ADAPTIVE WAVELET RESTORATION OF NOISY VIDEO SEQUENCES

Nasir Rajpoot *Zhen Yao* *Roland Wilson*

Department of Computer Science
University of Warwick
Coventry CV4 7AL, UK
email: nasir,yao,rgw@dcs.warwick.ac.uk

ABSTRACT

In this paper, we present a novel algorithm for restoration of noisy video sequences. A video sequence is first transformed into an optimal 3D wavelet domain using basis functions adapted to the contents of the sequence. Assuming that all the major spatiotemporal frequency phenomena present in the sequence produce high amplitude transform coefficients, a modified form of the BayesShrink thresholding method is used to suppress the noise. In order to reduce the effects of Gibbs phenomenon in the restored sequence, translation dependence is removed by averaging the restored instances of the shifted sequence. The algorithm yields promising results in terms of both objective and subjective quality of the restored sequence.

1. 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* [1, 2], has become increasingly popular in recent years. 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. Two methods are commonly used: *hard* and *soft* thresholding. In hard thresholding, all wavelet coefficients below a threshold are set to zero, assuming that the suppressed coefficients were the only ones most affected by noise. In the case of *soft* thresholding, all coefficients below a chosen threshold are set to zero, and the magnitude of the remaining coefficients is decreased by the threshold value, the assumption being that all coefficients were affected by noise. 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. This general approach to signal restoration can also be applied to the removal of noise in a noisy video sequence by thresholding the coefficients of three-dimensional (3D) spatiotemporal wavelet transform of the sequence. The case for thresholding in spatiotemporal 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. Recent attempts to solve the video restoration problem have included combined spatial and temporal wavelet denoising [3], and the use of thresholding in non-separable transform domains, such as oriented 3D ridgelets [4] and 3D complex wavelets [5].

Although wavelet shrinkage performs significantly better than most other commonly used denoising methods, visual quality of the restored image (or video) can sometimes suffer from *ringing* type of artifacts, vallyes around the edges, due to the Gibbs phenomenon. The shift-variant nature of the wavelet transform worsens the effect of the Gibbs phenomenon, resulting in unpleasant artifacts. Translation invariant (TI) wavelet denoising of Coifman and Donoho [6] was developed to counter such artifacts by averaging out the translation dependence. Another feature of wavelet denoising is that it imposes a fixed dyadic wavelet basis on all types of input signals. Not only can the use of dyadic wavelets result in a blurred reconstruction, it can also limit the analysis of a locally occurring phenomenon in the frequency domain. The solution to this problem lies in the use of basis functions which are well localized in frequency as well as time (for 1D signals in time). For a video sequence, basis functions (separable or otherwise) with good localization in both space-time and spatiotemporal frequency are sought. In another relevant paper [7], we have developed a non-separable 3D representation termed as the *planelet* basis and studied its application to the video denoising problem. In this paper, we present a novel algorithm based on the 3D extension of translation invariant denoising using an adaptive wavelet packet representation for restoration of noisy video sequences. Comparative results show that our algorithm achieves significant gains over the state-of-the-art denoising techniques in terms of both SNR and visual quality of the restored video sequence for all three standard test video sequences at different levels of noise.

2. WAVELET PACKETS IN 3D

The ability of wavelet packets to capture locally occurring frequency phenomena in a signal has led to their successful application to many problems including image coding [8, 9]. The fundamental idea is to relax the restricted decomposition of only the lowpass subband and allow the exploration of all frequency bands up to the maximum depth. 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 and the transform is invertible if appropriate dual filters $\{\tilde{h}_n\}$, $\{\tilde{g}_n\}$ are used on the synthesis side. These equations can be used to compute full wavelet packet (FWP) tree of the signal decomposition. However, this implies that a large number of combinations of basis functions is now available to completely represent the signal. A tree-prunning approach such as [10] can be used to efficiently select the *best basis* with respect to a cost function.

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. Given the goal here is to capture the significant spatiotemporal frequency phenomena in a video sequence, we used the Coifman-Wickerhauser entropy [10] as a cost function to select the best basis.

3. THE RESTORATION ALGORITHMS

The effect of the Gibbs phenomenon can be weakened by averaging the restored signal over a range of circular shifts [6]. For this reason, we apply soft threhsolding to the 3D wavelet packet coefficients of the shifted (in all three directions) noisy video sequence. A modified BayesShrink [11] method is using 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.

4. 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 corruped with additive white Gaussian noise, with the SNR of the noisy sequences being 0dB, 5dB, 10dB, and 15dB. Table 1 gives denoising results in terms of SNR for these noisy sequences using the following algorithms: TI soft thresholding in 2D wavelet domain (TIW2D), TI soft and hard thresholding in 3D wavelet domain (TIW3D), 3D wavelet packet (WP3D) with BayesShrink [11], both non-TI and TI 3D wavelet packet (TIWP3D) with the modified form of BayesShrink described in the previous section, and non-separable planelet [7] domain thresholding using SUREShrink [2] method. Comparative SNR curves for individual frames for two of the test sequences are provided in Figure 1. For all our experiments, the proposed algorithm produces by far the best results in terms of both overall and individual SNR. Some of the frames of the test sequences restored by our algorithm and TIW3D-Hard, a 3D realization of the algorithm in [6], are shown in Figure 2. While TIW3D restores clean and smooth version of the original frames, some of the details are restored by TIWP3D.

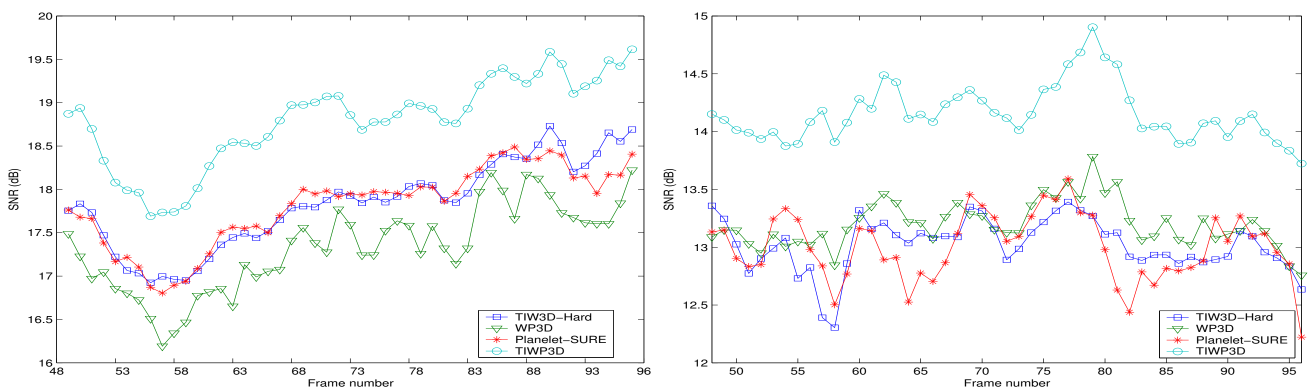
For comparison purposes, computational complexity for each of the algorithms considered is also provided in Table 1. It is clear from this table that the planelet algorithm of [7] is the least computationally expensive, whereas the TI implementations of 3D wavelet and 3D WP are towards the more expensive side with TIWP3D being the most expensive due to the additional one-off cost of best basis selection.

5. CONCLUSIONS

In this paper, a novel algorithm for restoration of noisy video sequences was presented. The algorithm works by finding an optimal wavelet packet representation for a sequence and averaging the results of thresholding by a modified form of BayesShrink [11] method on shifted transform coefficients. Experimental results suggest that the performance of the algorithm is by far the best as compared to other methods found in the literature, to the best of our knowledge. Although its computational cost may be a limiting factor in some applications, a less expensive version of the algorithm

Video Sequence	Noise (dB)	Denoising Algorithm (Transform+Thresholding)						
		TIW2D Soft	TIW3D Soft	TIW3D Hard	WP3D Bayes	WP3D Proposed	TIWP3D Proposed	Planelet SURE
Miss America	0	9.5	12.4	17.9	17.2	17.4	**18.9**	17.3
	5	12.6	15.6	19.5	19.0	19.3	**20.7**	19.6
	10	15.3	18.1	21.5	21.1	21.5	**23.0**	21.5
	15	18.1	20.3	23.9	23.1	23.8	**25.2**	23.5
Hall	0	5.3	11.0	14.7	14.8	15.0	**16.7**	14.8
	5	9.4	13.4	16.6	17.2	17.3	**18.9**	17.2
	10	11.7	15.5	19.0	19.5	19.8	**21.3**	19.5
	15	14.2	17.7	21.7	22.1	22.5	**24.1**	21.8
Football	0	5.9	9.4	11.9	11.9	12.0	**12.8**	12.1
	5	8.5	11.1	13.1	12.9	13.3	**14.3**	13.2
	10	11.0	12.5	15.0	13.9	15.2	**16.9**	14.7
	15	13.4	14.3	18.0	15.3	17.9	**20.0**	16.6
Computational Complexity	–	$O(N+l^2N)$	$O(N+l^3N)$	$O(N++l^3N)$	$O(N\log N)$	$O(N\log N)$	$O(N\log N++l^3N)$	$O(n)$

Table 1. SNR results for three standard video sequences and algorithms' complexity
N and n respectively denote sequece size and planelet window size, and l denotes length of the wavelet filter.



(a) (b)

Fig. 1. Frame-by-frame comparative results (a) *Miss America* and (b) *Football*

(WP3D) with the proposed thresholding method produces results comparable to the state-of-the-art at a lower cost. It is perhaps worth noting that while being the least expensive, the planelet method [7] produces SNR results which are still comparable to those of TIW3D-Hard. These results suggest that the localization of spatiotemporal frequency is a desirable feature of the domain in which video sequences are represented.

6. REFERENCES

[1] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

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

[3] A. Pizurica, V. Zlokolika, and W. Philips, "Combined wavelet domain and temporal video denoising," in *Proc. IEEE Intl. Conf. on Advanced Video and Signal based Surveillance (AVSS)*, July 2003.

[4] P. Carre, D. Helbert, and E. Andres, "3-D fast ridgelet transform," in *Proc. IEEE Intl. Conf. on Image Processing (ICIP)*, September 2003.

[5] 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.

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

[7] N.M. Rajpoot, R.G. Wilson, and Z. Yao, "Planelets: A new analysis tool for planar feature extraction," in

Fig. 2. Denoising results for three standard video sequences

(*a*) Frame# 90 of *Miss America* (*b*) Noisy (SNR=0dB) (*c*) TIW3D-Hard (SNR=17.9dB) (*d*) TIWP3D (SNR=18.9dB)
(*e*) Frame# 106 of *Hall* (*f*) Noisy (SNR=10dB) (*g*) TIW3D-Hard (SNR=19.0dB) (*h*) TIWP3D (SNR=21.3dB)
(*i*) Frame# 60 of *Football* (*j*) Noisy (SNR=5dB) (*k*) TIW3D-Hard (SNR=13.1dB) (*l*) TIWP3D (SNR=14.3dB)

Proc. 5th Intl. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), April 2004.

[8] F.G. Meyer, A.Z. Averbuch, and J-O. Strömberg, "Fast adaptive wavelet packet image compression," *IEEE Trans. on Image Processing*, vol. 9, May 2000.

[9] N.M. Rajpoot, R.G. Wilson, F.G. Meyer, and R.R. Coifman, "Adaptive wavelet packet basis selection for zerotree image coding," *IEEE Trans. on Image Processing*, vol. 12, December 2003.

[10] R.R. Coifman and M.V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. on Info. Th.*, vol. 38, March 1992.

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