BENEFITS OF CONSISTENCY IN IMAGE DENOISING WITH STEERABLE WAVELETS

Bugra Tekin, Ulugbek S. Kamilov, Emrah Bostan, and Michael Unser

Biomedical Imaging Group, EPFL, Switzerland

ABSTRACT

The steerable wavelet transform is a redundant image representation with the remarkable property that its basis functions can be adaptively rotated to a desired orientation. This makes the transform well-suited to the design of wavelet-based algorithms applicable to images with a high amount of directional features. However, arbitrary modification of the wavelet-domain coefficients may violate consistency constraints because a legitimate representation must be redundant. In this paper, by honoring the redundancy of the coefficients, we demonstrate that it is possible to improve the performance of regularized least-squares problems in the steerable wavelet domain. We illustrate that our consistent method significantly improves upon the performance of conventional denoising with steerable wavelets.

***Index Terms*—**Image denoising, sparse estimation, wavelet regularization, steerable wavelet transform

1. INTRODUCTION

The steerable wavelet transform is a multiscale image decomposition that was first introduced in [1, 2]. The key property of the transform is that its basis functions can be rotated to any orientation by forming suitable linear combinations of the primary directional wavelet components [3]. This makes the transform practical for the design of wavelet-based algorithms and provide rich directional information [4–9]. The steerability of the transform implies redundancy where the corresponding wavelets form a tight frame. Redundancy is desirable for image analysis and processing. However, for wavelet-domain estimation problems, it implies that not every set of wavelet coefficients can be perfectly inverted back to the original image domain.

In this work, we propose a simple technique for improving the performance of ℓ_1-based image denoising in the steerable-wavelet domain. Our technique, which we call *consistency*, refers to the fact that the solution obtained by the algorithm is constrained to the space spanned by the basis functions of the transform, which results in a certain norm equivalence between image-domain and wavelet-domain estimations. Previously, consistency was used in [10] to boost the performance of cycle spinning [11]. Our results indicate that the performance of consistent image denoising with steerable wavelets systematically improves the performance of conventional ℓ_1-based wavelet-domain denoising.

This work was supported by the European Commission under Grant ERC-2010-AdG 267439-FUN-SP.

2. WAVELET-BASED DENOISING

This paper focuses on the general denoising problem $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where $\mathbf{x} \in \mathbb{R}^n$ is the vector containing the samples of the unknown image of interest and $\mathbf{n} \in \mathbb{R}^n$ is the vector containing zero-mean white Gaussian noise. In wavelet-based denoising, the estimation is achieved in three steps.

- Perform the wavelet transform $\mathbf{u} = \mathbf{W}\mathbf{y} \in \mathbb{R}^m$, where, in general, $m \geq n$. The equality $m = n$ happens when the transform is non-redundant.
- Apply a denoising algorithm represented by $\mathbf{f}_{\boldsymbol{\lambda}} : \mathbb{R}^m \mapsto \mathbb{R}^m$ that associates the coefficients $\mathbf{u}$ to the estimate $\widehat{\mathbf{w}} = \mathbf{f}_{\boldsymbol{\lambda}}(\mathbf{u})$. The vector $\boldsymbol{\lambda}$ represents the set of parameters of the algorithm.
- Return to the image domain by applying the inverse transform to obtain $\widehat{\mathbf{x}} = \mathbf{W}^T\widehat{\mathbf{w}} \in \mathbb{R}^n$.

In the sequel, we restrict our attention to *tight* wavelet frames, which implies that the transform preserves inner products and is self-reversible

$$\forall \mathbf{x} \in \mathbb{R}^n, \qquad \mathbf{W}^T\mathbf{W}\mathbf{x} = \mathbf{x}. \tag{1}$$

This also implies that the transform is norm-preserving, with $\|\mathbf{x}\|_2^2 = \|\mathbf{W}\mathbf{x}\|_2^2$.

A common estimation approach is to solve the regularized least-squares problem

$$\widehat{\mathbf{w}} = \mathbf{f}_\lambda(\mathbf{u}) = \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{w} - \mathbf{u}\|_2^2 + \lambda\Phi(\mathbf{w}) \right\}, \tag{2}$$

where the function Φ is a *regularizer* that promotes solutions with certain desirable properties and $\lambda > 0$ is the *regularization parameter*. The least-squares term in (2) is traditionally justified by the fact that the transform $\mathbf{W}$ is norm-preserving. Nevertheless, as we show in the sequel, (2) is suboptimal for redundant $\mathbf{W}$. The popular wavelet-shrinkage algorithm is obtained by using the non-smooth convex function $\Phi(\cdot) =$

$\|\cdot\|_1$ that favors sparse wavelet-domain solutions and admits the closed-form solution $\widehat{\mathbf{w}} = \eta(\mathbf{u};\lambda)$. The soft-thresholding function η is applied componentwise on the detail wavelet coefficients.

3. STEERABLE WAVELET TRANSFORM

After describing the generic approach for wavelet-domain denoising, we now explain the particular type of *steerable* wavelet transforms. The concept relies on two fundamental elements: the Riesz transform and an isotropic mother wavelet frame.

3.1. Riesz Transform

Consider the function $f \in L_2(\mathbb{R}^d)$ and its Fourier transform $\mathcal{F}\{f\}(\boldsymbol{\omega}) = \hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} f(\mathbf{x}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}, \mathbf{x}\rangle} \mathrm{d}\mathbf{x}$. Then, the Riesz transform of f is a d-channel filterbank given by

$$\boldsymbol{\mathcal{R}} f(\mathbf{x}) = \begin{pmatrix} \mathcal{R}_1 f(\mathbf{x}) \\ \vdots \\ \mathcal{R}_d f(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} (h_1 * f)(\mathbf{x}) \\ \vdots \\ (h_d * f)(\mathbf{x}) \end{pmatrix} \tag{3}$$

where the filters $(h_n)_{n=1}^d$ are specified by their Fourier transform $\hat{h}_n(\boldsymbol{\omega}) = -\mathrm{j}\omega_n / \|\boldsymbol{\omega}\|$.

The Riesz transform commutes with a $(d \times d)$ rotation matrix $\mathbf{U}$, with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, as expressed by the following property of its impulse response:

$$\mathcal{F}\left\{\boldsymbol{\mathcal{R}}\{\delta\}(\mathbf{U}\cdot)\right\}(\boldsymbol{\omega}) = \mathbf{U}\mathcal{F}\left\{\boldsymbol{\mathcal{R}}\{\delta\}(\cdot)\right\}(\boldsymbol{\omega}). \tag{4}$$

There, $\delta(\mathbf{x}) = \prod_{n=1}^d \delta(x_n)$ is the multidimensional Dirac distribution. Hence, the rotated Riesz transform of an *isotropic function* f corresponds to the multiplication of $\mathbf{U}$ by the non-rotated Riesz transform

$$\boldsymbol{\mathcal{R}} f(\mathbf{U}\mathbf{x}) = \mathbf{U}\boldsymbol{\mathcal{R}} f(\mathbf{x}), \tag{5}$$

which demonstrates the steerability property of the Riesz transform. From an implementation point of view, the Riesz transform along the direction specified by the unit vector $\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{R}^d$ is computed as the smoothed version of a steerable gradient filter bank.

Other remarkable properties of the Riesz transform are shift and scale invariances, and inner-product preservation. The latter implies that, for $f, g \in L_2(\mathbb{R}^d)$, the Riesz transform satisfies the Parseval-like identity

$$\langle \boldsymbol{\mathcal{R}} f, \boldsymbol{\mathcal{R}} g\rangle_{L_2^d} = \sum_{n=1}^d \langle \mathcal{R}_n f, \mathcal{R}_n g\rangle_{L_2} = \langle f, g\rangle_{L_2}. \tag{6}$$

The essential consequence of (3)-(6) is that the Riesz transform maps any tight wavelet frame of $L_2(\mathbb{R}^d)$ into another one.

Finally, one obtains higher-order Riesz transforms by performing iterations of the first-order Riesz transform (details are explained in [9]). It is noteworthy that the higher-order Riesz transform also satisfies shift and scale invariances, and inner-product preservation.

3.2. Isotropic Mother-Wavelet Pyramid

A primary isotropic wavelet pyramid is required to preserve the relevance of the directional analysis performed by the Riesz transform. Moreover, the bandlimitedness of the wavelet bands must be enforced to ensure the isotropy of the primary together with the possibility of downsampling. At this point, we remind that a conventional orthogonal and separable wavelet transform does not meet these conditions. One of the popular wavelet designs that satisfy these constraints is the Simoncelli's wavelet

$$\hat{\psi}_{\mathrm{Sim}}(\boldsymbol{\omega}) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2\left(\frac{2\|\boldsymbol{\omega}\|}{\pi}\right)\right), & \frac{\pi}{4} \leq \|\boldsymbol{\omega}\| \leq \pi \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Therefore, we build a steerable wavelet pyramid by applying the Riesz transform to the mother wavelet. For instance, one constructs *steerable* gradient-like wavelets by applying the first-order Riesz transform to an isotropic pyramid. Since the primary wavelet is isotropic, the directionality information at each scale is conveyed only by the Riesz transform.

4. CONSISTENT WAVELET DENOISING

In this section, we develop a simple extension to conventional wavelet denoising with regularized least-squares. In particular, we are interested in applying the method to the steerable wavelet transform to improve its performance.

4.1. Constrained Formulation

As mentioned in Section 2, the tightness of the frame implies that (1) always holds. However, since $m > n$, there exist $\mathbf{w} \in \mathbb{R}^m$ for which $\mathbf{W}\mathbf{W}^T\mathbf{w} \neq \mathbf{w}$. This in turn implies that the solution obtained by solving (2) in the tight wavelet-frame domain is not *consistent*, in the sense that the wavelet transform of the final solution $\widehat{\mathbf{x}} = \mathbf{W}^T\widehat{\mathbf{w}}$ is not necessarily equal to $\widehat{\mathbf{w}}$ obtained from (2). However, since the desired solution always lives in the image domain, it makes sense to constrain the wavelet-domain estimate to be consistent, which also enforces the perfect invertibility of the transform. Such estimation can be performed by solving the constrained-optimization problem

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\min}\, \mathcal{J}(\mathbf{u}, \mathbf{w}) \quad \text{s.t.} \quad \mathbf{W}\mathbf{W}^T\mathbf{w} = \mathbf{w}, \tag{8a}$$

with

$$\mathcal{J}(\mathbf{u}, \mathbf{w}) = \frac{1}{2}\|\mathbf{w} - \mathbf{u}\|_2^2 + \lambda\Phi(\mathbf{w}). \tag{8b}$$

For the case of $\Phi(\cdot) = \|\cdot\|_1$, the optimization admits a simple interpretation as the estimation of a sparse and consistent wavelet-domain solution. In our recent work [10], a similar estimation procedure was used to improve the performance of cycle spinning, which works by considering different shifts of the signal to make the wavelet basis shift-invariant [11].

4.2. Algorithm

A practical implementation of the method can be obtained with the help of the augmented-Lagrangian approach [12], which casts the constrained-optimization (8) as a sequence of unconstrained problems. The idea is to replace the objective function $\mathcal{J}$ with the new penalty function defined as

$$\mathcal{L}(\mathbf{w}, \mathbf{x}) = \mathcal{J}(\mathbf{u}, \mathbf{w}) + \frac{\tau}{2}\|\mathbf{w} - \mathbf{W}\mathbf{x}\|_2^2 - \boldsymbol{\mu}^T(\mathbf{w} - \mathbf{W}\mathbf{x}), \quad (9)$$

where $\tau > 0$ is the penalty parameter and $\boldsymbol{\mu} \in \mathbb{R}^m$ is the vector of Lagrange multipliers. The condition $\mathbf{w} = \mathbf{W}\mathbf{x}$ asserted by the penalty constrains $\mathbf{w}$ to live in the column space of $\mathbf{W}$, which is equivalent to the consistency condition $\mathbf{w} = \mathbf{W}\mathbf{W}^T\mathbf{w}$. To minimize the objective (9), we alternate between solving the problem for $\mathbf{w}$ with $\mathbf{x}$ fixed and *vice versa*. For simplicity, we now assume $\Phi(\cdot) = \|\cdot\|_1$; however, the method can be readily extended to other regularizers as demonstrated in [10, 13].

Given the noisy wavelet-domain measurements $\mathbf{u} \in \mathbb{R}^m$, regularization parameter $\lambda > 0$, penalty parameter $\tau > 0$, and initial solution $\widehat{\mathbf{x}}^0$, the estimation proceeds as follows:

1. *Initialize:* Set $t = 0$ and $\boldsymbol{\mu}^0 = 0$.
2. *Update* $\widehat{\mathbf{w}}$*:* Minimize $\mathcal{L}$ with respect to $\mathbf{w}$ with $\mathbf{x}$ fixed

$$\begin{aligned}\widehat{\mathbf{w}}^{t+1} &= \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\min}\, \mathcal{L}(\mathbf{w}, \widehat{\mathbf{x}}^t) \\ &= \underset{\mathbf{w}\in\mathbb{R}^m}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{w} - \tilde{\mathbf{u}}\|_2^2 + \tilde{\lambda}\|\mathbf{w}\|_1 \right\} \\ &= \eta(\tilde{\mathbf{u}}, \tilde{\lambda}), \end{aligned} \quad (10)$$

   where $\tilde{\mathbf{u}} = (\mathbf{u} + \tau\mathbf{W}\widehat{\mathbf{x}}^t + \boldsymbol{\mu}^t)/(1+\tau)$ and $\tilde{\lambda} = \lambda/(1+\tau)$.

3. *Update* $\widehat{\mathbf{x}}$*:* Minimize $\mathcal{L}$ with respect to $\mathbf{x}$ with $\mathbf{w}$ fixed

$$\begin{aligned}\widehat{\mathbf{x}}^{t+1} &= \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\, \mathcal{L}(\mathbf{w}^{t+1}, \mathbf{x}) \\ &= \mathbf{W}^T\left(\widehat{\mathbf{w}}^{t+1} - \frac{\boldsymbol{\mu}^t}{\tau}\right). \end{aligned} \quad (11)$$

4. *Update* $\boldsymbol{\mu}$*:* Lagrange multipliers are updated according to the simple rule

$$\boldsymbol{\mu}^{t+1} = \boldsymbol{\mu}^t - \tau(\widehat{\mathbf{w}}^{t+1} - \mathbf{W}\widehat{\mathbf{x}}^{t+1}). \quad (12)$$

5. Set $t = t + 1$ and proceed to Step 2.

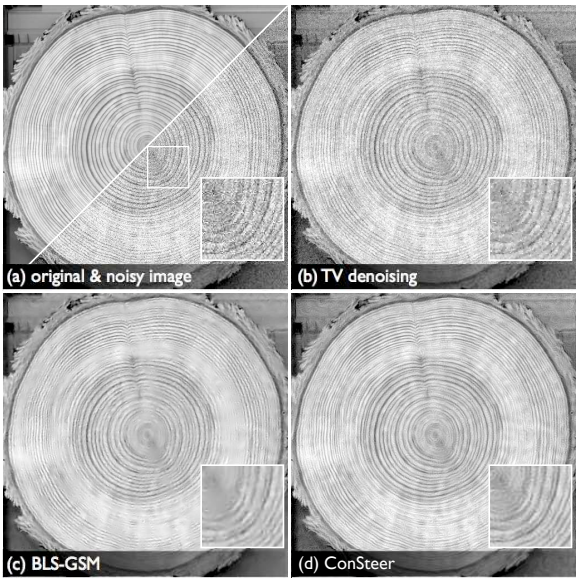

Fig. 1: *Comparison of denoising results for wood: (a) original and noisy image (SNR=15 dB); b) TV denoising (SNR=19.12 dB); (c) BLS-GSM denoising (SNR=22.46 dB); (d) ConSteer denoising (SNR=23.05 dB).*

For each iteration $t = 1, 2, 3, \ldots$, the proposed rules produce estimates $\widehat{\mathbf{x}}^t$ of the true signal $\mathbf{x}$. All updates are easy to compute since they reduce to evaluation of a wavelet-transform, pointwise soft-thresholding, and inverse wavelet-transform. Finally, note that, our method is minimizing a strictly convex function under linear constraints, which implies that the algorithm converges to the global minimum.

5. NUMERICAL EXAMPLES

We consider a simple simulation that illustrates the performance of our *consistent steerable-wavelet shrinkage* (ConSteer) algorithm. Specifically, we evaluate our method on three test images corrupted by additive white Gaussian noise. The denoising performance is compared in terms of the signal-to-noise ratio (SNR). Denoising results on the *wood* image for an input SNR value of 15 dB are shown in Figure 1.

We present in Table 1 the performance of five standard denoising algorithms: Haar-wavelet shrinkage, *total variation* (TV) denoising, traditional steerable-wavelet shrinkage, *Bayesian least squares Gaussian scale mixture* (BLS-GSM), and ConSteer. Each regularization parameter λ is optimized for the best SNR performance for each scale and orientation of the image representation. Three orientation bands have been chosen for the steerable pyramid used in BLS-GSM denoising as well as for the steerable-wavelet transform used in steerable-wavelet shrinkage and the ConSteer algorithms. For the ConSteer algorithm, tensor-based steering [9] proved to be more robust and computationally cheaper. It was carried out on the noisy data. The results presented in Table 1

Table 1: Evaluation and comparison of ConSteer.

Input SNR (dB)	10	15	20
		zoneplate	
Haar-wavelet shrink	11.68	15.85	20.55
TV denoising	13.38	17.12	21.35
Steerable-wavelet shrink	20.53	23.03	27.12
BLS-GSM	21.59	25.23	29.32
ConSteer	**22.07**	**26.38**	**30.13**
		wood	
Haar-wavelet shrink	14.68	17.98	21.71
TV denoising	16.30	19.12	22.54
Steerable-wavelet shrink	17.04	20.69	23.99
BLS-GSM	**19.9**	22.46	25.54
ConSteer	18.82	**23.05**	**26.17**
		barbara	
Haar-wavelet shrink	14.01	17.49	21.78
TV denoising	15.86	18.78	23.61
Steerable-wavelet shrink	17.31	20.03	22.68
BLS-GSM	**19.55**	**22.73**	**26.02**
ConSteer	18.67	21.54	24.73

lead to three conclusions. First, imposing consistency always improves the quality of denoising. This has already been observed in [10], where it is shown that consistent cycle spinning with Haar-wavelets is equivalent to TV denoising. Here, we see that the same observation is valid in the case of steerable wavelets, where ConSteer outperforms standard steerable-wavelet shrinkage for all input SNRs. Second, for certain images that are rich in directional cues, our simple ℓ_1-based ConSteer yields competitive performance with state-of-the-art methods like TV denoising and BLS-GSM. Third and last, our simulations indicate that the performance of ConSteer can be further improved by finding suitable strategies for steering. For example, we observed an additional 1-2 dB gain over the results in Table 1 when the steering direction was estimated on the original noiseless image.

6. CONCLUSION

We have presented a method that improves the performance of ℓ_1-based image denoising with steerable wavelets. The method is computationally simple and attains high-quality results that are competitive with other similar techniques. We believe that the ideas presented in this work are applicable to other state-of-the-art wavelet-domain algorithms. Future work will include possible extensions to other statistical priors and applications to extended inverse problems of the type $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$.

7. RELATION TO PRIOR WORK

The literature on image denoising is vast [5, 8–11]. Nevertheless, one may single out two classes of methods, such as total-variation [10, 14] and wavelet-domain approaches [5, 11]. In this paper, we have focused on wavelets, with a particular emphasis on Riesz-transform-based steerable wavelets [9]. The improved performance of our approach is due to consistency, which was introduced in [10] for cycle spinning. Our work thus extends and improves upon [9, 10] by considering directional wavelets with consistent denoising.

8. REFERENCES

[1] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. International Conference on Image Processing*, vol. 3, (Washington, DC), pp. 444–447, October 1995.

[2] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory*, vol. 38, pp. 587–607, March 1992.

[3] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Patt. Anal. and Machine Intell.*, vol. 13, pp. 891–906, September 1991.

[4] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 49–70, 2000.

[5] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in wavelet domain," *IEEE Trans. Image Process.*, vol. 12, pp. 1338–1351, November 2003.

[6] F. Denis and A. Baskurt, "Multidirectional curvilinear structures detection using steerable pyramid," *Journal of Electronic Imaging*, vol. 13, no. 4, pp. 756–765, 2004.

[7] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval with gaussianized steerable pyramids," *IEEE Trans. Image Process.*, vol. 15, pp. 2702–2718, September 2006.

[8] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, November 2004.

[9] M. Unser, N. Chenouard, and D. V. D. Ville, "Steerable pyramids and tight wavelet frames in $L_2(\mathbb{R}^d)$," *IEEE Trans. Image Process.*, vol. 20, p. 2705, October 2011.

[10] U. S. Kamilov, E. Bostan, and M. Unser, "Wavelet shrinkage with consistent cycle spinning generalizes total variation denoising," *IEEE Signal Process. Lett.*, vol. 19, pp. 187–190, April 2012.

[11] R. R. Coifman and D. L. Donoho, *Springer Lecture Notes in Statistics*, ch. Translation-invariant de-noising, pp. 125–150. Springer-Verlag, 1995.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 2 ed., 2006.

[13] E. Bostan, U. S. Kamilov, and M. Unser, "Reconstruction of biomedical images and sparse stochastic modelling," in *Proc. Int. Symp. Biomedical Imaging*, (Barcelona, Spain), May 2012. to appear.

[14] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, November 1992.