de l’algorithme de Chambolle, voir [3].

- *minimisation sur v (u et ξ sont fixés)* : on doit minimiser

$$v^{(i+1)} = \operatorname{argmin}_{\tilde{v}} \mu \|\Gamma \Psi \tilde{v}\|_{L^2}^2 + \frac{1}{2} \|(f - u^{(i+1)}) - \tilde{v}\|_{L^2}^2, \quad (9)$$

où $\Gamma = \Gamma(\xi^{(i+1)})$ (voir (5)).

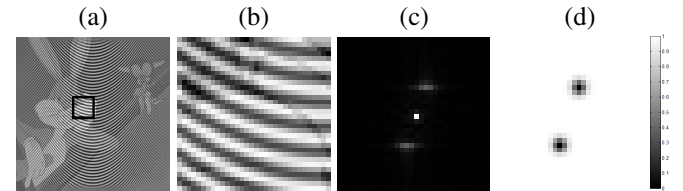


FIGURE 2 – Illustration de l’estimation des orientations : (a) l’image f , (b) la fenêtre autour d’un point x_p , (c) la transformée de Fourier locale correspondante et (d) les poids $\gamma_{p,k}(\xi)$ correspondants au ξ estimé à partir de la transformée.

En calculant le gradient de (9), on obtient que $v^{(i+1)}$ doit satisfaire :

$$(2\mu\Psi^*\Gamma^2\Psi + \text{Id})v^{(i+1)} = f - u^{(i+1)} \quad (10)$$

et donc, la solution peut être obtenue par une descente de gradient conjugué (remarquons que $A = \mu\Psi^*\Gamma^2\Psi + \text{Id}$ est symétrique positive).

3 Résultats numériques

La Figure 1 montre une image f (512×512) composée d’une partie géométrique ("cartoon") et d’une partie texture de type empreinte digitale et dégradée par un bruit gaussien additif. L’image bruitée f est alors décomposée en trois composantes u , v , et w en utilisant notre méthode avec les paramètres suivants $\lambda = 0.1$, $\mu = 0.3$, $q = 16$, $\Delta_x = 4$. Puisque u capture la

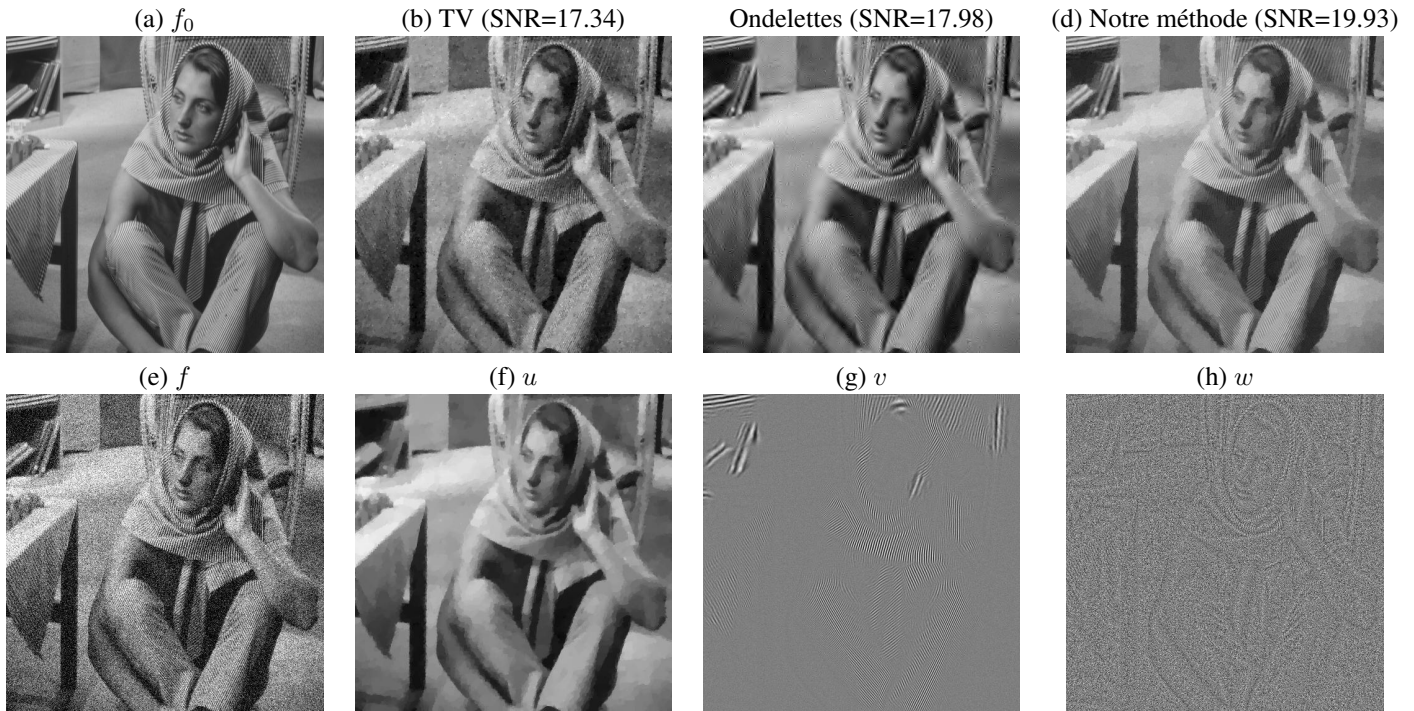


FIGURE 3 – Première ligne : (a) l'image originale f_0 , (b) débruitage par le modèle TV-L2 (λ est choisi pour obtenir le meilleur SNR, $\lambda = 0.1$), (c) débruitage par ondelettes invariantes par translation (le seuil est choisi pour obtenir le meilleur SNR) (d) notre résultat ($u + v$), qui correspond au meilleur SNR $u + v$. Seconde ligne : (e) la donnée bruitée f (SNR = 10.25) et la décomposition (f) la géométrie u , (g) la texture v et (h) le bruit w , $f = u + v + w$.

géométrie de l'image, v les motifs localement parallèles et w le bruit, on peut reconstruire une version débruitée de l'image initiale en additionnant u et v .

Reprenant le même principe, la Figure 3 montre un exemple de résultat obtenu par cette méthode de décomposition et débruitage pour l'image "Barbara". Nous comparons notre résultat avec deux autres méthodes de débruitage. Tous les paramètres sont choisis de manière à obtenir le meilleur rapport signal sur bruit (SNR). La décomposition entre structure et texture fournit une meilleure reconstruction de la texture et donc un meilleur rapport signal sur bruit. Il est important de remarquer que les méthodes de décomposition précédentes (telles que TV- G [4] ou TV- H^{-1} [6]) ne conviennent pas à un processus de débruitage. En effet, les normes G et H^{-1} sont petites pour des motifs contenant des hautes fréquences et sont donc également petites pour une grande partie du bruit. En outre la norme TV pénalise fortement les motifs oscillants et ces modèles ne sont donc pas capable de séparer efficacement la texture du bruit. Notre norme adaptative est au contraire petite pour des motifs qui présentent une certaine fréquence et est donc plus appropriée pour le débruitage.

Références

[1] J.-F. Aujol, G. Aubert, L. Blanc-Feraud, and A. Chambolle. Image decomposition into a bounded variation component

and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1) :71–88, January 2005.

- [2] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *IJCV*, 67(1) :111–136, 2006.
- [3] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20 :89–97, 2004.
- [4] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. American Mathematical Society, Boston, MA, USA, 2001.
- [5] M. Nikolova. A variational approach to remove outliers and impulse noise. *J. Math. Imaging Vis.*, 20(1-2) :99–120, 2004.
- [6] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Modeling & Simulation*, 1(3) :349–370, 2003.
- [7] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4) :259–268, 1992.
- [8] J.-L. Starck, M. Elad, and D.L. Donoho. Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics*, 132, 2004.