VIDEO DENOISING USING ORIENTED COMPLEX WAVELET TRANSFORMS

Fei Shi and Ivan W. Selesnick

Polytechnic University
Department of Electrical and Computer Engineering
6 Metrotech Center, Brooklyn, NY 11201

ABSTRACT

Video processing techniques using true 3D transforms are largely unexploited, partly because of the drawbacks of traditional separable 3D transforms. In this paper we use a new type of non-separable 3D wavelet transform for video denoising and overcome the motion-mixture problem by using oriented complex wavelets. This wavelet transform is a 3D version of Kingsbury's 1D and 2D dual-tree wavelet transforms. We also investigate video denoising techniques using a combination of both 2D and 3D oriented wavelet transforms. The results are compared with those obtained by separable wavelet transforms.

Keywords: Video denoising, 3D wavelet, dual-tree complex wavelet transform

1. INTRODUCTION

Although video sequences can be perceived as 3D data volumes that have strong correlation both in space and in time, little research is focused on true 3D decomposition of video. This is mainly because the structure of traditional separable decompositions, e.g., the separable 3D wavelet transform, usually do not provide an efficient representation of motion – an important type of information contained in video. We can demonstrate this drawback by interpreting a 3D separable wavelet as a video sequence. As a function of time, the wavelet displays no conspicuous motion. This is true for any separable wavelet transform.

For more efficient representations of video, we turn to non-separable wavelet transforms. In this paper, we adopt the 3D version of Kingsbury's dual-tree complex wavelet transform ($\mathbb{C}$WT) [1, 2], which has the orientation property that leads to a motion-selective subband decomposition. Thus, this transform takes into account the motion of image elements for video denoising, without explicitly using motion estimation.

Furthermore, for such special 3D data as video, it is not optimal to treat all the three dimensions equally. In this paper we also introduce an algorithm that combines both the 2D and 3D oriented wavelet transforms for video denoising. We compare both the PSNR values and the visual effects of denoising results using the separable 3D wavelet transform, the 2D $\mathbb{C}$WT, the 3D $\mathbb{C}$WT, and the combined method.

We gratefully acknowledge support from ONR (ONR grant N00014-03-1-0217). We also thank Shiyan Hu for reviewing earlier draft of the paper.

2. PRELIMINARIES ON THE ORIENTED COMPLEX WAVELET TRANSFORM

We begin with 2D to illustrate the orientation of the complex wavelet transform. For more details, refer to [3]. As shown in Fig. 1, the standard separable 2D wavelet transform has the checkerboard artifact that arises in the separable implementation. For the separable 2D wavelet transform, the 2D wavelet is given by $\psi_r(x,y) = \psi_r(x)\,\psi_r(y)$, where $\psi_r(\cdot)$ is a real wavelet. If we idealize the wavelet as high-pass, the 2D separable wavelet for the HH subband has four separate passbands. Therefore the wavelet mixes directions of $\pm 45°$.

The dual-tree $\mathbb{C}$WT overcomes this artifact by using analytic complex wavelets – wavelets that have one sided spectrums. Design techniques are addressed in [1, 4, 5]. The key idea is to design a pair of wavelets $\psi_h(x)$ and $\psi_g(x)$ such that $\psi_g(x) = \mathcal{H}\{\psi_h(x)\}$, where $\mathcal{H}$ denotes the Hilbert transform. Then $\psi(x) = \psi_h(x) + j\,\psi_g(x)$ is analytic. The 2D complex wavelet is given by $\psi(x,y) = \psi(x)\,\psi(y)$ and has only one passband. Therefore it's real part has two symmetric passbands and captures only one direction. For the other direction, consider the real part of the complex wavelet $\psi_2(x,y) = \psi(x)\,\overline{\psi(y)}$, where the overline represents complex conjugation. We illustrate the real part of one particular 2D dual-tree wavelet and its idealized spectrum in Fig. 2.

3. THE 3D ORIENTED COMPLEX WAVELET TRANSFORM

For the separable 3D wavelet transform, the wavelets mix even more orientations, as shown in Fig. 3. However, the problem can again be solved by the same principle used in the 2D dual-tree $\mathbb{C}$WT. Using the same Hilbert transform

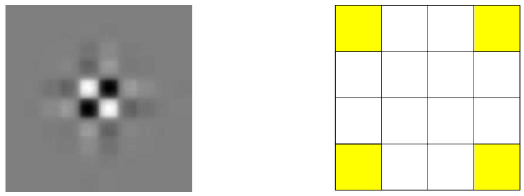

Fig. 1. Separable 2D wavelet of the HH subband and its idealized spectrum (shadow represents passband) .

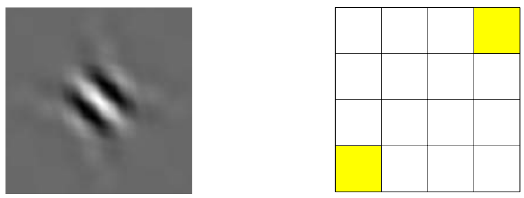

Fig. 2. Non-separable 2D complex wavelet (real part only) of the HH subband and its idealized spectrum.

pair $\psi_h(\cdot)$ and $\psi_g(\cdot)$, consider the 3D complex wavelet,

$$\psi(x, y, z) = \psi(x)\,\psi(y)\,\psi(z) = \\ (\psi_h(x) + j\,\psi_g(x))(\psi_h(y) + j\,\psi_g(y))(\psi_h(z) + j\,\psi_g(z)). \quad (1)$$

As in the 2D case, we take the real part of $\psi(x, y, z)$ to get an oriented real 3D wavelet. Define:

$$\psi_1 := \psi_h(x)\,\psi_h(y)\,\psi_h(z) \quad (2)$$
$$\psi_2 := \psi_g(x)\,\psi_g(y)\,\psi_h(z) \quad (3)$$
$$\psi_3 := \psi_g(x)\,\psi_h(y)\,\psi_g(z) \quad (4)$$
$$\psi_4 := \psi_h(x)\,\psi_g(y)\,\psi_g(z). \quad (5)$$

The real part of $\psi(x, y, z)$ in (1) can be written compactly as

$$\text{Real Part}\{\psi(x, y, z)\} = \psi_1 - \psi_2 - \psi_3 - \psi_4. \quad (6)$$

The resulting 3D dual-tree $\mathbb{C}$WT only captures one orientation (shown in Fig. 4) instead of four with the separable wavelet transform. For the other three orientations, we consider the real part of $\psi(x)\,\psi(y)\,\overline{\psi(z)}$, $\psi(x)\,\overline{\psi(y)}\,\psi(z)$, and $\psi(x)\,\overline{\psi(y)}\,\overline{\psi(z)}$. This gives the following orthonormal combination matrix of the four separable 3-D wavelet transforms:

$$\begin{aligned} \psi_a(x, y, z) &= 0.5\,(\psi_1 - \psi_2 - \psi_3 - \psi_4) \\ \psi_b(x, y, z) &= 0.5\,(\psi_1 - \psi_2 + \psi_3 + \psi_4) \\ \psi_c(x, y, z) &= 0.5\,(\psi_1 + \psi_2 - \psi_3 + \psi_4) \\ \psi_d(x, y, z) &= 0.5\,(\psi_1 + \psi_2 + \psi_3 - \psi_4)\,. \end{aligned} \quad (7)$$

The 3D dual-tree wavelets are motion-selective. This property of can be better illustrated by video clips of the wavelets, where both motion and edge information is displayed.

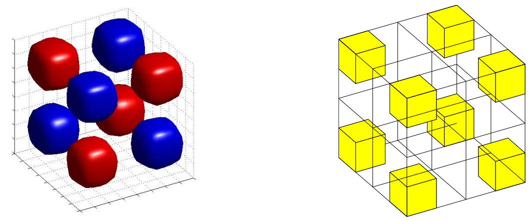

Fig. 3. Isosurface of a separable 3D complex wavelet (real part only) of the HHH subband and its idealized spectrum. The two colors represents surfaces of positive and negative values, respectively.

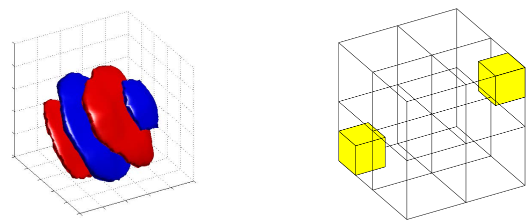

Fig. 4. Isosurface of a non-separable 3D complex wavelet (real part only) of the HHH subband and its idealized spectrum. The two colors represents surfaces of positive and negative values, respectively.

4. VIDEO DENOISING USING THE 3D DUAL-TREE TRANSFORM

In this paper we use soft thresholding with only the real part of the $\mathbb{C}$WT. This is enough to show the advantage of oriented wavelet transforms and requires less computations, though better performance is expected by using both the real and imaginary part. For the purpose of comparison, for each transform the optimal threshold is found using the original noise-free data. The same threshold is used for each subband of the respective types of wavelet transform. In practice, the threshold can be estimated from the noise data as in [6].

The test video consists of a stationary view with a person walking across the scene. Frames from the test video, the noisy video, and each of the processed videos are shown in Fig. 5 and 6. The two figures contain two typical types of video frames, one with only a stationary background and the other with a moving object. The PSNR value of the denoised video using the 2D $\mathbb{C}$WT is 27.25. In this case, the 2D $\mathbb{C}$WT is applied to each frame individually. Using the 3D separate wavelet transform, the PSNR value is 27.81, while using the 3D $\mathbb{C}$WT it is 28.99. The difference of the denoised videos is more visible when viewing the video sequence.

However the resulting videos also show the limitations of the oriented 3D dual-tree transform. For frames containing fast motion, the dual-tree 2D transform can give a superior result, which can be perceived in frame 21 when a person passes through the scene (Fig. 6). For this frame, the 2D dual-tree transform does best because for fast motion it is more difficult to exploit the temporal correlation of pixel values. The 2D and 3D oriented transforms each have their benefits for video denoising. Inspired by the work by Starck, Candés and Donoho [7] on image denoising using both the curvelet and the separable wavelet transforms, we investigate likewise the combined use of the 2D and 3D oriented wavelet transforms for video denoising, which will be introduced in the next section.

5. VIDEO DENOISING USING COMBINED 2D AND 3D DUAL-TREE TRANSFORM

Since video is a quite unique type of 3D data, it's intuitive to treat the third dimension – time, differently. Here we use an algorithm seeking the best representation of a video, using simultaneously the 2D and 3D dual-tree transforms. The idea is to seek the best basis that leads to the sparsest representation of the 3D data, using the 2D and 3D wavelets as candidate dictionaries. Various methods have been developed to achieve (or at least to approach) this [8, 9, 10, 11], among which are the Matching Pursuit method [8] and Basis Pursuit method [9]. Here we use the BPDN (Basis-Pursuit denoising) algorithm which is also adopted by Starck *et.al.* in [12]. It is implemented by a numerical Block-Coordinate-Relaxation method [13].

The algorithm is aimed at minimizing the target function

$$\min_{\{\underline{X}_2, \underline{X}_3\}} \sqrt{2}||W_2\underline{X}_2||_1 + ||W_3\underline{X}_3||_1 + \frac{1}{2\lambda}||\underline{X} - \underline{X}_2 - \underline{X}_3||_2^2, \tag{8}$$

where W_2 and W_3 are the 2D and 3D dual-tree wavelet transform, respectively, $\underline{X}_2$ and $\underline{X}_3$ are the part of video data best represented by the 2D and 3D transform, respectively, and $\underline{X}$ is the original (noisy) video data.

Under the occurrence of noise, we seek to minimize the l_1 norm of the coefficients with a penalty term of weighted square error between the sum of the two decomposition results and the given video. Note that the weight $\sqrt{2}$ is used to balance the 2D and 3D coefficients because using our dual-tree wavelet package, the norms of 2D wavelets are about $\sqrt{2}$ times those of 3D wavelets. (This is because both are tight frames and because the 3D $\mathbb{C}$WT is twice as expansive as the 2D $\mathbb{C}$WT.)

The numerical method can be described as follows:

1. Extract the low frequency part $\underline{X}_l$. Let $\underline{X} = \underline{X} - \underline{X}_l$.
2. Initialize threshold $T = \lambda \cdot L$, number of iterations N, $\underline{X}_2 = \underline{X}_3 = \underline{0}$.
3. Repeat updating $\underline{X}_2$ and $\underline{X}_3$ for N times:
   (a) $\underline{R} = \underline{X} - \underline{X}_3$;
   (b) $\underline{w}_2 = W_2(\underline{R})$;
   (c) $\underline{w}_2 = \text{soft}(\underline{w}_2, \sqrt{2}T), \underline{X}_2 = W^{-1}\underline{w}_2$;
   (d) $\underline{R} = \underline{X} - \underline{X}_2$;
   (e) $\underline{w}_3 = W_3(\underline{R})$;
   (f) $\underline{w}_3 = \text{soft}(\underline{w}_3, T), \underline{X}_3 = W^{-1}\underline{w}_3$.;
4. If $T > T_n$, $T = T - \lambda$, go to step 3.
5. Add back the low frequency part. $\underline{X}_3 = \underline{X}_3 + \underline{X}_l$

As mentioned by Starck *et.al.*, the advantages of this algorithm include the following: (a) There is no need to keep all the transform coefficients in memory, which is good especially for redundant transforms. (b) The algorithm has the capacity to include various constraints for optimization, which makes it flexible. The algorithm also automatically thresholds the coefficients for denoising purpose.

Since fast motion causes singularities in the time dimension, it is difficult to represent it efficiently by the 3D transform. We expect that after the 2D and 3D combined decomposition, $\underline{X}_2$ mainly contains the fast motion. Indeed it is true for our results, as shown in Fig.7.

For comparison with the work in the previous section, we also choose the best threshold T_n using the original noise-free data. (In practice T_n is about twice of the noise variance.) The PSNR value for our combined method is 28.46. Though it's not as good as the 3D method, the denoising result is surprisingly better visually. In Fig.5 and 6 we also show the denoising result of the two frames.

6. CONCLUSIONS AND FUTURE WORK

In this paper we use the oriented 3D $\mathbb{C}$WT for video denoising and obtain better results than the separable wavelet transform. The 3D $\mathbb{C}$WT is the 3D version of Kingsbury's dual-tree complex wavelet transform. It will also be of interest to extend to 3D other directional signal representation methods, e.g., curvelets [14, 15], directional filter banks and pyramids [16, 17], complex filter banks [18], the steerable pyramid [19, 20] and to investigate their use for video denoising. This paper also introduces a numerical optimization method for 2D and 3D combined wavelet decomposition and denoising. Without significantly degrading the PSNR value, it improves the visual quality of the denoised video. For this method, a better choice of parameters and constraints need further consideration.

7. REFERENCES

[1] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *Proc. Eighth IEEE DSP Workshop*, Utah, August 9-12, 1998.

[2] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Royal Society London A*, Sept. 1999.

[3] I. W. Selesnick and K.-L. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Wavelet Applications in Signal and Image Processing X (Proc. SPIE 5207)*, San Diego, 2003.

[4] N. G. Kingsbury, "A dual-tree complex wavelet transform with improved orthogonality and symmetry properties," in *Proc. IEEE Int. Conf. Image Processing*, Vancouver, Canada, September 10-13, 2000.

[5] I. W. Selesnick, "The design of approximate Hilbert transform pairs of wavelet bases," *IEEE Trans. on Signal Processing*, vol. 50, no. 5, pp. 1144–1152, May 2002.

[6] L. Sendur and I.W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. on Signal Processing*, vol. 50, pp. 2744–2756, 2002.

[7] J.-L. Starck, E. J. Candés, and D. L. Donoho, "Very high quality image restoration by combining wavelets and curvelets," in *Proc. SPIE*, 2001, Wavelet Applications in Signal and Image Processing IX.

[8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[9] S. Chen, D. Donoho, and M. Saunder, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[10] D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via l_1 minimization," in *Nat. Acad. Sci.*, 2003, vol. 100, pp. 2197–2202.

[11] X. Huo, *Sparse image reprsentation via combined transforms*, Ph.D. thesis, Stanford University, 1999.

[12] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Processing*, 2003, submitted.

[13] A. Bruce, S. Sardy, and P. Tseng, "Block coordinate relaxation methods for nonparametric signal de-noising," in *Proc. SPIE*, 1998, vol. 3391, pp. 75–86.

[14] E. J. Candés and D. L. Donoho, "Curvelets, multiresolution representation, and scaling laws," in *Proc. SPIE*, San Diego, July 2000, vol. 4119, Wavelet Applications in Signal and Image Processing VIII.

[15] M. N. Do and M. Vetterli, "Pyramidal directional filter banks and curvelets," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2001.

[16] R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. on Signal Processing*, vol. 40, no. 4, pp. 882–893, Apr. 1992.

[17] M. N. Do and M. Vetterli, "Contourlets: A directional multiresolution image representation," in *Proc. IEEE Int. Conf. Image Processing*, 2002.

[18] F. Fernandes, R. van Spaendonck, M. Coates, and S. Burrus, "Directional complex-wavelet processing," in *Proc. SPIE*, San Diego, July 2000, vol. 4119, Wavelet Applications in Signal and Image Processing VIII.

[19] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sept. 1991.

[20] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

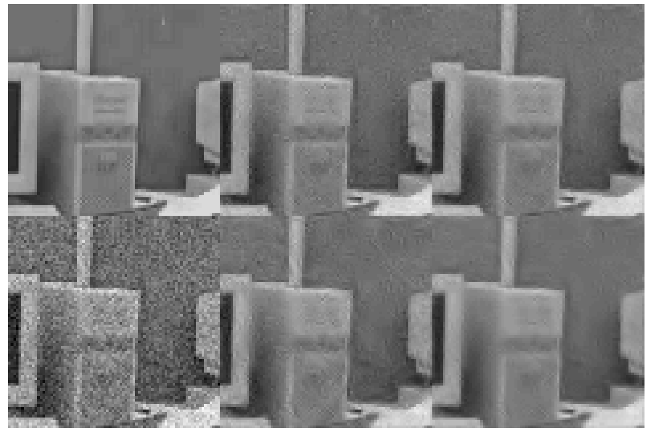

Fig. 5. Comparison of denoising results: frame 40 with stationary background. First row: original frame, 3D separable, 3D oriented. Second row: noisy frame, 2D oriented, combined 2D & 3D oriented.

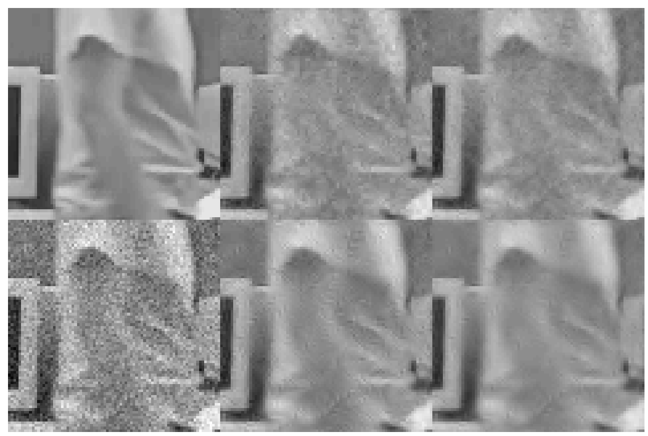

Fig. 6. Comparison of denoising results: frame 21 with moving person. Same ordering as in Fig. 5

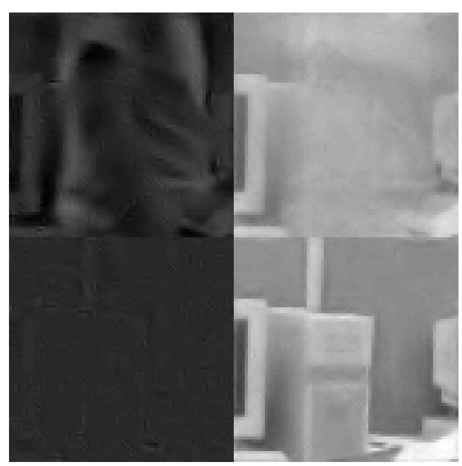

Fig. 7. Illustration of the 2D and 3D part separately. Frame 21 and 40. Left column: $\underline{X}_2$, right column: $\underline{X}_3$.