PLANELETS: A NEW ANALYSIS TOOL FOR PLANAR FEATURE EXTRACTION

Nasir Rajpoot *Roland Wilson* *Zhen Yao*

Department of Computer Science
University of Warwick
Coventry CV4 7AL, United Kingdom

e-mail: nasir,rgw,yao@dcs.warwick.ac.uk

ABSTRACT

Locally planar structures, formed by sweeping edges of objects, are commonly found in video sequences and convey most of the useful information. In this paper, the issue of efficient representation of such structures is addressed. We propose a novel representation tool which uses basis functions, termed as *planelets*, resembling planar structures and having compact support in space-time and spatiotemporal frequency. The representation is translation invariant, offers good directional selectivity, and can be computed efficiently. We show that the new representation, while being fast, produces video denoising results which compare favourably to one of the best known methods.

1. INTRODUCTION

Wavelets have gained significant popularity as a signal analysis and processing tool over the last decade or so. This is largely due to their ability to provide a localised, sparse representation of a signal (or image) which is inherently multiresolution in nature and can deal with point singularities. It should come as no surprise that wavelet-based solutions to many problems in the analysis and processing of video sequences have been proposed. These include denoising, coding, and motion compensation [1]. However, the performance of such algorithms is severely restricted due to the following observation. While the wavelet transform in higher dimensions can be conveniently computed separably, separability also seriously limits the ability of wavelets to efficiently represent higher dimensional features (such as lines in images or planes in 3D image volumes). For 2D images, non-separable representations such as ridgelets [2] and curvelets [3] have recently been developed, motivated by the same observation. Ridgelets have also been shown to be optimal for representing functions with linear singularities. Furthermore, the lack of frequency selectivity remains an elusive problem with most techniques operating in the wavelet domain.

In this paper, we present a novel representation designed specifically for efficiently representing 3D functions with planar singularities. Locally planar structures, such as moving luminance edges, are commonly found in video sequences and often convey most of the information. The new representation, termed as the *planelet* basis, has a combination of scale, translation, and directional characteristics which are well matched to the locally planar surfaces of interest in many applications. Extraction of such planar features may be useful in various applications, such as video denoising, video coding, geometry estimation [4] and tracking of objects in video sequences. The planelet representation offers translation invariance, good directional selectivity, and yet can be computed efficiently. The computational complexity of a planelet transform is $O(n)$, where n is the number of points in analysis window. In its current form, the representation provides a non-orthogonal basis and is redundant by less than 14%.

In the next section, the new basis is briefly described. The ability of planelets to extract planar surfaces is demonstrated in Section 3. Experimental results for restoring video sequences in very noisy environments show the superiority of our representation over the state-of-the-art method of denoising. The paper concludes with some remarks about the current work and directions for future work.

2. REPRESENTATION WITH PLANELETS

A prototypical planelet function in 1D is of the following form

$$f_{\xi,\omega,a}(x) = g(\frac{x-\xi}{a})\exp[-j\frac{\omega(x-\xi)}{a}] \quad (1)$$

where ξ, ω and a are respectively the location, frequency and scale parameters of the function. The function $g(.)$ is a window function, chosen alongwith the sampling interval to ensure invertibility of the discrete form of the transform. In 2D, the planelet basis can be regarded as a modification of the complex wavelet bases proposed in [5, 6], which show both translation invariance and directional selectivity, and may be used as an alternative to the ridgelet representation. In 3D, the basis comprises of the set of Cartesian products over ξ, ω at each scale a. That the continuous transform defined by (1) is invertible follows directly from the observation that it is simply the multiresolution Fourier transform (MFT) [5].

The discrete planelet transform (DPT), however, is significantly different from that described in [5]. It is a combination of two well known image transforms: the Laplacian pyramid [7] and the windowed Fourier transform (WFT). In some ways, it is similar to a 3D extension of the octave band Gabor representation proposed in [8], but avoids some of the more unpleasant numerical properties of the Gabor functions. The DPT of a video sequence x, in vector form, at scale m is given by

$$\hat{X}_m = \mathcal{F}_n(I - G_{m,m+1}G_{m+1,m})x_m \quad (2)$$

where $\hat{X}_m$ denotes the DPT at scale m, $\mathcal{F}_n$ is the WFT operator with window size $n \times n \times n$, I is the identity operator, x_m is the Gaussian pyramid representation of x at level m

$$x_m = \prod_{0}^{m-1} G_{l+1,l}x \quad (3)$$

and $G_{m,m+1}, G_{m+1,m}$ are the raising and lowering operators associated with transitions between levels in the Gaussian pyramid. Invertibility follows directly from equations (2) and (3):

Theorem 1 *The representation defined by equation (2) is invertible.*

Proof

First we note that the WFT operator $\mathcal{F}_n$ has an inverse, which can be denoted by $\mathcal{F}_n^{-1}$. Secondly, we know from Burt and Adelson that the Laplacian pyramid is invertible, since, trivially,

$$x_m = x_m - G_{m,m+1}x_{m+1} + G_{m,m+1}x_{m+1} \quad (4)$$

and the proof is completed by induction on m.

Importantly, although both the pyramid and WFT operators are Cartesian separable, the closeness of the Burt and Adelson filter to a Gaussian function gives the pyramid virtually isotropic behaviour, which can be exploited well by the high frequency resolution of a Fourier basis. The planelet basis functions resemble planar structures and have compact support in both space-time and spatiotemporal frequency.

3. PLANAR FEATURE EXTRACTION

Planelets provide an ideal tool for representing local planes in a video sequence (or an image volume, in general) due to their ability to localise planar surfaces which correspond to lines in the Fourier domain. The presence of planar surface in a local analysis window can be inferred by computing the eigenvalues of the local inertia tensor in the window and analysing them. The parameters for orientation of the local planar surface and translation from centre of the window can also be estimated by analysing the most significant coefficients in the locality. Consider a video sequence synthesised by moving the centre of a circle on a sinusoidal wave in the time direction. Nonlinear approximations of this sequence using only 0.07% of the wavelet and planelet coefficients are shown in Fig. 1(*a*) and 1(*b*), respectively. It is clear from this example that the planelet approximation of a video sequence containing locally planar surfaces can result in a smaller approximation error as compared to that using wavelets. Planelets, therefore, can also be used for a piecewise planar approximation of a video sequence. Moreover, the approximation can be made to be adaptive to the local scale of the planar surfaces.

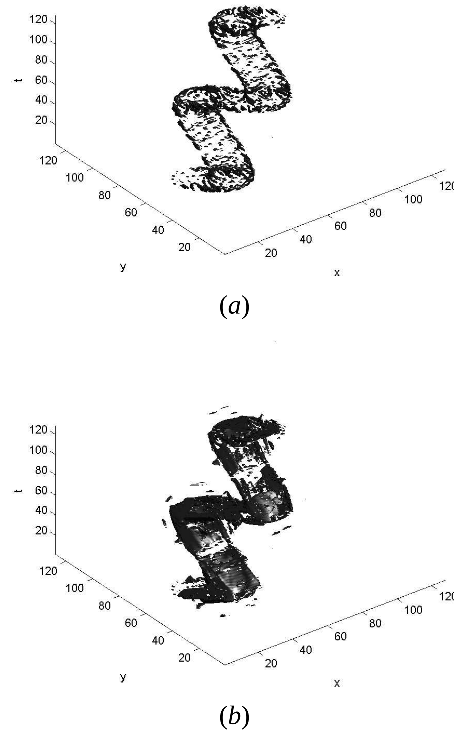

Fig. 1. Nonlinear approximation of a video sequence containing a moving circle using (*a*) wavelets (*b*) planelets

4. EXPERIMENTAL RESULTS

The ability of planelets to capture locally planar structures can be demonstrated in various applications, one of which is video denoising. In video sequences acquired in extremely noisy situations, it can be assumed that the coefficients which are relatively small in magnitude most likely correspond to the noise. This leads to a simple thresholding strategy in the planelet domain, which is akin to wavelet shrinkage method commonly used for denoising 2D images. Since the presence of additive Gaussian white noise means that almost all the planelet coefficients are affected by it, soft thresholding would provide an estimation of the original uncorrupted video sequence. The choice of threshold is crucial to the performance of a transform domain denoising algorithm [9]. We use either or both of the following thresholds: **(1)** a modification of the universal threshold proposed by Donoho and Johnstone [10]; $\theta_i = \mathcal{L}(\sigma)\sqrt{2\log n_i}$, where n_i denotes the number of coefficients at level i of the planelet decomposition, σ is the noise variance, and $\mathcal{L}(\sigma)$ is a suitably chosen function of σ, and **(2)** the *SUREshrink* (or simply *SURE*) threshold [1]. A translation invariant (TI) version of the planelet denoising was also implemented for a fair comparison with the translation invariant wavelet (TIW) denoising [11] in its 3D form.

The algorithms described above for video denoising were tested on four standard video sequences reduced to a resolution of 128^3: *Miss America*, *Football*, *Hall*, and *Tennis*. The image data was corrupted with additive Gaussian noise, and adaptive thresholding was applied to the transform coefficients of the noisy sequence represented in a 3-level planelet domain using a 16^3 window. Experimental results for the planelet denoising and the 3D TIW denoising [11] of the noisy sequences with signal-to-noise-ratio (SNR) values of 0dB, 5dB, 10dB, and 15dB are presented in Table 1. The function $\mathcal{L}(\sigma)$ was chosen to be $\mathcal{L}(\sigma) = a \log_{10} \sigma + b$ where $a, b \in \Re$ and $b = 2a$. A value of $a = 0.46$ was chosen empirically using least squares fitting. It can be observed from these results that the TI planelet denoising gives better performance than the 3D TIW denoising in terms of both visual quality and SNR gain. Denoising using TI planelet representation with *SURE* threshold generally outperforms the other two methods. Selected frames, for each of these sequences, restored by planelet denoising are shown in Fig. 2. Two types of artifacts were observed from non-TI results: blocking artifacts due to the use of a 16^3 window, and *fake textures* which sometimes persist within these windows, due to suppression of a significant amount of high frequency energy. The TI planelet denoising proves to be effective in removing both these kind of artifacts, and is particularly good in reconstructing some of the details that were smoothed out with the TIW method.

The computational complexity of our algorithm is $O(n)$ as compared to $O\left(N \log_2(N)\right)$ for TIW denoising, where n and N respectively denote the size of analysis window and the size of video sequence (resolution of each frame times the number of frames). While being faster by orders of magnitude, our algorithm still compares favourably to the 3D TIW for all our experiments.

Video Sequence	Noisy (dB)	TIW (dB)	Planelets (dB)		
			θ_i	*SURE*	*TI-SURE*
Miss America	0	17.9	17.1	17.3	18.1
	5	19.5	19.0	19.6	20.4
	10	21.5	20.8	21.5	22.4
	15	23.9	23.2	23.5	24.7
Football	0	11.9	11.7	12.1	12.5
	5	13.1	12.8	13.2	13.8
	10	15.0	14.4	14.7	15.6
	15	18.0	16.6	16.6	18.5
Hall	0	15.5	14.5	14.8	15.9
	5	16.6	16.7	17.2	18.2
	10	18.5	19.2	19.6	20.8
	15	20.6	21.7	21.8	23.9
Tennis	0	14.7	14.8	14.6	15.7
	5	16.6	16.5	16.7	17.6
	10	19.0	18.2	18.4	19.3
	15	21.7	19.9	20.1	21.4

Table 1. SNR (in dB) values for four standard video sequence

5. CONCLUSIONS

In this paper, planelets were proposed as an efficient representation tool for 3D functions with planar singularities. Such singularities are commonly found in video sequences in the form of moving luminance edges. It was shown that a piecewise planar approximation of a video sequence can be obtained by using a very small fraction of transform coefficients in the planelet domain. The ability of planelets to extract planar features from a video sequence makes them an attractive tool for analysis in various applications. For instance, decent restoration of video sequences captured in extreme noise is possible by thresholding of the planelet coefficients. The effects of oversampling and use of tapered windows remain to be investigated. Although the discrete planelet transform can be computed efficiently, its redundancy and high storage requirements may be of concern in some applications. Future work will address these issues and a further investigation into the usefulness of planelets in a wide range of video analysis applications.

6. REFERENCES

[1] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[2] E. Candes. *Ridgelets: Theory and Applications*. PhD thesis, Stanford University, Department of Statistics, 1998.

[3] E. Candes. Curvelets and curvilinear integrals. *Journal of Approximation Theory*, 113:59–90, 2000.

[4] A. Bhalerao and R. Wilson. A Fourier approach to 3D local feature estimation from volume data. In *Proceedings British Machine Vision Conference*, 2001.

[5] R.G. Wilson, A.D. Calway, and E.R.S. Pearson. A generalized wavelet transform for Fourier analysis: The multiresolution Fourier transform and its applications to image and audio signal analysis. *IEEE Transancation on Information Theory*, 38(2):674–690, March 1992.

[6] N.G. Kingsbury. Image processing with complex wavelets. *Phil. Trans. Royal Soc. A*, 1999.

[7] P.J. Burt and E.H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.

[8] M. Porat and Y.Y. Zeevi. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Trans. PAMI*, 10:452–468, 1988.

[9] M. Jansen. *Noise Reduction by Wavelet Thresholding*. Springer-Verlag, 2001.

[10] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 31:425–455, 1994.

[11] R.R. Coifman and D.L. Donoho. Translation-invariant denoising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, pages 125–150. Springer-Verlag, 1995.

[12] A. Kokaram. *Motion Picture Restoration*. Springer-Verlag, London, 1998.

Fig. 2. Denoising results for four standard video sequences

(*a*) Frame# 90 of *Miss America* (*b*) Noisy (SNR=0dB) (*c*) TIW (SNR=17.9dB) (*d*) TI-Planelets (SNR=18.1dB)

(*e*) Frame# 60 of *Football* (*f*) Noisy (SNR=5dB) (*g*) TIW (SNR=13.1dB) (*h*) TI-Planelets (SNR=13.8dB)

(*i*) Frame# 106 of *Hall* (*j*) Noisy (SNR=10dB) (*k*) TIW (SNR=18.5dB) (*l*) TI-Planelets (SNR=20.8dB)

(*m*) Frame# 57 of *Tennis* (*n*) Noisy (SNR=15dB) (*o*) TIW (SNR=21.7dB) (*p*) TI-Planelets (SNR=21.4dB)