Tetrolet Shrinkage with Anisotropic Total Variation Minimization for Image Approximation

Jens Krommweh[a], Jianwei Ma[*,b,c]

[a]*Faculty of Mathematics, University of Duisburg-Essen, Campus Duisburg, 47048 Duisburg, Germany*
[b]*School of Aerospace, Tsinghua University, Beijing 100084, China*
[c]*Centre of Geoscience, Ecole des Mines de Paris, 35 rue Saint-Honore 77305, Fontainebleau, France*

Abstract

In this paper, an anisotropic total variation (ATV) minimization is combined with the new adaptive tetrolet transform for discontinuity-preserving image processing. In order to suppress the pseudo-Gibbs artefacts and to increase the regularity, the conventional shrinkage results are further processed by a total variation (TV) minimization scheme, in which only the insignificant tetrolet coefficients of the image are changed by the use of ATV constrained projection, instead of previous TV projections. Numerical experiments of piecewise-smooth images show the good performance of the proposed hybrid method to recover the shape of edges and important detailed directional components, in comparison to some existing methods.

Key words: anisotropic total variation minimization, directional wavelets, tetrolet transform, post-processing method, image approximation, data compression, sparse representation
2000 MSC: 42C40, 65T60, 94A12, 94A08

1. Introduction

The wavelet transform is one of the most popular tools in image processing due to its promising properties for singularity analysis and efficient computational complexity. However most natural images/signals exhibit discontinuities across curves (so-called line or curve singularities). Although applications of wavelets have become increasingly popular in scientific and engineering fields, traditional wavelets perform well only for representing point singularities, since they ignore the geometric properties of directed structures and do not exploit the regularity of edges. Therefore, wavelet-based compression, denoising, or structure extraction become computationally inefficient for geometric features with line and surface singularities.

The recently proposed geometric wavelets, e.g., curvelets [3, 16], contourlets [12], bandelets [28], shearlets [19], can improve this drawback of traditional wavelets to some extent. Typically,

---

curvelets proposed by Candès, Donoho, Demanet, and Ying [2, 3, 4] allow an almost optimal sparse representation of objects with C^2-singularities. For a smooth object f with discontinuities along C^2-continuous curves, the best m-term approximation f_m by curvelet thresholding obeys $\|f - f_m\|_2^2 \leq Cm^{-2} (\log m)^3$, while for wavelets the decay rate is only m^{-1}. Unfortunately, like wavelets, the geometric wavelets still suffer from the pseudo-Gibbs artefacts and shift/rotation variance, though they attempt to overcome these problems at least partially [14, 24].

The TV-synthesized wavelet transform can effectively overcome these problems. After Chambolle et al. [6] have examined the close relationship between wavelet-based algorithms and variational problems, in the wide field of image processing, there have been proposed many approaches combining TV methods and wavelet transforms to improve the performances. Besides [5, 9], we mention the work by Chan and Zhou who combine wavelet shrinkage and TV minimization for image compression in [8]. In [32] Starck et al. separate images into texture and piecewise smooth parts, exploiting both the variational and the sparsity mechanisms. Their method combines a basis pursuit denoising algorithm and a TV regularization scheme, too.

In this paper, we focus on a method originally proposed by Durand and Froment in [13]. The main idea is to fix the significant coefficients after a classical wavelet thresholding and to change the small coefficients in an iterative minimization process such that the TV norm of the image becomes minimal. This method is really effective to reduce the pseudo-Gibbs oscillating artefacts in the high pass part. In our previous work, we successfully applied the TV minimization/diffusion in a denoising framework for complex wavelets [21], ridgelets [25], wave atoms [23], and curvelets [22, 24, 29]. The hybrid methods show good performance for real applications on denoising of engineering surfaces and images with textures.

In this work, we extend the TV minimization post-processing scheme from Durand and Froment [13] with some essential modifications. Firstly, we propose the usage of the anisotropic total variation instead of the traditional isotropic one. The anisotropy ensures smoothing in flat regions while edges are preserved. The ATV is not new for image processing, but in this work it is the first time to be used to reduce the pseudo-Gibbs artefacts resulted from wavelet shrinkage.

The second innovation is the combination of the post-processing TV minimization step with the tetrolet transform, which is a recently proposed adaptive geometric wavelet filter bank [20]. This transform was especially designed for sparse representation. Therefore we thirdly apply our hybrid method to image approximation while the minimization approach by Durand and Froment was originally proposed in the image denoising framework.

Due to their small support tetrolets do not essentially suffer from oscillating Gibbs artefacts but rather from their non-smoothness and the blocking artefacts (what we call pseudo-Gibbs artefacts). We shall see that the ATV minimization is also able to overcome this drawback while preserving the anisotropic edges and features, by smoothing the reconstructed image changing the small high pass coefficients in a suitable way.

The paper is organized as follows. In Section 2 we recall the rough idea of the tetrolet trans-
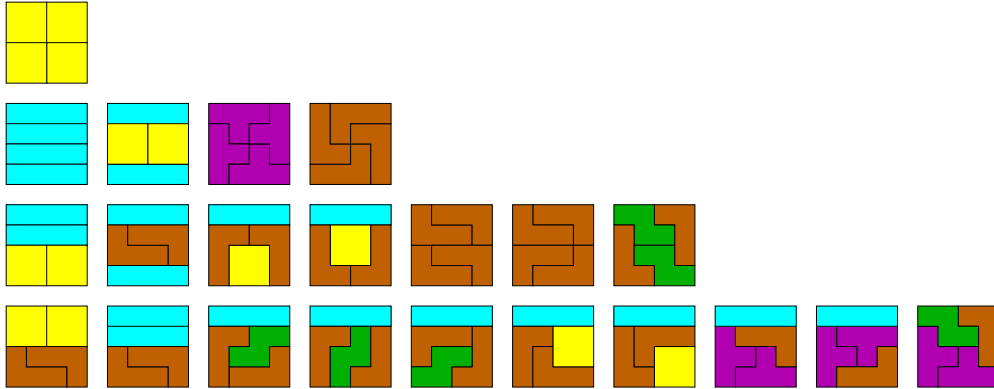
Figure 1: The five free tetrominoes.

Figure 2: The 22 fundamental forms tiling a 4×4 board disregarding rotations and reflections.

form on which the hybrid method is based. Then, in the next section a description of the TV minimization post-processing step is given. In Section 4 we extend this algorithm by proposing an ATV functional which leads to a new powerful algorithm. Finally, in the last section we present some numerical results applying the proposed algorithm to image approximation.

2. The Adaptive Tetrolet Transform

In order to apply our hybrid method to image compression we combine the ATV minimization post-processing step with a suitable wavelet transform. We choose the tetrolet transform recently introduced by one of the authors in [20]. It is a non-redundant adaptive geometric wavelet transform (that means it is adapted to the local image geometry) especially designed for sparse representation.

We shortly recall the tetrolet transform. The underlying idea is simple but enormously effective. The construction is similar to the idea of digital wedgelets where Haar functions on wedge partitions are considered [11, 17].

In the two-dimensional classical Haar case, the low-pass filter and the high-pass filters are just given by the averaging sum and the averaging differences of each four pixel values which are arranged in a 2×2 square. Obviously, the fixed blocking by the dyadic squares is very inefficient because the local structures of an image are disregarded. Our idea is, to allow more general partitions such that the local image geometry is taken into account. Namely, we use tetromino partitions. Tetrominoes became popular through the famous computer game classic 'Tetris', they are shapes made by connecting four equal-sized squares, each joined together with at least one other square along an edge [18]. Disregarding rotations and reflections there are five different shapes, the so called *free tetrominoes* (see Figure 1). It is easy to find out, that there are 117 solutions

for disjoint covering of a 4×4 board with four tetrominoes (see Figure 2). Thus, we divide the given image into blocks of size 4×4 and consider the 117 admissible tetromino partitions there. Among them, we compute in each block the optimal tetromino partition which is adapted to the image geometry in this block. Hence, we define on these geometric shapes Haar-type wavelets, called *tetrolets*, which form a local orthonormal basis.

In the following we describe the computation of the tetrolet decomposition in more detail.

Let $I = \{(m_1, m_2) : m_1, m_2 = 0, \ldots, N-1\} \subset \mathbb{Z}^2$ be the index set of a digital image $\mathbf{a} = (a[m_1, m_2])_{(m_1, m_2) \in I}$ with $N = 2^J$, $J \in \mathbb{N}$. We determine a *4-neighborhood* of an index $(m_1, m_2) \in I$ by $N_4(m_1, m_2) := \{(m_1-1, m_2), (m_1+1, m_2), (m_1, m_2-1), (m_1, m_2+1)\}$. An index that lies at the boundary has three neighbors, an index at the vertex of the image has two neighbors.

We consider disjoint partitions of the index set I with subsets I_ν that satisfy two conditions:

1. each subset I_ν contains four indices, i.e., $\#I_\nu = 4$, and
2. every index of I_ν has a neighbor in I_ν, i.e. $\forall (m_1, m_2) \in I_\nu \ \exists (m_1', m_2') \in I_\nu : (m_1', m_2') \in N_4(m_1, m_2)$.

We call such subsets I_ν *tetromino*. Now, we restrict ourselves to a 4×4 block $Q \subset I$ and consider the tetrolet representation locally. Let the image block Q be disjointly covered by four tetrominoes $\{I_0, I_1, I_2, I_3\}$ and L should be a bijective mapping which maps the four pixel pairs (m_1, m_2) in I_ν to the scalar set $\{0, 1, 2, 3\}$, i.e., it brings the pixels into a unique order. Then, for $\nu = 0, 1, 2, 3$ the discrete basis functions are defined by

$$
\phi_{I_\nu}[m_1, m_2] := \begin{cases} 1/2, & (m_1, m_2) \in I_\nu, \\ 0, & \text{else,} \end{cases}
$$

and

$$
\psi_{I_\nu}^l[m_1, m_2] := \begin{cases} \epsilon[l, L(m_1, m_2)], & (m_1, m_2) \in I_\nu, \\ 0, & \text{else,} \end{cases}
$$

for $l = 1, 2, 3$. Due to their tetromino support the $\psi_{I_\nu}^l$ are called *tetrolets* and ϕ_{I_ν} is the corresponding scaling function. The function values $\epsilon[l, L(m_1, m_2)]$ in the tetrolet definition come from the Haar wavelet transform matrix

$$
W := (\epsilon[l, m])_{l, m=0}^3 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.
$$

With these functions we can decompose the image block $\mathbf{a}|_Q = (a[m_1, m_2])_{(m_1, m_2) \in Q}$ into a detail part and an approximation part

$$
a[m_1, m_2] = \sum_{l=1}^3 \sum_{\nu=0}^3 d_\nu^l[m_1, m_2] \, \psi_{I_\nu}^l[m_1, m_2] + \sum_{\nu=0}^3 c_\nu[m_1, m_2] \, \phi_{I_\nu}[m_1, m_2],
$$

4

if $\{I_0, I_1, I_2, I_3\}$ is a tetromino tiling of the 4×4 block $Q \subset I$. The approximation and tetrolet coefficients are

$$c_\nu[m_1, m_2] = \langle a[\cdot, \cdot], \phi_{I_\nu} \rangle = \sum_{(m', n') \in I_\nu} \frac{1}{2} a[m', n'],$$

$$d_\nu^l[m_1, m_2] = \langle a[\cdot, \cdot], \psi_{I_\nu}^l \rangle = \sum_{(m', n') \in I_\nu} \epsilon[l, L(m', n')] \, a[m', n'],$$

if $(m_1, m_2) \in I_\nu$. Obviously, the absolute size of the tetrolet coefficients depends on the underlying tetromino support. In order to get a sparse representation we choose among the 117 possibilities to cover the block Q with tetrominoes the covering where the tetrolet coefficients d_ν^l have minimal l_1 norm.

The rough structure of the tetrolet filter bank algorithm is described in Table 1, a following shrinkage procedure leads to a sparse representation of images. See [20] for more details.

Algorithm 1: Adaptive Tetrolet Decomposition
Input: Image $f = (f_p)_{p \in I}$ with index set $I = \{0, \ldots, N-1\}^2$, $N = 2^J, J \in \mathbb{N}$.
1. Divide the image into 4×4 blocks.
2. Find the sparsest tetrolet representation in each block .
3. Rearrange the low- and high-pass coefficients of each block into a 2×2 block.
4. Store the tetrolet coefficients (high-pass part).
5. Apply step 1 to 4 to the low-pass image for further decomposition.
Output: Tetrolet coefficients of the decomposed image.

Table 1: Adaptive tetrolet decomposition algorithm.

3. Project Gradient Descent Scheme

There are a lot of hybrid methods to combine wavelet and variational techniques in image processing (see e.g. [5, 8, 9, 32]). Considering the variational part, we focus on TV minimization which was introduced by Rudin, Osher, and Fatemi (ROF-model) in the framework of image processing [30]. It has turned out that the usage of a TV norm is very convenient for image denoising because the TV regularization technique reduces the noise while edges are preserved, in contrast to the traditional L_2 norm, where discontinuities are smoothed [7].

For a function $f : \Omega \to \mathbb{R}$ with $\Omega \subset \mathbb{R}^2$ and $|\nabla f(x)| \in L_1(\mathbb{R}^2)$ the *isotropic TV functional* is defined by

$$TV(f) = \int_\Omega |\nabla f(x)| \, dx. \tag{3.1}$$

5

The variational formulation of the denoising problem leads to gradient descent schemes. Due to the nondifferentiability of $TV(f)$ one often replaces (3.1) by the regularized functional

$$TV(f) = \int_\Omega \sqrt{|\nabla f(x)|^2 + \beta^2}\, dx, \tag{3.2}$$

with a small parameter $\beta > 0$. Another possibility is to use a subgradient instead of a gradient.

In the context of image compression, Chan and Zhou apply in [8] an iterative TV minimization algorithm in order to remove the pseudo Gibbs phenomenon which occurs with wavelet thresholding. In the following we consider a very similar TV minimization technique proposed by Durand and Froment [13] in the image denoising framework. Their subgradient descent scheme has already been successfully used in combination with complex wavelets [21], complex ridgelets [25] and curvelets [22].

We follow the one dimensional discrete approach in [13] skipping the proofs for a better readability. In doing so we immediately transfer the idea from the denoising framework in the approximation framework. An extension to the two dimensional case is straightforward.

For a signal $f = (f_0, \ldots, f_{N-1}) \in \mathbb{R}^N$ the *discrete total variation of f* is given by

$$TV(f) = \sum_{k=0}^{N-2} |f_{k+1} - f_k|, \tag{3.3}$$

see [26, chapter 2]. We assume that $N = 2^{J'}$ with $J' \in \mathbb{N}$, and using the DWT we decompose the signal f into high pass parts and one low pass part. For this purpose we consider the continuous function f^c in the approximation space V_0 that is linked with the given signal f by

$$f^c = \sum_{k=0}^{N-1} f_k\, \phi_{0,k}, \tag{3.4}$$

where $f_k = \langle f^c, \phi_{0,k} \rangle$ holds (if $\{\phi_{0,k}\}_{k=0}^{N-1}$ forms an orthonormal basis of V_0). Applying J decomposition levels with $0 < J < \log_2(N)$ we have

$$f^c = \sum_{j=-1}^{-J} \sum_{k=0}^{2^j N - 1} d_{j,k}(f^c)\, \psi_{j,k} + \sum_{k=0}^{2^{-J} N - 1} c_{-J,k}(f^c)\, \phi_{-J,k}$$

with the wavelet coefficients $d_{j,k}(f^c) := \langle f^c, \psi_{j,k} \rangle$ and the approximation coefficients $c_{-J,k}(f^c) := \langle f^c, \phi_{-J,k} \rangle$. After a nonlinear wavelet shrinkage the function $\tilde{f}^c$ has the form

$$\tilde{f}^c = \sum_{j=-1}^{-J} \sum_{k=0}^{2^j N - 1} S_\lambda(d_{j,k}(f^c))\, \psi_{j,k} + \sum_{k=0}^{2^{-J} N - 1} c_{-J,k}(f^c)\, \phi_{-J,k} \tag{3.5}$$

with a thresholding operator S_λ. A common choice of S_λ is the global hard thresholding operator

$$S_\lambda(x) = \begin{cases} x, & |x| \geq \lambda, \\ 0, & |x| < \lambda, \end{cases} \tag{3.6}$$

6

with a suitable threshold parameter λ. Finally, the reconstructed function in (3.5) can be displayed in V_0 again by

$$\tilde{f}^c = \sum_{k=0}^{N-1} \tilde{f}_k \, \phi_{0,k}$$

where the coefficients $\tilde{f}_k = \langle \tilde{f}^c, \phi_{0,k} \rangle$ provide the approximated signal $\tilde{f} = (\tilde{f}_0, \ldots, \tilde{f}_{N-1}) \in \mathbb{R}^N$ of f.

Let

$$M_{f^c,\lambda} := \{(j,k) : |d_{j,k}(f^c)| \geq \lambda\}$$

be the index set of the retained coefficients after the shrinkage procedure. Further, let be $X := \mathbb{R}^N$ the set of signals with length N. These signals have finite TV norm. Then we can define a subspace $U = U_{M_{f^c,\lambda}} \subset X$ which contains the signals with finite TV norm which essentially correspond with $\tilde{f}$,

$$U := \{u \in X : d_{j,k}(u^c) = d_{j,k}(f^c) \; \forall(j,k) \in M_{f^c,\lambda}, c_{-J,k}(u^c) = c_{-J,k}(f^c) \; \forall k\}.$$

Here, u^c is the continuous function that corresponds to the discrete signal u in terms of (3.4). The signals $u \in U$ differ from the given $\tilde{f}$ at most in the small coefficients which do not contain important information about f, i.e.

$$U = \{\tilde{f}\} + V,$$

with

$$V := \{v \in X : d_{j,k}(v^c) = 0 \; \forall(j,k) \in M_{f^c,\lambda}, c_{-J,k}(v^c) = 0 \; \forall k\}. \tag{3.7}$$

Now, the idea is to find among these signals $u \in U$ that one with minimal TV norm. Hence, we solve the variational problem

Problem 3.1. *Find $u^* \in U$ such that $TV(u^*) = \min_{u \in U} TV(u)$.*

Due to the convexity of the TV functional, the problem has a solution in the subspace U. This finally brings us to the main theorem that a solution can be computed by a *subgradient descent scheme* with a projection onto the linear space V.

Theorem 3.2. *An approximation of the solution u^* of Problem 3.1 is given by the algorithm*

$$u_{k+1} = u_k - t_k \, P_V(g_{TV}(u_k)), \tag{3.8}$$

where P_V is the orthogonal projection onto V, $g_{TV}(u_k)$ is a subgradient of $TV(u_k)$, and the time step t_k satisfies the conditions

$$t_k > 0, \quad \lim_{k \to \infty} t_k = 0, \quad \sum_{k \in \mathbb{N}} t_k = \infty \tag{3.9}$$

in order to obtain convergence.

The proof of the theorem can be found in [13]. It is based on the basic convergence results for subgradient methods from [31].

7

4. ATV Minimization Post-processing

In this section we extend the project gradient descent scheme in (3.8) by proposing a discrete formulation of the ATV norm $ATV(f)$ for a given signal f. In order to avoid a regularization we handle again with the subgradient $g_{ATV}(f)$. Instead of an ordinary wavelet transform we combine the resulting post-processing method with the efficient tetrolet transform for image approximation.

From (3.8) we obtain the descent scheme

$$f^{(k+1)} = f^{(k)} - t_k P_V(g_{ATV}(f^{(k)})), \qquad (4.1)$$

where the sequence of step size $(t_k)_{k \in \mathbb{N}}$ satisfies the conditions (3.9). Since (3.7) the projection P_V onto the subspace V for a signal $g \in X$ is given by the inverse wavelet transform. Let g^c be the continuous function that corresponds to g in the sense of (3.4) and $M_{g^c, \lambda}$ should be the index set of the significant coefficients. Then, we get with

$$\sum_{(j,k) \notin M_{g^c,\lambda}} d_{j,k}(g^c)\, \psi_{j,k}$$

a function which is reconstructed from the insignificant wavelet coefficients, that means from

$$c_{-J,k}(g^c) := 0, \qquad d_{j,k}(g^c) = \begin{cases} 0, & (j,k) \in M_{g^c,\lambda}, \\ d_{j,k}(g^c), & \text{else.} \end{cases} \qquad (4.2)$$

If we represent this function with the basis function $\phi_{0,k}$ in the start space V_0, the coefficients provide the entries of the discrete projection $P_V(g)$.

It remains to calculate the gradient $g_{ATV}(f)$ in (4.1). The original (isotropic) TV functional appearing in the ROF-model can be generalized in a natural way by using anisotropic energies [15]. This leads to a significant improvement since the smoothing procedure stops at edges.

Generalizing the discrete one-dimensional TV in (3.3), for two dimensional data $f := (f_p)_{p \in I}$ we define the *discrete isotropic TV* for a symmetric neighborhood $N(p)$ of the pixel p by

$$TV(f) = \sum_{p \in I} \sum_{p' \in N(p)} |f_{p'} - f_p|. \qquad (4.3)$$

Usual neighborhoods are the 4-star-neighborhood $N_4(p) = \{p' \in I : \|p - p'\|_2^2 = 1\}$ or the 8-neighborhood $N_8(p) = \{p' \in I : \|p - p'\|_2^2 \leq 2\}$. The *discrete ATV functional* is the weighted sum [33]

$$ATV(f) = \sum_{p \in I} \sum_{p' \in N(p)} \sqrt{w_{p,p'}}|f_{p'} - f_p|, \qquad (4.4)$$

where the weights $w_{p,p'} \geq 0$ cause the anisotropic smoothing taking the local image geometry into account. That means, $w_{p,p'}$ depends on the intensity difference $|f_p - f_{p'}|$. More precisely, the

smoothing process across discontinuities (i.e. for large $|f_p - f_{p'}|$) should be stopped while smoothing in flat regions (i.e. for small $|f_p - f_{p'}|$) should be encouraged. Therefore we require that $w_{p,p'}$ is monotone decreasing for $f_p - f_{p'} \geq 0$, and $w_{p,p'} \to 0$ for $|f_p - f_{p'}| \to \infty$. Secondly, $w_{p,p'}$ is symmetric, i.e. $w_{p,p'} = w_{p',p}$ for all $p, p' \in I$. Thirdly, we normalize the weights setting $w_{p,p} = 1$.

There are many possibilities to choose the weights in such fashion, see [10]. A popular choice are the weights from the bilateral filter [34] which contain of an intensity and a spatial component

$$w_{p,p'} = w_{p,p'}(f_p - f_{p'}, p - p') = \exp\left(-\frac{1}{\sigma_i^2}(f_p - f_{p'})^2\right) \cdot \exp\left(-\frac{1}{\sigma_s^2}\|p - p'\|_2^2\right).$$

The first term is the *intensity weight* which punishes pixel values $f_{p'}$ from another image region (and thus ensures edge preserving smoothing), the second term is the *spatial weight* which punishes pixels p' that are far away from p in the spatial domain. The parameters σ_i and σ_s are control parameters.

Then, we compute the subgradient at the point $p \in I$ applying the chain rule

$$
\begin{aligned}
g_{ATV}(f)|_p &= \frac{\partial ATV(f)}{\partial f_p} \\
&= \sum_{p' \in N(p)} \sqrt{w_{p,p'}} \operatorname{sgn}(f_{p'} - f_p)(-1) + \frac{\partial}{\partial f_p}\left(\sum_{p' \in N(p)\backslash p} \sum_{p'' \in N(p')} \sqrt{w_{p',p''}}|f_{p''} - f_{p'}|\right) \\
&= \sum_{p' \in N(p)} \sqrt{w_{p,p'}} \operatorname{sgn}(f_p - f_{p'}) + \sum_{p' \in N(p)\backslash p}\left(\sum_{p'' \in N(p')} \sqrt{w_{p',p''}} \frac{\partial}{\partial f_p}|f_{p''} - f_{p'}|\right).
\end{aligned}
$$

The expression $\frac{\partial}{\partial f_p}|f_{p''} - f_{p'}|$ in the last term vanishes for all $p'' \neq p$, i.e.

$$
\begin{aligned}
g_{ATV}(f)|_p &= \sum_{p' \in N(p)} \sqrt{w_{p,p'}} \operatorname{sgn}(f_p - f_{p'}) + \sum_{p' \in N(p)\backslash p} \sqrt{w_{p',p}} \operatorname{sgn}(f_p - f_{p'}) \\
&= 2 \sum_{p' \in N(p)} \sqrt{w_{p,p'}} \operatorname{sgn}(f_p - f_{p'}), \tag{4.5}
\end{aligned}
$$

with symmetric weights $w_{p',p} = w_{p,p'}$. In order to apply Theorem 3.2 we have finally to prove that the derived formula gives a subgradient of the ATV norm. Indeed, this is the case:

Lemma 4.1. *The expression in (4.5) is a subgradient of ATV, i.e., it holds*

$$ATV(h) \geq ATV(f) + \langle g_{ATV}(f), h - f\rangle \quad \forall h \in X.$$

Proof. It holds $|y| \geq |x| + \operatorname{sgn}(x)(y - x), \forall x, y \in \mathbb{R}$. Let $y := \sqrt{w_{p,p'}}(h_{p'} - h_p)$ and $x := \sqrt{w_{p,p'}}(f_{p'} - f_p)$. Then we get

$$\sqrt{w_{p,p'}}|h_{p'} - h_p| \geq \sqrt{w_{p,p'}}|f_{p'} - f_p| + \operatorname{sgn}(f_{p'} - f_p) \cdot \sqrt{w_{p,p'}}(h_{p'} - h_p - f_{p'} + f_p).$$

9

Now we sum over $p \in I$ and $p' \in N(p)$ and obtain

$$
\begin{aligned}
ATV(h) &\geq ATV(f) + \sum_{p \in I} \sum_{p' \in N(p)} \text{sgn}(f_{p'} - f_p) \cdot \sqrt{w_{p,p'}}((h_{p'} - f_{p'}) - (h_p - f_p)) \\
&= ATV(f) + \sum_{p \in I} 2 \sum_{p' \in N(p)} \sqrt{w_{p,p'}} \, \text{sgn}(f_{p'} - f_p) \cdot (h_p - f_p) \\
&= ATV(f) + \sum_{p \in I} g_{ATV}(f)|_p \cdot (h - f)|_p \\
&= ATV(f) + \langle g_{ATV}(f), h - f \rangle.
\end{aligned}
$$

$\square$

Summing up, we have the following algorithm.

Algorithm 2: ATV Minimization of Tetrolet Coefficients

Input: Image $f = (f_p)_{p \in I}$ with index set $I = \{0, \ldots, N-1\}^2$, $N = 2^J$, $J \in \mathbb{N}$.

1. Compute the discrete tetrolet coefficients.
2. Apply thresholding and record indices of retained coefficients in $M_{f^c, \lambda}$.
3. Reconstruct $\tilde{f}$ from the retained tetrolet coefficients by inverse tetrolet transform.
4. Set $f^{(0)} := \tilde{f}$. For $k = 0, 1, \ldots$ minimize the ATV norm of $f^{(k)}$ by the following iterative scheme
   (i) Compute the subgradient $g_{ATV}(f^{(k)})$ of $f^{(k)}$ by (4.5).
   (ii) Decompose $g_{ATV}(f^{(k)})$ into the tetrolet coefficients $d_{j,k}(g_{ATV}(f^{(k)}))$ and $c_{-J,k}(g_{ATV}(f^{(k)}))$.
   (iii) Compute $P_V(g_{ATV}(f^{(k)}))$ by the inverse tetrolet transform from the coefficients according to (4.2).
   (iv) Set $f^{(k+1)} = f^{(k)} - t^{(k)} P_V(g_{ATV}(f^{(k)}))$.

Output: Image $f^{(k)}$.

Table 2: ATV minimization of tetrolet coefficients.

Finally, we mention that there exist other possibilities to define the discrete TV functional. Our approach in (4.3) was a discretization of the continuous TV functional in (3.1). A discretization of the *regularized* TV functional in (3.2) leads to the anisotropic discrete TV [33]

$$
ATV(f) = \sum_{p \in I} \sqrt{\sum_{p' \in N(p)} w_{p,p'}|f_{p'} - f_p|^2 + \beta^2}. \tag{4.6}
$$

In this case, instead of a subgradient $g_{ATV}(f)$ we get a more complex gradient

$$
\nabla_f ATV(f)|_p = \sum_{p' \in N(p)} w_{p',p} \left(Z(p, p') + Z(p', p'') \right) (f_p - f_{p'}) \tag{4.7}
$$

10

with $Z(p_1, p_2) = \left(\sum_{p_2 \in N(p_1)} w_{p_1,p_2}(f_{p_1} - f_{p_2})^2 + \beta^2 \right)^{-1/2}$ for $p_1, p_2 \in I$. We obtain this gradient formula at the point $p \in I$ applying the chain rule

$$\nabla_f ATV(f)|_p = \frac{\partial ATV(f)}{\partial f_p}$$

$$= \frac{\partial}{\partial f_p} \sqrt{\sum_{p' \in N(p)} w_{p,p'}(f_p - f_{p'})^2 + \beta^2}$$

$$+ \frac{\partial}{\partial f_p} \left(\sum_{p' \in N(p)\backslash p} \sqrt{\sum_{p'' \in N(p')} w_{p',p''}(f_{p'} - f_{p''})^2 + \beta^2} \right)$$

$$= Z(p, p') \left(\sum_{p' \in N(p)} w_{p,p'}(f_p - f_{p'}) \right) + \sum_{p' \in N(p)\backslash p} Z(p', p'') \sum_{p'' \in N(p')} w_{p',p''} \frac{\partial}{\partial f_p}(f_{p'} - f_{p''})^2.$$

The expression $\frac{\partial}{\partial f_p}(f_{p'} - f_{p''})^2$ in the last term vanishes for all $p'' \neq p$, i.e.

$$\nabla_f ATV(f)|_p = Z(p, p') \left(\sum_{p' \in N(p)} w_{p,p'}(f_p - f_{p'}) \right) + \sum_{p' \in N(p)\backslash p} Z(p', p'') w_{p',p}(f_{p'} - f_p)(-1)$$

$$= \sum_{p' \in N(p)} w_{p',p} \left(Z(p, p') + Z(p', p'') \right) (f_p - f_{p'}).$$

with symmetric weights $w_{p',p} = w_{p,p'}$.

5. Application to Image Approximation

Originally, the TV minimization post-processing in [13] was proposed for noise removal, but image approximation is also based on a wavelet shrinkage which leads to pseudo-Gibbs artefacts in the high pass part. Therefore we apply the previously derived hybrid method to image approximation. In order to do this we combine the ATV minimization post-processing step with an appropriate underlying wavelet transform, namely the tetrolet transform. In contrast to the curvelet [3, 16] or contourlet transform [12] the tetrolet transform is non-redundant and adaptive and thus particular suitable for sparse image representation.

The tetrolet transform does not essentially suffer from oscillating Gibbs artefacts due to the very small support of the tetrolets. A major problem are the blocking artefacts and the non-smoothness of the basis functions with approximation of images with higher regularity than piecewise constant. The proposed ATV minimization is able to overcome this drawback, too. While we fix the essential tetrolet coefficients we change the small ones by the iterative TV minimization process in order to smooth the image regions and preserve the edges.

5.1. Perfect reconstruction property for linear signals

In this subsection we illustrate that the proposed post-processing scheme increases the regularity of the approximated signal which yields even to perfect reconstruction under several assumptions. It is a first attempt to understand the theoretical properties of the combined tetrolet shrinkage with ATV minimization.

The underlying tetrolet transform is a generalized adaptive Haar wavelet transform. Hence, after a shrinkage procedure one obtains a piecewise constant approximation of the given image. Restricting ourselves to the one-dimensional case (and thus to classical Haar wavelet transform), we consider a linear input signal $f = (f[k])_{k=0}^{N-1}$ with $f[k] = a\,k + b$ and its piecewise constant Haar wavelet approximation $\tilde{f}$. In the following we analyse whether we are able to improve $\tilde{f}$ by the ATV minimization post-processing step such that we achieve perfect reconstruction of f.

Theorem 5.1. *Let $f = (f[k])_{k=0}^{N-1}$ be a linear signal with $f[k] = a\,k + b$. If we apply Algorithm 2 to f with a global shrinkage parameter $2^{-1/2}a < \lambda < 2a$, after only one single post-processing TV minimization step with $t^{(0)} = a/2$ we obtain perfect reconstruction of f using the neighborhood $N(k) = \{k-1, k+1\}$ and isotropic weights $w_{k,k'} = 1$, $k, k' \in \{0, \ldots, N-1\}$.*

Proof. We prove the perfect reconstruction property by a straightforward application of Algorithm 2.

Step 1. We start in the space V_0 with the continuous function f^c which corresponds to the discrete signal f

$$f^c = \sum_{k=0}^{N-1} c_{0,k}(f^c)\, \phi_{0,k}$$

with $c_{0,k}(f^c) = \langle f^c, \phi_{0,k} \rangle = f[k]$. The discrete Haar wavelet transform leads in the j-th decomposition level, $j = -1, \ldots, -J$, for $k = 0, \ldots, 2^j N - 1$ to

$$c_{j,k}(f^c) = \frac{1}{\sqrt{2}}(c_{j+1,2k+1}(f^c) + c_{j+1,2k}(f^c)),$$

$$d_{j,k}(f^c) = \frac{1}{\sqrt{2}}(c_{j+1,2k+1}(f^c) - c_{j+1,2k}(f^c)).$$

Particularly, in the first level the coefficients have the form for $k = 0, \ldots, \frac{N}{2} - 1$

$$c_{-1,k}(f^c) = \frac{1}{\sqrt{2}}(c_{0,2k+1}(f^c) + c_{0,2k}(f^c)) = \frac{1}{\sqrt{2}}(f[2k+1] + f[2k]) = \frac{1}{\sqrt{2}}(a(4k+1) + 2b),$$

$$d_{-1,k}(f^c) = \frac{1}{\sqrt{2}}(c_{0,2k+1}(f^c) - c_{0,2k}(f^c)) = \frac{1}{\sqrt{2}}(f[2k+1] - f[2k]) = \frac{1}{\sqrt{2}}a.$$

Step 2. If we choose the threshold λ such that $2^{-1/2}a < \lambda < 2a$, the wavelet coefficients $d_{-1,k}(f^c)$ in the first level are set to zero. All other coefficients $c_{-1,k}(f^c)$ are retained.

Step 3. The reconstruction leads to the Haar wavelet approximation $\tilde{f}^c$, whose coefficients in the space V_0 provide the discrete approximation $\tilde{f} = (\tilde{f}[k])_{k=0}^{N-1}$ of the input signal f.
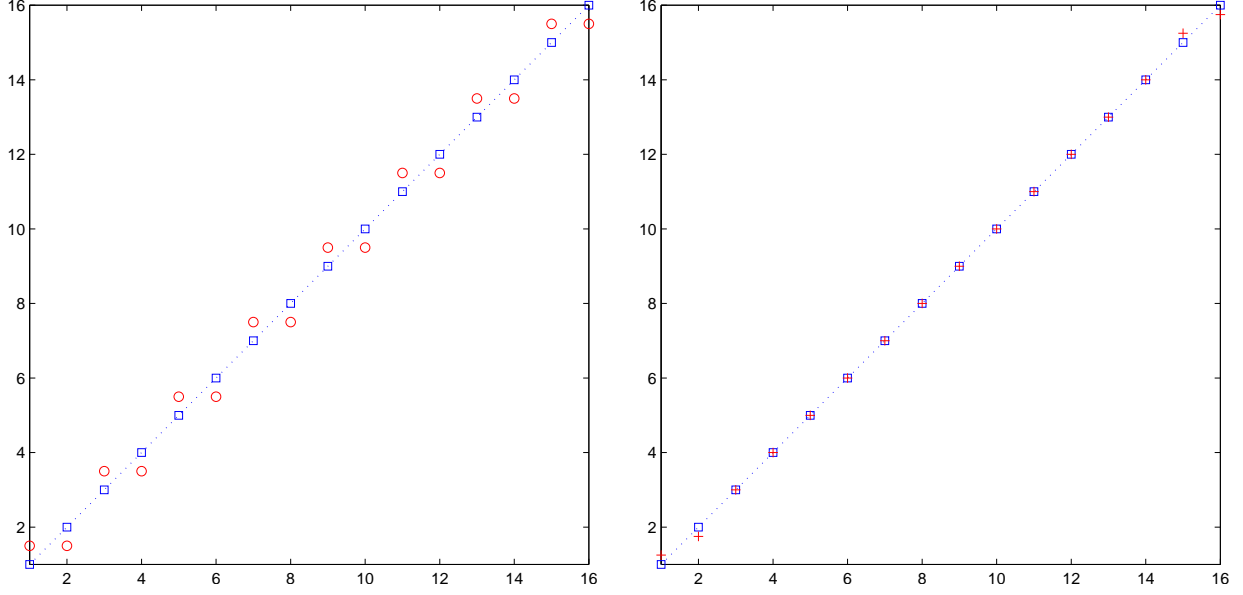
12

Figure 3: Approximation of a linear signal with post-processing. (a) Input (□) and Haar wavelet approximation (o), (b) Input (□) and smoothed approximation (+) after one single TV minimization step.

The reconstruction formula for the coefficients are

$$c_{0,2k+1}(f^c) = \frac{1}{\sqrt{2}}(c_{-1,k}(f^c) + d_{-1,k}(f^c)),$$

$$c_{0,2k}(f^c) = \frac{1}{\sqrt{2}}(c_{-1,k}(f^c) - d_{-1,k}(f^c)), \quad k = 0, \dots, \frac{N}{2} - 1.$$

Since the wavelet coefficients $d_{-1,k}(f^c)$ were cut in the shrinkage procedure, the terms are simplified such that we have for $k = 0, \dots, \frac{N}{2} - 1$

$$\tilde{c}_{0,2k+1}(f^c) = \tilde{c}_{0,2k}(f^c) = \frac{1}{\sqrt{2}}c_{-1,k}(f^c) = \frac{1}{2}(a(4k + 1) + 2b).$$

This means for $k = 0, \dots, N - 1$

$$\tilde{c}_{0,k}(f^c) = \begin{cases} \frac{1}{2}(a(2k + 1) + 2b), & k \text{ even} \\ \frac{1}{2}(a(2k - 1) + 2b), & k \text{ odd} \end{cases} = \begin{cases} \frac{2k+1}{2}a + b, & k \text{ even} \\ \frac{2k-1}{2}a + b, & k \text{ odd} \end{cases}. \tag{5.1}$$

Then we get the continuous approximation $\tilde{f}^c$ of f^c

$$\tilde{f}^c = \sum_{k=0}^{N-1} \tilde{c}_{0,k}(f^c)\, \phi_{0,k},$$

which leads to the discrete approximation $\tilde{f} = (\tilde{f}[k])_{k=0}^{N-1}$ of f, because it is $\tilde{f}[k] = \tilde{c}_{0,k}(f^c)$. See an illustration in Figure 3(a).

13

Step 4. Set $f^{(0)} := \tilde{f}$. We minimize the TV norm of $f^{(0)}$ with only one iteration step of the post-processing step. Note that it holds

$$\operatorname{sgn}(f^{(0)}[k] - f^{(0)}[k+1]) = \begin{cases} 0, & k \text{ even} \\ -1, & k \text{ odd} \end{cases}. \tag{5.2}$$

(i) The subgradient $g_{ATV}(f^{(0)})$ at the point k is given in (4.5) as

$$g_{ATV}(f^{(0)})[k] = 2\left(\sqrt{w_{k,k-1}}\operatorname{sgn}(f^{(0)}[k] - f^{(0)}[k-1]) + \sqrt{w_{k,k+1}}\operatorname{sgn}(f^{(0)}[k] - f^{(0)}[k+1])\right).$$

Because the values $f[k] = a\,k + b$ in $f = (f[k])_{k=0}^{N-1}$ are sampling points from a linear function without any jump, it is reasonable to choose in this case isotropic weights $w_{k,k+1} = 1$ for all k. Thus, the subgradient reduces to

$$g_{ATV}(f^{(0)})[k] = 2\left(\operatorname{sgn}(f^{(0)}[k] - f^{(0)}[k-1]) + \operatorname{sgn}(f^{(0)}[k] - f^{(0)}[k+1])\right).$$

(ii) Decomposition of the subgradient into Haar wavelet coefficients leads to

$$\begin{aligned} d_{-1,k}(g^c_{ATV}(f^{(0)})) &= \frac{1}{\sqrt{2}}(g_{ATV}(f^{(0)})[2k+1] - g_{ATV}(f^{(0)})[2k]) \\ &= \sqrt{2}\operatorname{sgn}(f^{(0)}[2k+1] - f^{(0)}[2k+2]) - \sqrt{2}\operatorname{sgn}(f^{(0)}[2k] - f^{(0)}[2k+1]). \end{aligned} \tag{5.3}$$

(iii) The inverse Haar wavelet transform from the insignificant coefficients gives

$$\sum_{k=0}^{N/2-1} d_{-1,k}(g^c_{ATV}(f^{(0)}))\,\psi_{-1,k}.$$

This is the corresponding continuous function to the discrete projection $P_V(g_{ATV}(f^{(0)}))$ with $V = \{v \in X : d_{j,k}(v^c) = 0\ \forall j < -1, c_{-1,k}(v^c) = 0\ \forall k\}$. A pointwise consideration for $k = 0, \ldots, N-1$ is

$$P_V(g_{ATV}(f^{(0)}))|_k = \begin{cases} -\frac{1}{\sqrt{2}}d_{-1,\frac{k}{2}}(g^c_{ATV}(f^{(0)})), & k \text{ even} \\ \frac{1}{\sqrt{2}}d_{-1,\frac{k-1}{2}}(g^c_{ATV}(f^{(0)})), & k \text{ odd} \end{cases}.$$

With (5.3) and the property (5.2) of $f^{(0)}$ we get

$$\begin{aligned} P_V(g_{ATV}(f^{(0)}))|_k &= \begin{cases} -\operatorname{sgn}(f^{(0)}[k+1] - f^{(0)}[k+2]) + \operatorname{sgn}(f^{(0)}[k] - f^{(0)}[k+1]), & k \text{ even} \\ \operatorname{sgn}(f^{(0)}[k] - f^{(0)}[k+1]) - \operatorname{sgn}(f^{(0)}[k-1] - f^{(0)}[k]), & k \text{ odd} \end{cases} \\ &= \begin{cases} 1, & k \text{ even} \\ -1, & k \text{ odd} \end{cases}. \end{aligned}$$

(iv) Considering the single iteration step

$$f^{(1)} = f^{(0)} - t^{(0)}P_V(g_{ATV}(f^{(0)}))$$

14

at the point $k = 0, \ldots, N - 1$, we get with (5.1) and the step size $t^{(0)} = a/2$

$$f^{(1)}[k] = f^{(0)}[k] - t^{(0)} \begin{cases} 1, & k \text{ even} \\ -1, & k \text{ odd} \end{cases} = \begin{cases} \frac{2k+1}{2}a + b - a/2, & k \text{ even} \\ \frac{2k-1}{2}a + b + a/2, & k \text{ odd} \end{cases} = ak + b,$$

which means perfect reconstruction. □

If we consider signals with jumps of high amplitude, the choice of anisotropic weights is essential in order to suppress the smoothing at jumps.

5.2. Numerical results

In Figure 3 the perfect reconstruction property is illustrated for a linear signal $f = (f[k])_{k=0}^{15}$ with $f[k] = k$. The Haar wavelet approximation $f^{(0)}$ in 3(a) can be smoothed by applying one single TV minimization step of the proposed post-processing scheme according to Theorem 5.1. The resulting approximation $f^{(1)}$ provides perfect reconstruction apart from the boundaries, see 3(b).

Figure 4 shows the input images that are tested for our method. Figure 5 and Figure 6 present the smoothing effect of the ATV minimization combined with the tetrolet approximation. We compare our results with the biorthogonal 9-7 filter. For the clock image in Figure 5 the PSNR value only slightly increases after 5 iterations but the visual quality of the image is considerably improved (notice the shadow of the book). For the piecewise smooth image in Figure 6 we notice after 10 iterations the strong smoothing which essentially piecewise constant tetrolet approximation. Here, the results for the tetrolet transform and for the 9-7 filter are in the same scale. Note, that the adaptive tetrolets already have anisotropic properties in themselves while the tensor wavelets in the 9-7 filter do not have any directional sensitivity. Therefore, the application of the anisotropic post-processing technique leads to a greater PSNR gain with the underlying 9-7 filter, namely the PSNR increases from 36.07 dB to 38.24 dB, than with the tetrolet transform where the PSNR only increases from 36.59 dB to 38.06 dB.

In both figures we have chosen as computational parameters a step size $t_k = \frac{1}{k+1}$ (according to (3.9)) and the weight control parameters $\sigma_s = 2, \sigma_i = 100$ with an 8-neighborhood. Furthermore, we have taken the regularized discrete TV functional in Ea. (4.6).

6. Conclusions

In this paper, an effective hybrid method is proposed for image compression by combining the ATV minimization with tetrolet shrinkage. This algorithm can be easily extended for any other wavelet transforms.

We used classic gradient method for computing the minimization of regularized TV norm in our algorithm. Some advanced algorithms, e.g., first-order methods by Nesterov [27] and Aujol [1], and fast algorithm by Wang et al. [35] can be incorporated into our framework to improve the
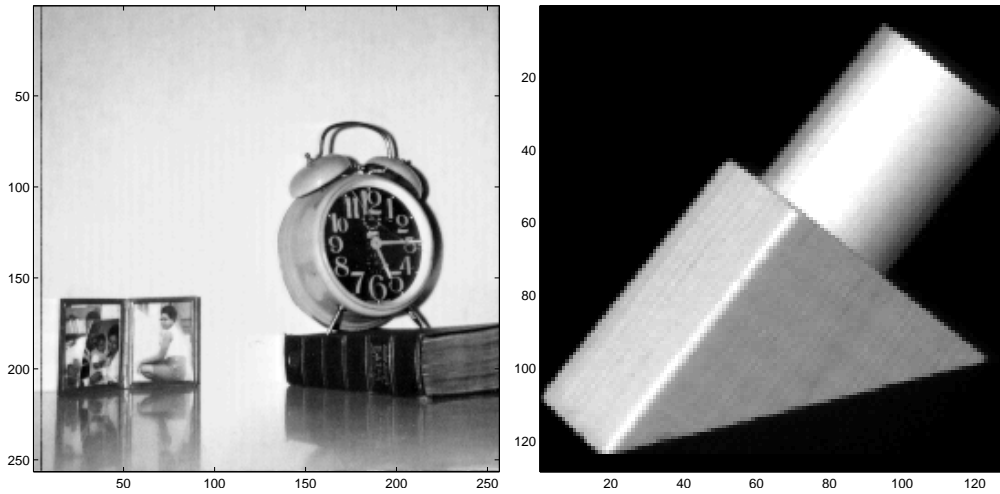
Figure 4: Input images.

performance of the TV step. In a next work, we intend to combine the ATV minimization with an underlying curvelet transform for image denoising with texture images. Furthermore, we will present some theoretical results to measure the approximation improvement using the proposed post-processing algorithm.

References

[1] J. Aujol, Some first-order algorithms for total variation based image restoration, *J. Math. Imaging Vis.*, **34**, 307–327 (2009).

[2] E. Candès, D. Donoho, Curvelets - a surprisingly effective nonadaptive representation for objects with edges, In: A. Cohen, C. Rabut, L. Schumaker (Eds.), *Curves and Surface Fitting: Saint-Malo 1999*, Vanderbilt Univ. Press, Nashville, 2000, 105–120.

[3] E. Candès, D. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, *Comm. Pure and Appl. Math.*, **56**, 216–266 (2004).

[4] E. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, *Multiscale Model. Simul.*, **5** (3), 861–899 (2006).

[5] E. Candès, F. Guo, New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction, *Signal Process.*, **82** (11), 1519–1543 (2002).
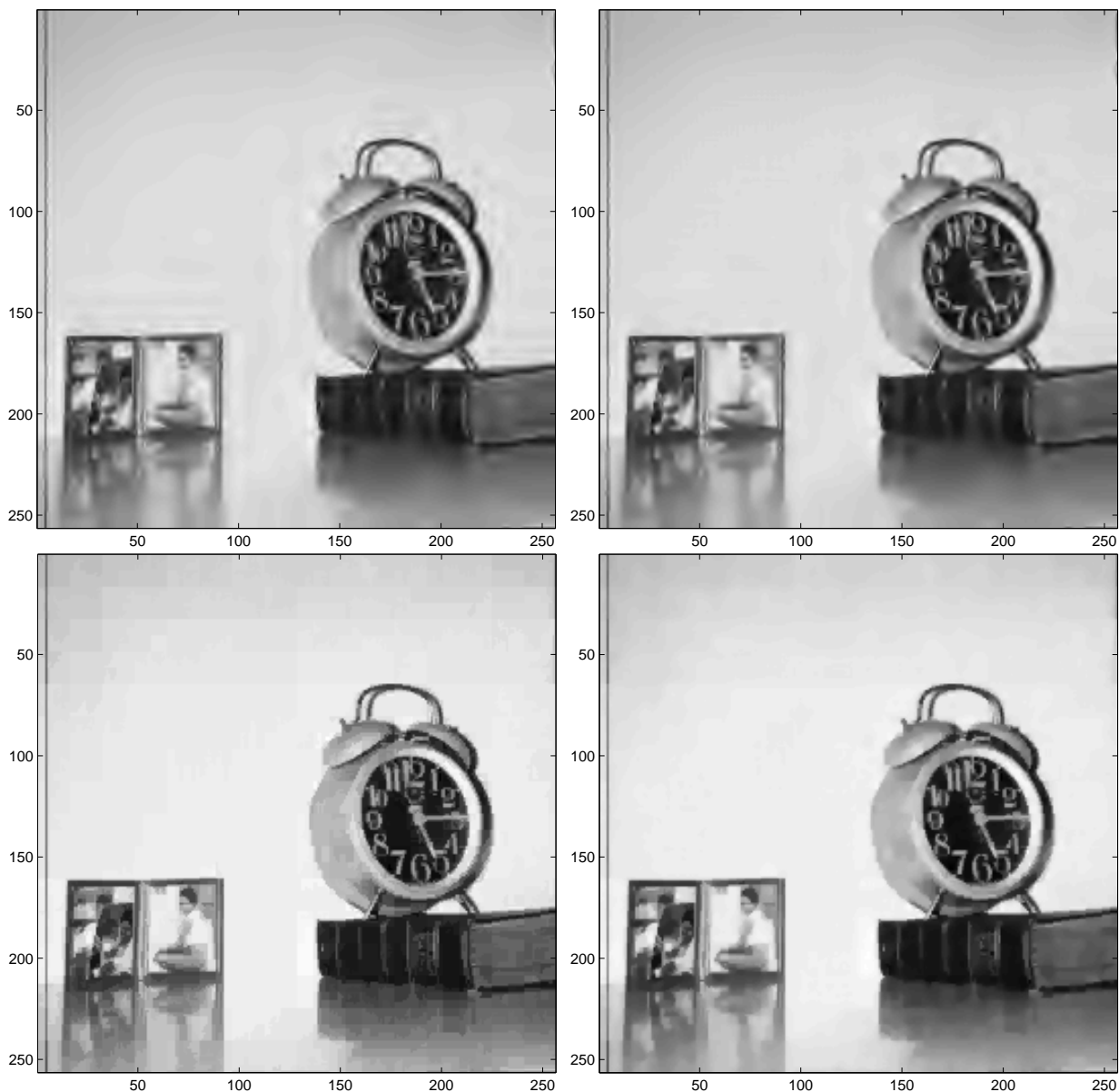
16

Figure 5: Approximation for the clock image with 2048 coefficients. (a) Biorthogonal 9-7 shrinkage, PSNR=30.07 DB. (b) Biorthogonal 9-7 shrinkage with ATV minimization, PSNR=30.62 DB. (c) Tetrolet shrinkage, PSNR=32.61 DB. (d) Tetrolet shrinkage with ATV minimization, PSNR=33.08 DB.

[6] A. Chambolle, R. De Vore, N. Lee, B. Lucier, Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Trans. Image Process.*, **7** (3), 319–335 (1998).

[7] T. Chan, S. Osher, J. Shen, The digital TV filter and nonlinear denoising, *IEEE Trans. Image Process.*, **10**, 231–241 (2001).

[8] T. Chan, H. Zhou, Total variation improved wavelet thresholding in image compression, in: *Proc. of ICIP' 2000* **16** (5), 1289–1302 (2000).

[9] R. Coifman, A. Sowa, Combining the calculus of variations and wavelets for image enhancement, *Appl. Comput.*
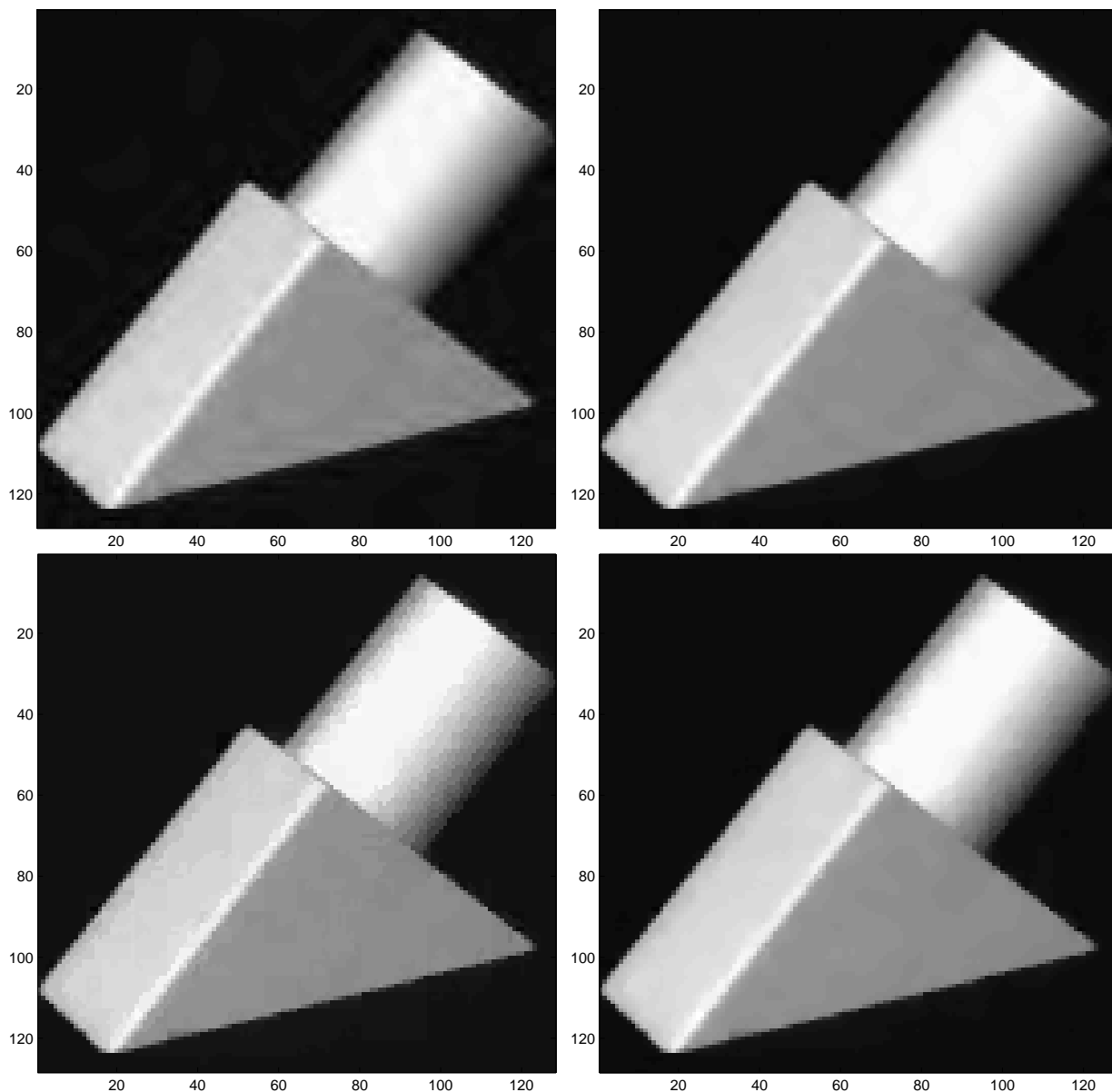
Figure 6: Approximation for a piecewise smooth image with 1024 coefficients. (a) Biorthogonal 9-7 shrinkage, PSNR=36.07 dB. (b) Biorthogonal 9-7 shrinkage with ATV minimization, PSNR=38.24 dB. (c) Tetrolet shrinkage, PSNR=36.59 dB. (d) Tetrolet shrinkage with ATV minimization, PSNR=38.06 dB.

    Harmon. Anal., **9**, 1–18 (2000).

[10] S. Didas, P. Mrázek, J. Weickert, Energy-based image simplification with nonlocal data and smoothness terms, in: Algorithms for Approximation – Proceedings of the 5th International Conference, Chester, 51–60, Springer, Berlin, Heidelberg, New York, 2007.

[11] D. Donoho, Wedgelets: nearly minimax estimation of edges, *Ann. Statist.*, **27** (3), 859–897 (1999).

[12] M. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, *IEEE Trans. Image Process.*, **14** (12), 2091–2106 (2005).

[13] S. Durand, J. Froment, Reconstruction of wavelet coefficients using total variation minimization, *SIAM J. Sci. Comput.*, **24**, 1754–1767 (2003).

[14] G. Easley, D. Labate, F. Colonna, Shearlet based total variation for denoising, *IEEE Trans. Image Process.*, **18** (2), 260–268 (2009).

[15] S. Esedoglu, S. Osher, Decomposition of images by the anisotropic Rudin-Osher-Fatemi model, *Comm. Pure Appl. Math.*, **57**, 1609–1626 (2004).

[16] M. Fadili, J.-L. Starck, Curvelets and ridgelets, in: R. Meyers (Ed.), Encyclopedia of Complexity and Systems Science, Vol 3, 1718–1738, Springer, New York, 2009.

[17] H. Führ, L. Demaret, F. Friedrich, Beyond wavelets: New image representation paradigms. Book chapter, in: M. Barni and F. Bartolini (Eds.), Document and Image Compression, Taylor and Francis, Boca Raton, FL, 2006.

[18] S. Golomb, Polyominoes, Princeton University Press, Princeton, NJ, 1994.

[19] K. Guo, D. Labate, Optimally sparse multidimensional representation using shearlets, *SIAM J. Math. Anal.*, **39**, 298–318 (2007).

[20] J. Krommweh, Tetrolet transform: A new adaptive Haar wavelet algorithm for sparse image representation, J. Vis. Commun. (2010), doi:10.1016/j.jvcir.2010.02.011

[21] J. Ma, Towards artifact-free characterization of surface topography using complex wavelets and total variation minimization, *Appl. Math. Comput.*, **170** (2), 1014–1030 (2005).

[22] J. Ma, Curvelets for surface characterization, *Appl. Phys. Lett.*, **90**, 054109 (2007).

[23] J. Ma, Characterization of textural surfaces using wave atoms, *Appl. Phys. Lett.*, **90**, 264101 (2007).

[24] J. Ma, G. Plonka, Combined curvelet shrinkage and nonlinear anisotropic diffusion, *IEEE Trans. Image Process.*, **16** (9), 2198–2206 (2007).

[25] J. Ma, M. Fenn, Combined complex ridgelet shrinkage and total variation minimization, *SIAM J. Sci. Comput.*, **28** (3), 984–1000 (2006).

[26] S. Mallat, A wavelet tour of signal processing, Academic Press, San Diego, CA, 1998.

[27] Y. Nesterov, Smooth minimization of non-smooth functions, *Math. Program.*, **103** (1), 127–152 (2005).

[28] E. Le Pennec, S. Mallat, Sparse geometric image representations with bandelets, *IEEE Trans. Image Process.*, **14** (4), 423–438 (2005).

[29] G. Plonka, J. Ma, Nonlinear regularized reaction-diffusion filters for denoising of images with textures, *IEEE Trans. Image Process.*, **17** (8), 1283–1294 (2008).

[30] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D*, **60**, 259–268 (1992).

[31] N. Shor, Minimization methods for nondifferentiable functions, Springer Series in Computational Mathematics 3, Springer, Berlin, 1985.

[32] J. Starck, M. Elad, D. Donoho, Image decomposition via the combination of sparse representation and a variational approach, *IEEE Trans. Image Process.*, **14** (10), 1570–1582 (2005).

[33] V. Ta, S. Bougleux, A. Elmoataz, and O. Lezoray, Nonlocal anisotropic discrete regularization for image, data filtering and clustering, Tech. Rep., Université de Caen Basse-Normadie, Caen, France, 2007.

[34] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in *Proc. IEEE Int. Conf. Computer Vision*, New Delhi, India, 1998, pp. 836–846.

[35] Y. Wang, J. Yang, W. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imaging Sciences*, **1** (3), 248–272 (2008).