Adaptive Wavelet Restoration of Noisy Video Sequences

Nasir Rajpoot **Zhen Yao** **Roland Wilson**

Department of Computer Science, University of Warwick, Coventry, UK

emails: `{nasir,yao,rgw}@dcs.warwick.ac.uk`

Overview

Our approach to the problem of denoising video sequences assumes that any pre-dominant spatiotemporal structures present in a video sequence can be captured even in the presence of noise, provided an appropriate analysis tool is used. We employ adaptive 3D wavelet packet transform to represent a video sequence, and use an optimal threshold (which requires no prior knowledge of the noise variance) to kill coefficients that may be more influenced by noise. The results are quite promising, in terms of both SNR and visual quality.

Motivation : The case for thresholding in spatiotemporal adaptive wavelet domain is supported by the fact that certain errors in motion estimation can be overcome by including the temporal direction in the realm of wavelet domain.

Introduction : It is often desirable to remove noise from video sequences captured in noisy environments or corrupted by noise during transmission, in broadcast and surveillance applications to name only a few. Noise removal by thresholding in the wavelet domain, a method also known as the *wavelet shrinkage*, has become increasingly popular in recent years. Regardless of which thresholding method is employed for denoising the signal, the algorithm is fast and offers the advantage that both compression and restoration of a signal can be achieved simultaneously.

Contributions : We present a novel algorithm based on the 3D extension of translation invariant denoising using an adaptive wavelet packet representation of restoration of noisy video sequences. The novelty of the algorithm lies in:

(a) extending the adaptive wavelet packet representation to include the temporal direction, and

(b) applying a modified threshold to the transform coefficients,

resulting in significant gains over the state-of-the-art denoising techniques (please see the results section).

Method

The **wavelet thresholding** approach works in three steps: taking the discrete wavelet transform of a noisy signal, thresholding the wavelet coefficients, and taking the inverse discrete wavelet transform to estimate the original signal.

1. *Discrete 3D Wavelet Packet Transform* :
The discrete wavelet packet transform (DWPT) of a 1D signal x of length N can be computed as follows

$$\begin{aligned} w_{2n,d,l} &= \sum_k g_{k-2l}\, w_{n,d-1,k} & l &= 0,1,\ldots,N2^{-d}-1 \\ w_{2n+1,d,l} &= \sum_k h_{k-2l}\, w_{n,d-1,k} & l &= 0,1,\ldots,N2^{-d}-1 \\ w_{0,0,l} &= x_l & l &= 0,1,\ldots,N-1 \end{aligned}$$

where $d = 1,2,\ldots,J-1$ is the scale index, with $J = \log_2 N$, n and l respectively denote the frequency and position indices, $\{h_n\}$ and $\{g_n\}$ correspond to the lowpass and highpass filters respectively for a two-channel filter bank.

The 3D DWPT can be computed by applying above equations separably in all three directions to get the FWP decomposition up to the coarsest resolution of subbands. The best basis can be selected in $O(N \log N)$ time, where N denotes the number of samples (frame resolution times the number of frames) in the video sequence.

2. *Modified BayesShrink Thresholding* :
A modified BayesShrink [3] method is used to compute the optimal value of threshold adaptively for each subband. Threshold θ_b for a subband of length N in an L-level WP decomposition is given by

$$\theta_b = \sqrt{\log N/L}\left(\frac{\sigma^2}{\sqrt{\max(\sigma_b^2 - \sigma^2, 0)}}\right)$$

where σ_b^2 is the subband variance, and σ^2 is the noise variance. If σ^2 is not known, a robust median estimate for noise standard deviation $\hat{\sigma}$ is obtained as follows

$$\hat{\sigma} = \mathcal{E}\{\hat{\Sigma}\}, \quad \hat{\sigma}_i = \frac{Median(|Y_i|)}{0.6745}$$

where $\hat{\sigma}_i \in \hat{\Sigma}$, $Y_i \in \{\mathcal{Y}\}$, set of all HHH bands in the decomposition tree, and the mean $\mathcal{E}$ is taken only on the smaller half of the sorted $\hat{\Sigma}$ excluding the smallest value.

3. *Inverse 3D Wavelet Packet Transform* :
The transform is invertible if the geometry of 3D basis used in forward DWPT is known and appropriate dual filters $\{\tilde{h}_n\}$, $\{\tilde{g}_n\}$ are used on the synthesis side.

Experimental Results

The above algorithm was tested against a number of other algorithms for restoration of several standard video sequences, three of which are included here: *Miss America*, *Hall*, and *Football*, all at a resolution of 128^3. The video sequences were corrupted with additive white Gaussian noise, with the SNR of the noisy sequences being 0dB, 5dB, and 10dB. Comparative SNR curves for individual frames for the test sequences are provided in Figure 1.

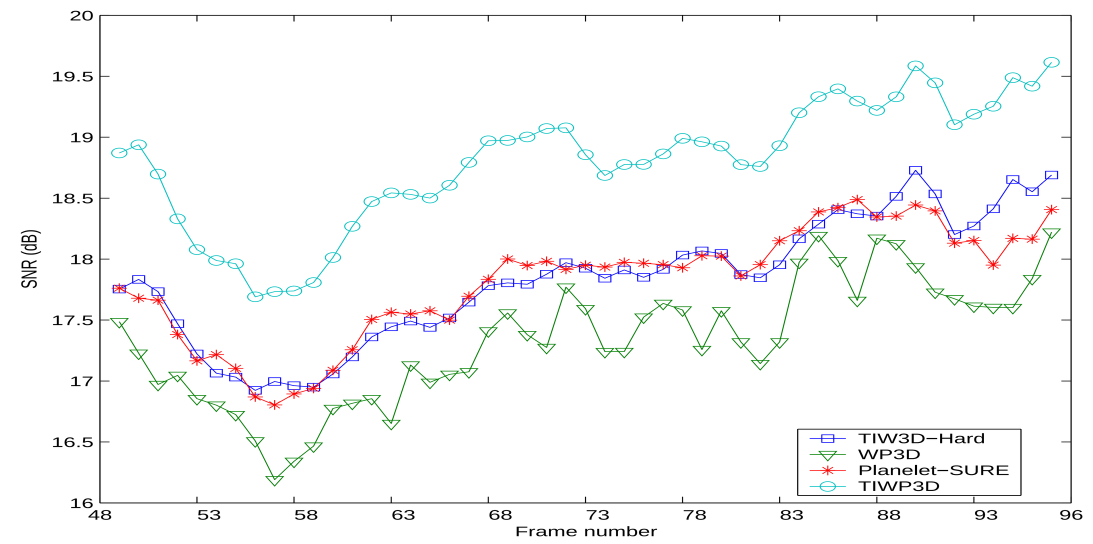
Comparative results for following algorithms are provided. translation-invariant (TI) hard thresholding in 3D wavelet domain (**TIW3D**), 3D wavelet packet (**WP3D**) with modified *BayesShrink*[3] as described in the previous section, non-separable planelet [4] domain thresholding using SUREShrink method, and TI 3D wavelet packet (TIWP3D) with the modified *BayesShrink*. For comparison purposes, computational complexity for each of the algorithms considered is also provided in Table 1.

Discussion: Our experimental results show that the proposed algorithm outperforms the state-of-the-art in terms of both SNR and visual quality. In terms of computational complexity, the planelet algorithm of [4] is the least computationally expensive, whereas the TI implementations of 3D wavelet and 3D WP are towards the more expensive side with TIWP3D being the most expensive due to the additional one-off cost of best basis selection.

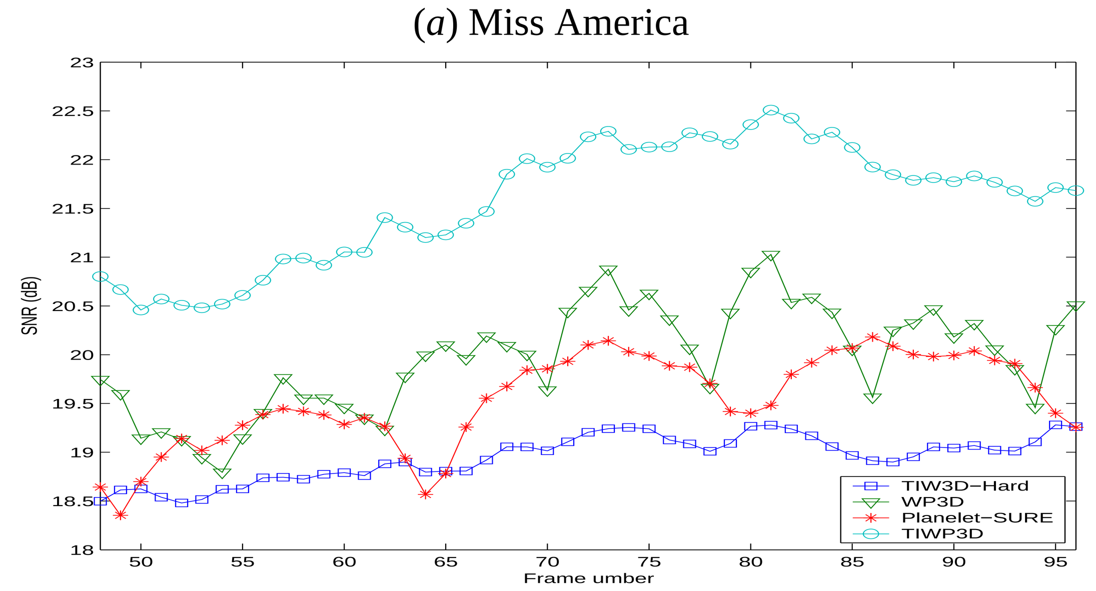
TIW3D-Hard	WP3D	Planelet-SURE [4]	TIWP3D
$O(N + l^3 N)$	$O(N \log N)$	$O(n)$	$O(N \log N + l^3 N)$

TABLE 1. Computational complexity of the tested algorithms

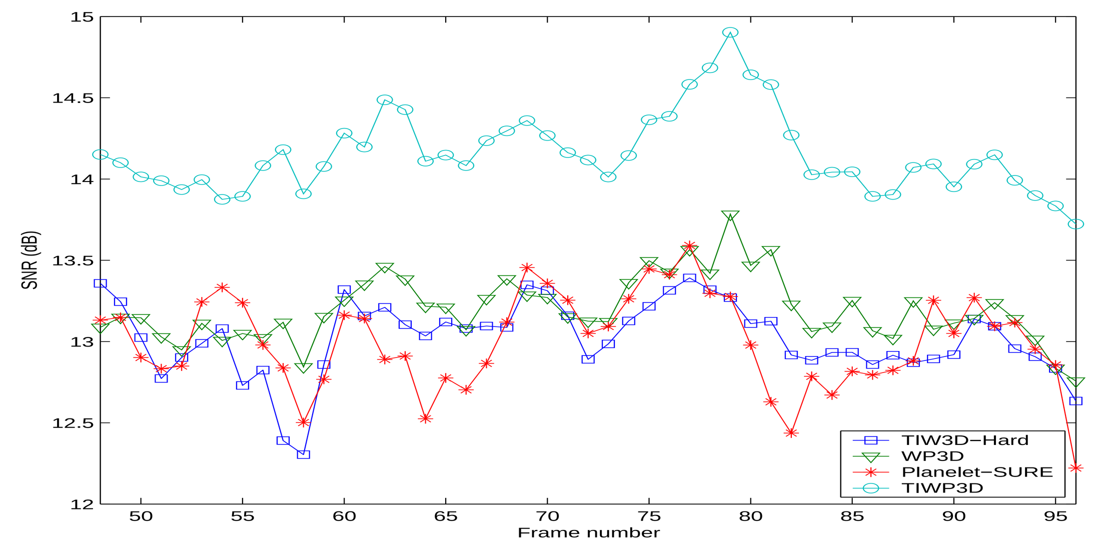
N and n respectively denote sequence size and planelet window size, and l denotes length of the wavelet filter.



(*a*) Miss America

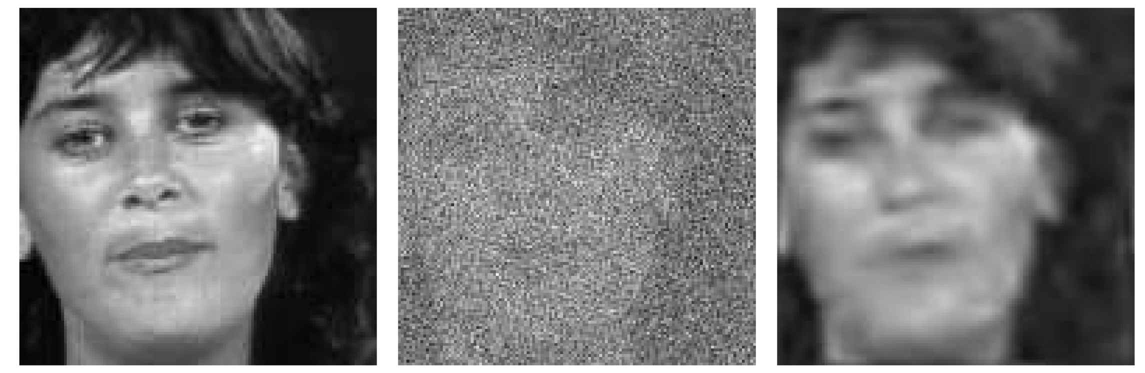


(*b*) Hall



(*c*) Football

FIGURE 1. *Frame-by-frame denoising results*



(*a*) Original (*a*) Noisy (0dB) (*a*) TIWP3D (18.9dB)

(*a*) Original (*a*) Noisy (10dB) (*a*) TIWP3D (21.3dB)

(*a*) Original (*a*) Noisy (5dB) (*a*) TIWP3D (14.3dB)

FIGURE 2. *Visual denoising results*

Summary

Conclusions:

- Our results show that adaptive wavelet packets in 3D are well suited to represent pre-dominant spatiotemporal structures in video sequences, even in the presence of noise.

Future Directions:

- Simultaneous coding and denoising of video sequences
- Solution for high computational/memory complexity

References

[1] R.R. Coifman and D.L. Donoho, "Translation-invariant denoising," in *Wavelets and Statistics. Lecture Notes in Statistics*, 1995.

[2] I.W. Selesnick and K.Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Proc. SPIE Wavelets X*, August 2003.

[3] G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. on Image Processing*, vol. 9, September 2000 .

[4] N.M. Rajpoot, R.G. Wilson, and Z. Yao, "Planelets: A new analysis tool for planar feature extraction," in *Proc. 5th Intl. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, April 2004.