Motion-Compensated Wavelet Video Denoising

Fu Jin, Paul Fieguth, and Lowell Winger

Dept. of Systems Design Engineering, Univ. of Waterloo
Waterloo, Ontario, Canada N2L 3G1
{fjin, pfieguth,lwinger}@uwaterloo.ca

Abstract. We study a low complexity, motion-compensated recursive video denoising scheme in the wavelet domain. To preserve *weak* spatial textures, both spatial and temporal filtering are carried out in the wavelet domain. We justify the proposed approach by evaluating the relative increase of mean squared error (MSE) with respect to the optimal estimator. Various multiresolution motion estimation schemes are studied to allow us to exploit the high temporal correlation present in most video. Kalman filtering is then applied to the wavelet coefficients along motion trajectories to efficiently suppress noise.

We show experimentally that our wavelet-based recursive denoising compares favorably with other wavelet-based denoising approaches. Specifically, we can preserve both strong and weak spatial details while removing noise.

Keywords: Video denoising, motion compensation, wavelet transform, Kalman filtering.

1 Introduction

With the maturity of digital video capturing devices and broadband transmission networks, numerous applications have been emerging. These include teleconferencing, remote surveillance, multimedia services and digital television, to name a few. However, the video signal is almost always corrupted by noise from the capturing devices or during transmission due to random thermal or other electronic noises. Usually, noise reduction can considerably improve visual quality and benefit the subsequent processing tasks, such as video compression.

There are many existing video denoising approaches in the spatial domain [1,2,4,9], which can roughly be divided into two classes:

Temporal-only: A temporal-only approach utilizes only the temporal correlations, neglecting spatial information. Since video signals are strongly correlated along motion trajectories, motion estimation/compensation is normally employed. In those cases where motion estimation is not accurate, motion detection may be used to avoid blurring. These techniques can preserve spatial details well, but the resulting images usually still contain removable noise since spatial correlations are neglected.

A. Campilho, M. Kamel (Eds.): ICIAR 2004, LNCS 3211, pp. 572–579, 2004.

Spatio-temporal: More sophisticated methods exploit both spatial and temporal correlations, such as simple adaptive weighted local averaging [3], *3-D* order-statistic algorithms [2], *3-D* Kalman filtering [4] and *3-D* Markov random models [5].

Although there have been many papers addressing the application of wavelet transforms to image denoising, comparatively few have addressed wavelet-based video denoising. Roosmalen *et. al.* [6] proposed video denoising by thresholding the coefficients of a specific *3-D* wavelet representation and Selesnick *et. al.* [7] found an *efficient 3-D* orientation-selective wavelet transform, *3-D complex* wavelet transforms, which avoided the time-consuming motion estimation process. The main drawbacks of the *3-D* wavelet transforms include a long time latency and the inability to adapt to fast motions.

In most video processing applications a long latency is unacceptable, so recursive approaches are widely employed. Pizurica *et. al.* [8] proposed sequential *2-D* spatial and *1-D* temporal denoising, in which they first do sophisticated wavelet-based image denoising for each frame and then recursive *temporal* averaging. However, *2-D* spatial filtering tends to introduce artifacts and to remove *weak* details along with the noise.

In this paper, we propose wavelet-domain recursive video denoising. We use a *2-D* spatial wavelet, and filter recursively to preserve low latency. But unlike [8], our temporal filtering works on the wavelet coefficients themselves (instead of on the spatial pixels). In such a way we minimize spatial blurring to preserve the *weak* spatial details in still areas. The key to processing directly in the wavelet domain is an efficient shift-invariant transform, allowing spatial motion to be meaningfully reflected in the wavelet coefficients. We study several schemes for robust wavelet-domain motion estimation and then proceed to motion-compensated temporal filtering in the *wavelet domain* using an adaptive Kalman filter.

2 Wavelet-Based Video Denoising

In standard wavelet-based image denoising [9] the *2-D* wavelet transform is used to get a compact representation. Thus it would seem natural to select *3-D* wavelets for video denoising [6,7]. However, there are compelling reasons to choose a *2-D* spatial wavelet transform with recursive temporal filtering for video denoising:

1. There is a clear asymmetry between the spatial and temporal axes, in terms of correlation and resolution. A recursive approach is naturally suited to this asymmetry, whereas a *3-D* wavelet is not.
2. Recursive filtering can significantly reduce time delay and memory requirements.
3. For autoregressive models the optimal estimator can be achieved recursively.
4. Motion information can be efficiently exploited with recursive filtering.

A. Problem Formulation

The video denoising problem can be modeled as follows: given video measurements y with spatial indices i, j and temporal index k

$$y(i,j,k) = x(i,j,k) + v(i,j,k), \;\; i,j = 1,2,...,N, \; k = 1,2,...,M \tag{1}$$

corrupted by i.i.d Gaussian noise v, we need to estimate the true image sequence x. Define $\mathbf{x}(k)$, $\mathbf{y}(k)$ and $\mathbf{v}(k)$ to be the column-stacked images at time k, then (1) becomes

$$\mathbf{y}(k) = \mathbf{x}(k) + \mathbf{v}(k), \; k = 1,2,...,M \tag{2}$$

We propose to do denoising in the wavelet domain; let H be the *2-D* wavelet transform matrix, then (2) is transformed as

$$H\mathbf{y}(k) = H\mathbf{x}(k) + H\mathbf{v}(k) \tag{3}$$

Denoting wavelet coefficient vector as $\mathbf{y}_H(k)$, (3) is rewritten as

$$\mathbf{y}_H(k) = \mathbf{x}_H(k) + \mathbf{v}_H(k) \tag{4}$$

Since we seek a recursive temporal filter we assert an autoregressive form for the signal model

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + B\mathbf{w}(k+1) \tag{5}$$

thus

$$\mathbf{x}_H(k+1) = A_H \cdot \mathbf{x}_H(k) + B_H \cdot \mathbf{w}_H(k+1) \tag{6}$$

where $A_H \equiv HAH^{-1}, \;\; B_H \equiv HBH^{-1}$. It should be noted the wavelet domain state model still has an autoregressive form. Therefore, *optimal* filtering can be achieved in a recursive way in the wavelet domain.

B. An Example: Recursive Image Filtering in the Spatial and Wavelet Domains

As a quick proof of principle, we can denoise *2-D* images using a recursive *1-D* wavelet procedure, analogous to denoising *3-D* video using *2-D* wavelets. We do not propose this as a superior approach to image denoising, rather as a measure of promise in the video case. We use an autoregressive image model and do *1-D* wavelet transform on each column, followed by recursive filtering column by column. We assess estimator performance in the sense of relative increase of MSE:

$$\delta_{MSE} = \frac{MSE - MSE_{optimal}}{MSE_{optimal}} \tag{7}$$

where $MSE_{optimal}$ is the MSE of the optimal Kalman filter.

Table 1. Percentage increase δ_{MSE} in estimation error relative to the optimal estimator. Estimating wavelet coefficients independently introduces only slight error.

SNR(dB)	10	0	-10
δ_{MSE} (spatial)	1.88%	14.10%	18.65%
δ_{MSE} (wavelet)	0.1%	1.09%	6.51%

Table 2. Comparison of PSNR (in dB) of the proposed method and *3-D* complex DWT for sequence *Paris*.

PSNR (Original)	28.2	22.1	18.6	16.1
PSNR (Proposed method)	37	30	27	26
PSNR (*3-D* complex DWT [7])	35	28	24	23

We use a common image model

$$x(i,j) = \rho_v x(i-1,j) + \rho_h x(i,j-1) - \rho_v \rho_h x(i-1,j-1) + w(i,j), \quad \rho_h = \rho_v = 0.95 \tag{8}$$

which is a causal MRF model and can be converted to a vector autoregressive model [10].

The optimal recursive filtering requires the joint processing of entire columns. As this would be completely impractical in the video case, for reasons of computational complexity we recursively filter the wavelet coefficients *independently*, ignoring inter-coefficient relationships. As shown in Table 1, scalar processing in the wavelet domain leads to only very moderate increases in MSE relative to the optimum, whereas this is not at all the case in the spatial domain. It should be noted that the wavelet-based scalar processor is comparable to the optimal filter when $SNR > 0dB$, a condition satisfied in many practical applications.

3 The Denoising System

The success of *1-D* wavelet image denoising motivates the extension to the video case. There are three crucial aspects: (1) the choice of *2-D* wavelet transform, (2) wavelet-domain motion estimation, and (3) the recursive filtering applied to the motion-compensated wavelet coefficients. These steps are detailed below.

2-D **wavelet transform**: We apply a *2-D* wavelet transform to each frame, rather than a *3-D* transform for the whole image sequence, and the coefficients are then filtered recursively.

A huge number of wavelet transforms have been developed (e.g., orthogonal / non-orthogonal, real-valued/complex-valued, decimated / redundant). For image denoising problems, three criteria are desired:

1. Shift invariance: to suppress frequency aliasing and related artifacts;
2. Direction selectivity: of importance when image has dominant oriented features;
3. Low complexity.

The *2-D* dual-tree complex wavelet proposed by Kingbury [7] satisfies these requirements very well. Unfortunately it is not convenient for motion estimation since the motion information is related to the coefficient phase, which is a non-linear function of translation. Alternatively, specially designed *2-D* wavelet transforms (e.g., curvelet, contourlet) are sensitive to feature directions but are too complex for computation. In this paper, we choose to use an over-complete wavelet representation proposed by Mallat *et. al.* [11]. Although this wavelet representation does not have very good directional selectivity (e.g., it mixes 45-degree and −45-degree features), several researchers have used it for natural image denoising with impressive results (e.g.[8]).

Multiresolution Motion Estimation (MRME): We utilize the well-used block matching (BM) technique for motion estimation. Compared with other motion estimation approaches, such as optical flow and pixel-recursive methods, block matching is straightforward to compute and less sensitive to noise.

Single-resolution block matching approaches have been studied extensively [1] and are successfully used in modern video compression standards. Multi-resolution block matching (MRBM) was first proposed by Zhang *et. al.* [1] for wavelet-based video coding, and recent developments can be found in [12]. The basic idea of standard MRBM is to start block matching at the coarsest level, using this estimate as a prediction for the next finer scale. Oddly, a critically decimated wavelet was used [12], which implies that the motion is not a constant function of scale. A much more sensible choice of wavelet, used in this paper, is the overcomplete transform which is shift-invariant, leading to consistent motion as a function of scale, except in the vicinity of motion boundaries. Clearly, this high inter-scale relationship of motion should be exploited to improve accuracy. We developed a sequence of four approaches of increasing accuracy:

1. The standard MRME scheme [12].

2. Block matching separately on each level, combined by median filtering.

3. Joint block matching simultaneously at all levels:

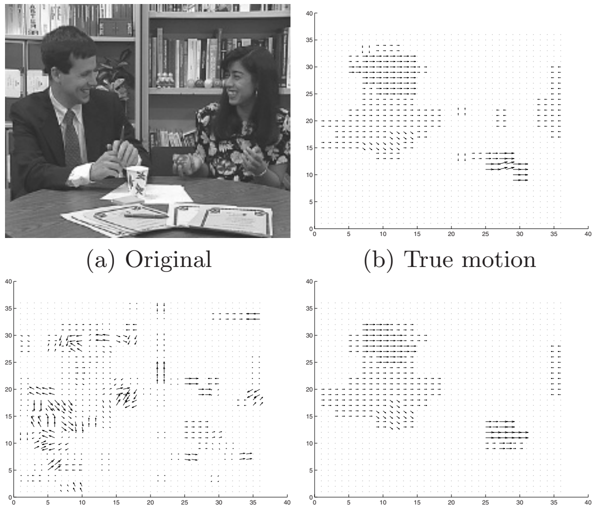
Let $\epsilon^l(\mathbf{i}, \mathbf{v})$ denote the displaced frame difference (DFD) at position $\mathbf{i}$ of level l with displacement $\mathbf{v}$). Then the total DFD over all levels is defined as

$$\epsilon(\mathbf{i}, \mathbf{v}) = \sum_{l=1}^{J} \epsilon^l(\mathbf{i}, \mathbf{v}) \tag{9}$$

and the displacement field $\mathbf{v}(\mathbf{i}) = [v_x(\mathbf{i}), v_y(\mathbf{i})]$ is found by minimizing $\epsilon(\mathbf{i}, \mathbf{v})$.

4. Block matching with smoothness constraint:

The above schemes did not assert any spatial smoothness or correlation in the motion vectors, which we expect in real-world sequences. This is of considerable importance when the additive noise levels are large, leading to irregular estimated motion vectors. Therefore, we introduce an additional



(a) Original (b) True motion

(c) Standard (d) Smoothness-constrained

Fig. 1. Compare MRBM results of the standard approach (c) and the proposed one (d). The standard does not exploit the smoothness property (*a prior* knowledge) of motion fields. Therefore, it performs poorly in the presence of noise. For sharp comparison, our proposed approach give much better results.

smoothness constraint and perform BM by solving the optimization problem

$$arg\min_{\mathbf{v}} \left\{ \sum_{\mathbf{i}} \left[\gamma \cdot \epsilon(\mathbf{i}, \mathbf{v}) + (|v_x(\mathbf{i}) - \frac{1}{M} \sum_{m \in N_b(\mathbf{i})} v_x(\mathbf{i} + \mathbf{m})| \right.\right.$$
$$\left.\left. + |v_y(\mathbf{i}) - \frac{1}{M} \sum_{n \in N_b(\mathbf{i})} v_y(\mathbf{i} + \mathbf{n})|) \right] \right\} \qquad (10)$$

where N_b is the neighborhood set of the element $\mathbf{i}$ and M is the number of elements in N_b. $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N]^T$ is the motion field to be estimated. γ controls the tradeoff between $\epsilon(\mathbf{i}, \mathbf{v})$ (see (9)) and smoothness.

Experimentally, we have found Approach 4 to be the most robust to noise and yield reasonable motion estimates. In the simulation results given below we will use this latter approach for motion estimation.

Wavelet coefficient filtering: The key to our approach is to support both spatial and temporal filtering, as appropriate. Specifically, when the motion information is unambiguous, i.e.,

$$|x_H(m, n, k) - \rho(x_H(m, n, k-1))| < 2\sqrt{2}\sigma_{v_H}, \qquad (11)$$

where $\rho(\cdot)$ is the motion compensation function, we restrict the filtering to be purely temporal to avoid any spatial blurring. However, when motion estimates are poor, i.e.,

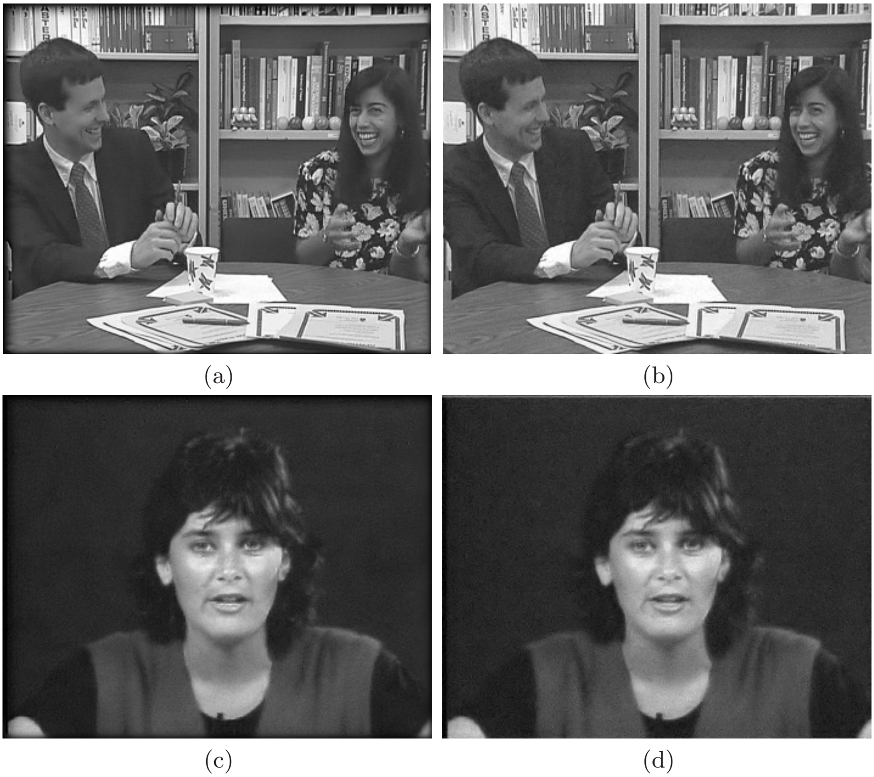


Fig. 2. Denoised images : (a) & (c) the proposed method (b) & (d) *3-D* complex DWT. $\sigma_{v_H} = 10$ for *Paris*, $\sigma_{v_H} = 20$ for *MissAmerica*.

$$|x_H(m, n, k) - \rho(x_H(m, n, k-1))| \geq 2\sqrt{2}\sigma_{v_H}, \qquad (12)$$

we apply wavelet shrinkage to exploit spatial correlation. Although this introduces some blurring, wavelet shrinkage is state-of-the-art in static image denoising.

4 Experimental Results

We have tested all four of the MRBM approaches listed in the previous section. Due to space limit we show only the results of standard MRBM (hierarchical prediction-updating approach) and our proposed smoothness-constrained motion estimation in Fig. 1. It is clear that MRBM becomes much less sensitive to noise if the smoothness constraint is applied, so we propose to use this method for the subsequent video denoising tests.

To show denoising performance of the proposed method we have experimented on several standard video sequences (e.g., *Paris, Salesman, MissAmer-*

ica). We compare the denoising results of the proposed method with those from the *3-D* complex DWT [7] (http://taco.poly.edu/WaveletSoftware/), in terms of Peak SNR (PSNR) in Table 2, and our method gives a consistent improvement in performance. Several denoised images are shown in Fig. 2, where a careful examination of spatial textures (in the face and table top) shows the supperiority of our proposed approach. In particular, there are fewer aliasing artifacts, and stationary objects (such as the bookshelf) are denoised much more crispy.

References

1. Brailain J., Kleihorst R., Efstratiadis S., Katsaggelos A. and Lagendijk A.: Noise reduction filters for dynamic image sequences: A review. Proc. of IEEE **83**(9) (1995) 1272–1292
2. Arce G: Multistage order statistic filters for image sequence processing. IEEE Trans. Acoust., Speech, Signal Proc. **39**(5) (1991) 1147–1163
3. Ozkan M. and Sezan I. and Tekalp A.: Adaptive motion-compensated filtering. IEEE Trans. on CSVT.,**3**(8) (1993) 277–290
4. Kim J. and Woods J.: Spatio-temporal adaptive *3-D* Kalman filter for video. IEEE Trans. Imag Proc.,**6**(3) (1997) 414–424
5. Brailean J. and Katsaggelos A.: Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences. IEEE Trans. Imag Proc.,**4**(9) (1995) 1236–1251
6. Roosmalen P., Westen S., Lagendijk R. and Biemond J.: Noise reduction for image sequences using an oriented pyramid thresholding technique. Proc. Intern. Conf. Image Proc., (1996) 375–378
7. Selesnick I and Li K.: Video denoising using *2-D* and *3-D* dual-tree complex wavelet transforms. Proc. SPIE Wavelet Applications in Signal and Image Processing X,2003,700–707
8. Pizurica A., Zlokolica V. and Philips W.: Combined wavelet domain and temporal video denoising. Proc. Advanced Video and Signal Based Surveillance,2003,334–341
9. Chang S., Bin Y. and Vetterli M.: Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Trans. Image Proc., **9**(9), (2000) 1522–1531
10. Rosenfeld A. and Kak A.: Digital picture processing. Acadamic Press (1982)
11. Mallat S. and Zhong S.: Characterization of signals from multiscale edges. IEEE Trans. PAMI, **14**(9),(1992) 710–732
12. Zan J, Ahmad M.O. and Swamy S.: New techniques for multi-resolution motion estimation. IEEE Trans. CSVT, **12**(9),(2002) 793–802
13. Lee J.:Digital image enhancement and noise filtering by use of local statistics. IEEE Trans. PAMI, **2**(6),(1980) 165–168