VIDEO PROCESSING USING THE 3-DIMENSIONAL SURFACELET TRANSFORM

Yue Lu and Minh N. Do

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
Email: {yuelu, minhdo}@uiuc.edu; Web: www.ifp.uiuc.edu/∼{yuelu, minhdo}

ABSTRACT

Motion estimation is a common ingredient in many state-of-the-art video processing algorithms, serving as an effective way to capture the spatial-temporal correlation in video signals. However, the robustness of motion estimation often suffers from problems such as ambiguities of motion trajectory (i.e. the aperture problem) and illumination variances. In this paper, we explore a new framework for video processing based on the recently proposed surfacelet transform. Instead of containing an explicit motion estimation step, the surfacelet transform provides a motion-selective subband decomposition for video signals. We demonstrate the potential of this new technique in a video denoising application.

1. INTRODUCTION

Motion estimation is an effective way to capture the spatial-temporal correlation in video signals, and is widely used in many state-of-the-art video processing algorithms. Its importance has been confirmed by the ubiquitous existence of motion estimation/compensation steps in the latest series of video coding standards, from MPEG-1 to H.264.

However, the robustness of motion estimation often suffers from problems such as ambiguities of motion trajectory (i.e. the aperture problem), illumination variances, and noise in the video sequences. Meanwhile, for practical applications, there is always a trade-off between the accuracy of the motion vectors and computational complexity.

In this paper, we explore a new framework for video processing based on the recently proposed surfacelet transform [1]. Instead of containing an explicit motion estimation step, the surfacelet transform provides a motion-selective subband decomposition for video signals. Similar approaches to video processing using various motion-selective 3-D transforms have been previously proposed by several researchers [2]–[4]. A potential advantage of the surfacelet transform is that its directional resolution can be refined by invoking more levels of decomposition. In practice, we usually choose to have 192 or more directional subbands at finer scales of the surfacelet transform, which can potentially lead to a more accurate separation of different motion information in the video signals.

This paper is organized as follows. Section 2 briefly introduces the key ideas and construction of the surfacelet transform. We show in Section 3 why the surfacelet transform can potentially provide an efficient representation for video signals without explicit motion estimations. We present experimental results on video denoising in Section 4 and conclude the paper in Section 5.

This work was supported by the US National Science Foundation under Grant CCR-0237633 (CAREER).

2. THE SURFACELET TRANSFORM

In this section, we briefly introduce the filter bank construction of the surfacelet transform, with emphasis on the key ideas and properties. More details of this transform can be found in [1].

2.1. The 3-D Directional Filter Banks

In 1992, Bamberger and Smith [5] proposed the directional filter bank (DFB) for an efficient directional decomposition of 2-D signals. The DFB is implemented via an l-level tree-structured decomposition that leads to 2^l subbands with wedge-shaped frequency partitioning as shown in Figure 1(a). The directional-selectivity and efficient structure of the DFB makes it an attractive candidate for many image processing applications. By combining the DFB with the Laplacian pyramid, Do and Vetterli [6] constructed the *contourlets*, which provides a directional multiresolution transform for sparse image representation.

The first step in the construction of the surfacelet transform is extending the 2-D DFB to higher dimensions. For example, in 3-D, we want to achieve the frequency partitioning as shown in Figure 1(b), where the ideal passbands of the component filters are rectangular-based pyramids radiating out from the origin at different orientations and tiling the entire frequency space. We can see this is a natural extension from the wedge-shaped frequency partitioning in 2-D. In the rest of the paper, we use NDFB to represent the new directional filter banks in 3-D and beyond.

To obtain the first level of decomposition in NDFB, we employ a three-channel undecimated filter bank shown in Fig-

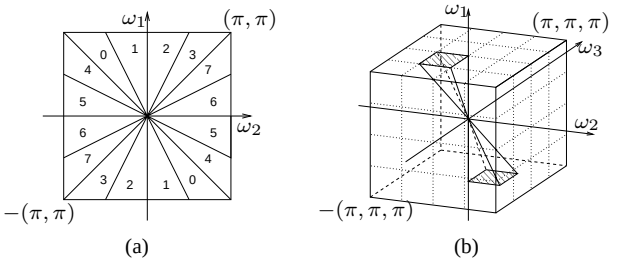


Fig. 1. (a) Frequency partitioning of the directional filter bank with 3 levels of decomposition. (b) Frequency partitioning of the NDFB in 3-D. The ideal passbands of the component filters are rectangular-based pyramids radiating out from the origin at 3×2^l ($l \geq 0$) different orientations and tiling the entire frequency space.

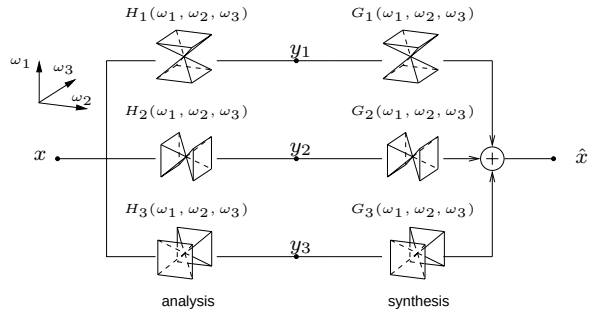


Fig. 2. The first level of decomposition: a three-channel undecimated filter bank in 3-D. The ideal frequency-domain supports of the component filters are hourglass-shaped regions, with their corresponding dominant directions aligned with the ω_1, ω_2, and ω_3 axes, respectively.

ure 2. This filter bank decomposes the 3-D frequency spectrum of the input signal into three hourglass-shaped subbands, with their dominant directions aligned with the ω_1, ω_2, and ω_3 axes, respectively.

Figure 3 shows the block diagram of subsequent levels of decompositions on one of the three branches. After the 3-D hourglass filter, we sequentially decompose the signal by two 2-D filter banks, with the first one, denoted as $\mathrm{IRC}_{12}^{(l_2)}$, operating along the (n_1, n_2)-plane and the second one, $\mathrm{IRC}_{13}^{(l_3)}$, along the (n_1, n_3)-plane.

The 2-D filter bank $\mathrm{IRC}_{12}^{(l_2)}$, which stands for Iteratively Resampled Checkerboard filter bank, has a binary-tree structure with l_2 (≥ 0) levels of decomposition, and therefore has 2^{l_2} different output branches. The second filter bank $\mathrm{IRC}_{13}^{(l_3)}$ has the same construction as $\mathrm{IRC}_{12}^{(l_2)}$, but operates along a different signal plane, i.e, $(n_1, n_2) \rightarrow (n_1, n_3)$, and with a different decomposition depth, i.e. $l_2 \rightarrow l_3$. Note that we attach an $\mathrm{IRC}_{13}^{(l_3)}$ to every output channel of $\mathrm{IRC}_{12}^{(l_2)}$, so we have a total of $2^{l_2+l_3}$ output channels in Figure 3. If the IRC filter banks are designed suitably, then the hourglass-shaped filter provided in the first level can be further divided into finer pyramid-shaped subbands as shown in Figure 1(b). We leave the details of the construction of the IRC filter banks to [1].

In summary, the NDFB filter bank described above has the following useful properties:

1. **Directional decomposition.** The NDFB decomposes 3-D signals into 3×2^l ($l \geq 0$) directional subbands, as shown in Figure 1(b). The number of directional subbands can be increased by iteratively invoking more levels of decomposition through a simple expansion rule.

2. **Construction.** The NDFB has an efficient tree-structured implementation using iterated filter banks.

3. **Perfect reconstruction.** The original signal can be exactly reconstructed from its transform coefficients in the absence of noise or other processing.

4. **Small redundancy.** The NDFB is 3-times expansive in the 3-dimensional case.

2.2. The Multiscale Pyramid

The 3-D directional frequency partitioning makes NDFB a suitable tool in capturing surface singularities within 3-D signals. To efficiently capture and represent local surface singularities with different sizes, we construct the surfacelet transform as a multiresolution version of NDFB. This strategy is analogous to the contourlet construction [6], in which the original 2-D DFB is combined with a multiscale decomposition. However, an important distinction is that instead of using the Laplacian pyramid as in contourlets, we employ a new multiscale pyramid structure for the surfacelet transform, as shown in Figure 4, which is conceptually similar to the one used in the steerable pyramid [7].

In Figure 4, we use $L_i(\boldsymbol{\omega})$ ($i = 0, 1$) to represent the lowpass filters and $D_i(\boldsymbol{\omega})$ ($i = 0, 1$) to represent the highpass filters in the multiscale decomposition. $S(\boldsymbol{\omega})$ is an anti-aliasing filter used to cancel the aliasing caused by the upsampling operations. The NDFB is attached to the highpass branch at the finest scale and bandpass branches at coarser scales. To have more levels of decomposition, we can recursively insert at point a_{n+1} a copy of the diagram contents enclosed by the dashed rectangle.

In the new multiscale pyramid depicted in Figure 4, the lowpass filter $L_0(\boldsymbol{\omega})$ in the first level is downsampled by a non-integer factor of 1.5 (upsampling by 2 followed by downsampling by 3) along each dimension. Although this fractional sampling factor makes the new pyramid slightly more redundant than the Laplacian pyramid (e.g. 1.34 versus 1.14 in redundancy ratios in 3-D), we find the added redundancy to be very useful in reducing the frequency-domain aliasing of the NDFB. We leave a detailed explanation for this issue to [8].

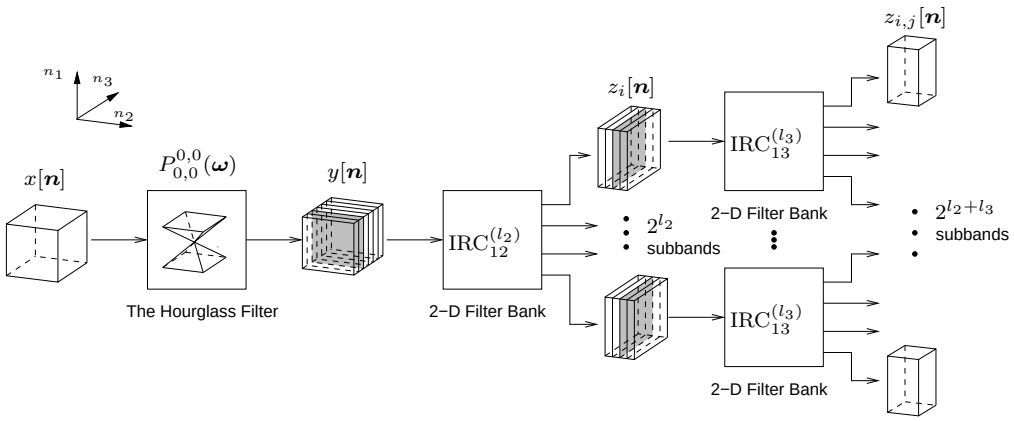


Fig. 3. One branch of the proposed filter bank structure of the NDFB in 3-D.

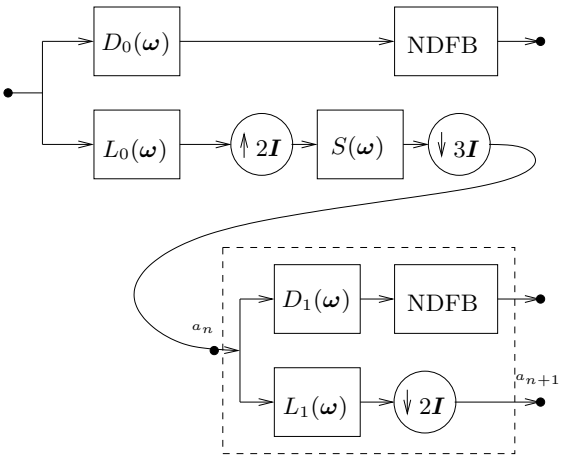


Fig. 4. The block diagram of the analysis part of the proposed surfacelet transform. The NDFB filter banks are attached to the highpass subband of the multiscale pyramid at each scale.

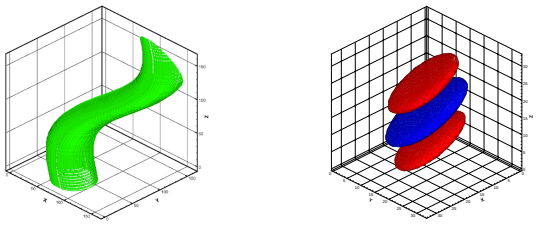
3. THE SURFACELET TRANSFORM FOR VIDEO PROCESSING

In this section, we explain, from two different points of view, why the surfacelet transform can potentially provide an efficient representation for video signals.

3.1. Spatial Domain

Video signals can be regarded as a special type of 3-D volumetric data, with two spatial dimensions and one temporal dimension. Moving objects in the video tend to carve out smooth surfaces in the 3-D spatial/temporal space. For instance, as shown in Figure 5(a), a moving circle in a video forms a smooth tube in the 3-D space.

We show in Figure 5(b) the isosurface of a surfacelet basis element in the spatial domain. The blue (or dark) colored isosurface is extracted at half of the most positive value and the red (or light) colored isosurface at half of the most negative value. We can see that surfacelets in the spatial domain are localized surface patches, smooth along the tangent planes and oscillatory along various normal directions. With these properties, we envision that surfacelets can efficiently capture and represent discontinuities living on the smooth surfaces in the 3-D spatial-temporal space.



(a) (b)

Fig. 5. (a) A smooth tube in the 3-D spatial-temporal space carved out by a moving circle in the video. (b) The isosurface of a surfacelet in the spatial domain.

3.2. Frequency Domain

Consider a simple translational motion model with a constant speed. The frames in the video are related as

$$I(x, y, t) = I(x - v_x t, y - v_y t, 0),$$

where $I(x, y, t)$ is the image intensity at location (x, y) and time t; while (v_x, v_y) is the constant motion vector. Although real-life video sequences usually contain much more complicated motions, the above model holds approximately in any local spatial-temporal region.

The Fourier transform $\hat{I}(\omega_x, \omega_y, \omega_t)$ of the 3-D volume can be written as

$$\hat{I}(\omega_x, \omega_y, \omega_t) = 2\pi \cdot \hat{I}_0(\omega_x, \omega_y) \cdot \delta(\omega_x v_x + \omega_y v_y + \omega_t), \quad (1)$$

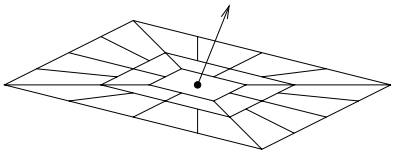


Fig. 6. The 3-D spectrum $\hat{I}(\omega_x, \omega_y, \omega_t)$ of the video signal is supported on a tilted plane with normal direction $(v_x, v_y, 1)^T$.

where $\hat{I}_0(x, y)$ is the 2-D Fourier transform of the initial frame at time 0. It follows from (1) that the 3-D spectrum is supported on a tilted 2-D plane whose normal direction is $(v_x, v_y, 1)^T$. As shown in Figure 6, the frequency partitioning of the surfacelet transform, when intersected with the tilted plane, generates a contourlet-like [6] directional multiresolution decomposition for the 2-D spectrum $\hat{I}_0(\omega_x, \omega_y)$. Consequently, when applied to video signals, the surfacelet transform can be seen as a motion-adaptive contourlet transform, whose basis images can represent 2-D edges moving at different directions.

4. VIDEO DENOISING

To demonstrate the potential of the surfacelet transform for video processing, we show in this section a video denoising application, in which we use the surfacelet transform to remove the white Gaussian noises added to several video signals, including "Mobile", "Coastguard" and "Tempete". For benchmark, we also show the denoising performance of the 3-D undecimated separable wavelet transform (UDWT) and the real-valued dual-tree wavelet transform (DTWT) [9, 10]. Note that DTWT can also provide a directional decomposition for 3-D signals, and has been previously applied in video processing applications [4].

In the experiment, we use 4 levels of decomposition for all transforms. For the surfacelet transform, the number of directional subbands for each scale, from fine to coarse, is set to be 192, 192, 48 and 12. For the UDWT, we use the "symlet" of length 16. For a fair comparison, we employ the hard thresholding denoising method for all 4 transforms, by truncating the transform coefficients of the noisy sequences with a threshold T. We choose $T = 3 \cdot \hat{\sigma}$, where $\hat{\sigma}$ is the estimated noise standard deviation at the corresponding subband. Although not being the best denoising algorithm available, this simple hard thresholding scheme can often be a good indication of the potential of different transforms.

Table 1 shows the PSNR (in dB) of the denoised test sequences by using different transforms. In calculating the PSNR values, the first and last 5 frames are excluded to rule out the boundary-effect. To help interpret the results, we also list the redundancy ratio of each transform in the table, since the redundancy of a transform indicates its computational and memory efficiency, which is an important practical issue to consider in the context of 3-D signal processing.

Fig. 7. Denoised frames from the "Mobile" sequence. From left to right, top to bottom: noiseless, UDWT, DTWT, and Surfacelets.

We can see from the table that Surfacelet outperforms the other two transforms by a large margin (from 0.79 dB to 1.36 dB) for the tested sequences. In particular, despite being much less redundant than UDWT (4.02 versus 29 in redundancy ratios), the surfacelet transform still achieves significantly better results. This performance gain is mainly due to the fact that Surfacelet can separate and capture motion information in its various directional subbands, while UDWT tends to mix different directional information into one subband. A potential advantage of the surfacelet transform over DTWT is that Surfacelet can refine its angular resolution (i.e. have more directional subbands) by invoking more levels of decomposition. In practice, we usually choose to have 192 or more directional subbands at finer scales, in contrast to the fixed 28 directional subbands provided in the DTWT.

Figure 7 shows one frame of the denoised video sequence by using different transforms. We can see that image details are best preserved by the surfacelet transform. This difference in denoising quality is much more conspicuous when viewing the video sequences.

5. CONCLUSIONS

In this paper, we explored a new framework for video processing using the surfacelet transform, which provides a motion-selective subband decomposition for video signals. Desirable properties of this transform include refinable angular resolution which leads to a more accurate separation of different motion information; an efficient tree-structured implementa-

Table 1. PSNR values of the denoised sequences by using UDWT, DTWT, and the surfacelet transform. The redundancy of each transform is shown in parentheses.

	Mobile			Coastguard			Tempete		
σ	30	40	50	30	40	50	30	40	50
UDWT (29)	24.08	23.00	22.24	25.95	24.95	24.20	27.14	26.05	25.24
DTWT (4)	24.64	23.44	22.60	26.06	25.02	24.23	27.19	26.07	25.26
Surfacelets (4.02)	**25.93**	**24.79**	**23.96**	**26.85**	**25.90**	**25.18**	**28.00**	**27.01**	**26.29**

tion; and a relatively low redundancy. The potential of the surfacelet transform is demonstrated by a video denoising application.

6. REFERENCES

[1] Y. Lu and M. N. Do, "Multidimensional directional filter banks and surfacelets," *IEEE Trans. Image Proc.*, to appear.

[2] E. Chang and A. Zakhor, "Subband video coding based on velocity filters," in *Proc. IEEE International Symposium on Circuits and Systems*, May 1992.

[3] F. A. Mujica, J.-P. Leduc, R. Murenzi, and M. J. T. Smith, "A new motion parameter estimation algorithm based on the continuous wavelet transform," *IEEE Trans. Image Proc.*, vol. 9, no. 5, pp. 873–888, May 2000.

[4] I. W. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Proc. of SPIE conference on Wavelet Applications in Signal and Image Processing X*, San Diego, USA, August 2003.

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

[6] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Proc.*, vol. 14, no. 12, December 2005.

[7] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inform. Th., Special Issue on Wavelet Transforms and Multiresolution Signal Analysis*, vol. 38, no. 2, pp. 587–607, March 1992.

[8] Y. Lu and M. N. Do, "A new contourlet transform with sharp frequency localization," in *Proc. IEEE Int. Conf. on Image Proc.*, Atlanta, USA, October 2006.

[9] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 10, pp. 234–253, 2001.

[10] I. W. Selesnick, "The double-density dual-tree DWT," *IEEE Trans. Signal Proc.*, vol. 52, no. 5, pp. 1304–1314, May 2004.