

Wavelet Video Denoising with Regularized Multiresolution Motion Estimation

Fu Jin, Paul Fieguth, and Lowell Winger

Department of Systems Design Engineering, Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1

Received 1 September 2004; Revised 23 June 2005; Accepted 30 June 2005

This paper develops a new approach to video denoising, in which motion estimation/compensation, temporal filtering, and spatial smoothing are *all* undertaken in the wavelet domain. The key to making this possible is the use of a shift-invariant, overcomplete wavelet transform, which allows motion between image frames to be manifested as an equivalent motion of coefficients in the wavelet domain. Our focus is on minimizing spatial blurring, restricting to temporal filtering when motion estimates are reliable, and spatially shrinking only insignificant coefficients when the motion is unreliable. Tests on standard video sequences show that our results yield comparable PSNR to the state of the art in the literature, but with considerably improved preservation of fine spatial details.

Copyright © 2006 Fu Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

With the maturity of digital video capturing devices and broadband transmission networks, many video applications have been emerging, such as teleconferencing, remote surveillance, multimedia services, and digital television. 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 facilitate the subsequent processing tasks, such as video compression.

There are many existing video denoising approaches in the spatial domain [1–4], which can roughly be divided into two or three classes.

Temporal-only

An approach utilizes only the temporal correlations [1], 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 [1, 5] 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.

Spatio-temporal

More sophisticated methods exploit both spatial and temporal correlations, such as simple adaptive weighted local averaging [6], 3D order-statistic algorithms [2], 3D Kalman filtering [3], and 3D Markov models [7]. However, due to the high structural complexity of natural image sequences, accurate modeling remains an open research problem.

Spatial-only, a third alternative, would apply 2D spatial denoising to each video frame, taking advantage of the vast image denoising literature. Work in this direction shows limited success, however, because 2D denoising blurs spatial details, and because a spatial-only approach ignores the strong temporal correlations present in video.

Recently, many wavelet-based image denoising approaches have been proposed with impressive results [4, 8–10]. However, it is interesting to note that although there have been many papers addressing wavelet-based image denoising, comparatively few have addressed wavelet-based *video* denoising. Roosmalen et al. [11] proposed video denoising by thresholding the coefficients of a specific 3D wavelet representation and Selesnick and Li [12] employed an efficient 3D orientation-selective wavelet transform, 3D *complex* wavelet transforms, which avoids the time-consuming motion estimation process. The main drawbacks of the 3D 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. [5] proposed sequential 2D spatial and 1D temporal denoisings, in which they first do sophisticated wavelet-based image denoising for each frame and then recursive *temporal* averaging. However, 2D spatial filtering tends to introduce artifacts and to remove weak details along with the noise. Due to difficulties in estimating motion in noise, only simple motion detection was used in [5] to utilize temporal correlation between frames.

Given the strong decorrelative properties of the wavelet transform and its effectiveness in image denoising, we are highly motivated to consider spatial-temporal video filtering, but entirely in the wavelet domain. That is, to maintain low latency, we employ a frame-by-frame recursive temporal filter but, unlike [5], perform all filtering in the wavelet domain. However for wavelet-domain motion estimation/compensation to be possible, predictable image motion must correspond to a predictable motion of the corresponding wavelet coefficients. The key, therefore, to wavelet-based video denoising is an efficient, shift-invariant, overcomplete wavelet transform. The benefits of such an approach are clear.

- (1) The recursive, frame-by-frame approach implies low latency.
- (2) The wavelet decorrelative property allows very simple, scalar temporal filtering.
- (3) Where motion estimates are unreliable, wavelet shrinkage can provide powerful denoising.

The remaining challenges, the design of a robust approach to wavelet motion estimation and the selection of a particular spatial-temporal denoising scheme, are studied in this paper.

2. WAVELET-BASED VIDEO DENOISING

In standard wavelet-based image denoising [4], a 2D wavelet transform is used because it leads to a sparse, efficient representation of images, thus it would seem natural to select 3D wavelets for video denoising [11, 12]. As already discussed, however, there are compelling reasons to choose a 2D spatial wavelet transform with recursive temporal filtering.

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

2.1. Problem formulation

Given video measurements y , corrupted by i.i.d. Gaussian noise v , with spatial indices i, j and temporal index k ,

$$y(i, j, k) = x(i, j, k) + v(i, j, k), \quad (1)$$

$$i, j = 1, 2, \dots, N, \quad k = 1, 2, \dots, M,$$

our goal is to estimate the true image sequence x . Define $\mathbf{x}(k)$, $\mathbf{y}(k)$, and $\mathbf{v}(k)$ to be the column-stacked video frames at time k , then (1) becomes

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

We propose to denoise in the wavelet domain. Let H be a 2D wavelet transform operator, then (2) is transformed as

$$\mathbf{y}_H(k) = \mathbf{x}_H(k) + \mathbf{v}_H(k), \quad (3)$$

where $\mathbf{y}_H(k) = H\mathbf{y}(k)$, $\mathbf{x}_H(k) = H\mathbf{x}(k)$, and $\mathbf{v}_H(k) = H\mathbf{v}(k)$ denote the respective vectors in the transformed domain.

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

$$\mathbf{x}(k+1) = A(k)\mathbf{x}(k) + B(k)\mathbf{w}(k+1) \quad (4)$$

for some white, stochastic driving process $\mathbf{w}(k)$, thus

$$\mathbf{x}_H(k+1) = A_H(k)\mathbf{x}_H(k) + B_H(k)\mathbf{w}_H(k+1). \quad (5)$$

The inference of A_H and B_H , in general a complicated system-identification problem, is simplified for video by assuming that each frame is related equal to its predecessor, subject to some motion field

$$d(i, j, k) = [d_x(i, j, k), d_y(i, j, k)]^T. \quad (6)$$

Given a shift-invariant, undecimated wavelet transform H , the wavelet coefficients are subject to the same motion as the image itself, thus the dynamic model (5) simplifies as

$$x_H^l(i, j, k+1) = x_H^l(i + d_x(i, j, k), j + d_y(i, j, k), k) + 0 \cdot w_H^l(i, j, k+1) \quad (7)$$

at wavelet level l . It should be noted that (7) approximates motion as locally translatory and is not able to handle zooming and occlusions. In our proposed approach, we assess the validity of (7) for all wavelet coefficients; when (7) is found to be invalid, we make no assumption regarding the temporal relationship in the dynamic model (5):

$$\mathbf{x}_H(k+1) = 0 \cdot \mathbf{x}_H(k) + B_H(k)\mathbf{w}_H(k+1). \quad (8)$$

That is, we have a purely spatial problem, to which standard shrinkage methods can be applied.

2.2. An example: recursive image filtering in the spatial and wavelet domains

As a quick proof of principle, we can denoise 2D images using a recursive 1D wavelet procedure, analogous to denoising 3D video using 2D wavelets. We do not propose this as a superior approach to image denoising, rather as a simple test of recursive wavelet-based denoising, to motivate related approaches in the case of video denoising. We use an autoregressive image model and apply a 1D wavelet transform to each column,

TABLE 1: Percentage increase δ_{MSE} in estimation error relative to the optimal estimator, based on filtering each coefficient independently. In the wavelet case, the independence assumption introduces only slight error when the input PSNR is relatively large (e.g., 10 dB).

SNR(dB)	10	0	-10
δ_{MSE} (spatial)	99.1%	209.6%	91.6%
δ_{MSE} (overcomplete wavelet [13])	9.1%	24.1%	36.7%
δ_{MSE} (orthogonal Daubechies length-4 wavelet)	8.2%	21.1%	32.3%

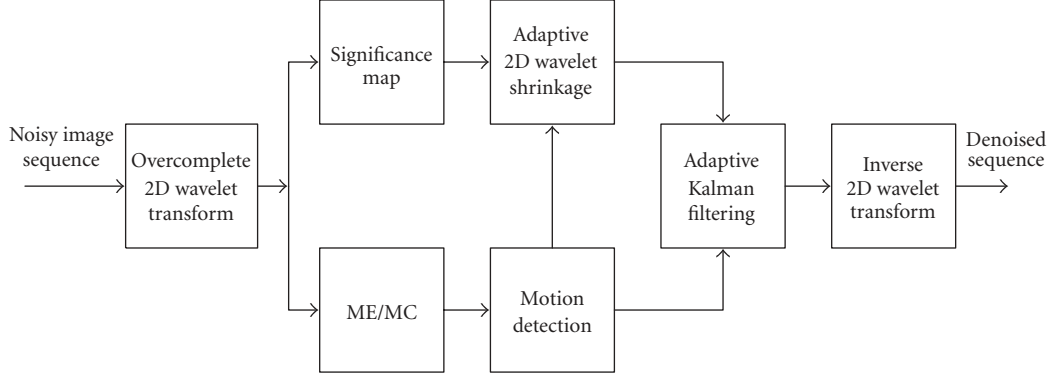


FIGURE 1: Video denoising system.

followed by recursive filtering column by column. We assess the estimator performance in the sense of relative increase of MSE:

$$\delta_{\text{MSE}} = \frac{\text{MSE} - \text{MSE}_{\text{optimal}}}{\text{MSE}_{\text{optimal}}}, \quad (9)$$

where $\text{MSE}_{\text{optimal}}$ is the MSE of the optimal Kalman filter. For the purpose of this example, we use a common image model

$$\begin{aligned} x(i, j) = & \rho_v x(i-1, j) + \rho_h x(i, j-1) - \rho_v \rho_h x(i-1, j-1) \\ & + w(i, j), \quad \rho_h = \rho_v = 0.95, \end{aligned} \quad (10)$$

which is a causal Markov random field (MRF) model and can be converted to a vector autoregressive model [14].

The optimal recursive filtering requires the joint processing of entire image columns, for image denoising, or of entire images, for video denoising. As this would be completely impractical in the video case, for reasons of computational complexity we recursively filter the coefficients *independently*, an assertion which is known to be false, especially for overcomplete (undecimated) wavelet transforms. However, as shown in Table 1, scalar processing in the wavelet domain leads to only very moderate increases in MSE relative to the optimum, even for the strongly correlated coefficients of the overcomplete wavelet transform, whereas this is not at all the case in the spatial domain. We conclude, therefore, that it is reasonable in practice to process the wavelet coefficients independently, with much better performance than such an approach in the spatial domain. It should be noted that the wavelet-based scalar processor is comparable to the optimal filter when $\text{SNR} > 10$ dB, a condition satisfied in many practical applications.

3. THE DENOISING SYSTEM

The success of 1D wavelet denoising of images motivates the extension to the 2D wavelet denoising of video. The block diagram of the proposed video denoising system is illustrated in Figure 1, where the presence of separate temporal and spatial smoothing actions is clear. There are four crucial aspects: (1) the choice of 2D wavelet transform, (2) wavelet-domain motion estimation, (3) adaptive spatial smoothing, and (4) recursive temporal filtering. These steps are detailed below.

2D wavelet transform

A huge number of wavelet transforms have been developed: orthogonal/nonorthogonal, real-valued/complex-valued, decimated/redundant. However, for video denoising, we desire a wavelet with low complexity, directionality selectivity, and, crucially, shift-invariance. The shift-invariance, necessary for motion estimation in the wavelet domain, eliminates all orthogonal or critically decimated wavelets from consideration, so the use of an overcomplete transform is critical.

The 2D dual-tree complex wavelet proposed by Kingsbury, [12] satisfies these requirements very well, unfortunately it is less convenient for motion estimation since the motion information is related to the coefficient phase, which is a nonlinear function of translation. Alternatively, specially designed 2D wavelet transforms (e.g., curvelet, contourlet) are sensitive to feature directions, but are computationally complex for computation. In this paper, we choose to use an overcomplete wavelet representation proposed by Mallat and Zhong [13], which, although it does not have very good directional selectivity, has been used for natural image

denoising with impressive results [9, 15]. However, unlike [9, 15], the wavelet transform employed in this paper has two (instead of three) orientations per scale.

Multiresolution motion estimation

Motion estimation is required to relate two successive video frames to allow temporal smoothing. A wide variety of methods have been studied, however, we will focus on block matching [1, 6], which is simpler to compute and less sensitive to noise in comparison with other approaches, such as optical flow and pixel-recursive methods.

Although regular block matching has widely been studied and used in video processing, *multiresolution* block matching (MRBM) is a much more recent development, but one which appears very naturally in our context of multi-level wavelets.

Multiresolution block matching was proposed by Zhang et al. [16, 17] for wavelet-based video coding, where the basic idea 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 [17], which implies that the interframe relationship between the wavelet coefficients varies from scale to 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 interscale relationship of motion should be

exploited to improve accuracy. We evaluated two traditional multi resolution motion estimation (MRME) methods and following these ideas, we developed two new approaches.

- (1) The standard MRME scheme [16].
- (2) Block matching separately on each level, combined by median filtering [17].
- (3) Joint block matching simultaneously at all levels: let $\epsilon^l(i, j, k, \mathbf{d}(i, j, k))$ denote the displaced frame difference (DFD) of level l . Then the total DFD over all levels is defined as

$$\epsilon(i, j, k, \mathbf{d}(i, j, k)) = \sum_{l=1}^J \epsilon^l(i, j, k, \mathbf{d}(i, j, k)) \quad (11)$$

and the displacement field $\mathbf{d}(i, j, k) = [d_x(i, j, k), d_y(i, j, k)]$ is found by minimizing $\epsilon(i, j, k, \mathbf{d}(i, j, k))$.

- (4) Block matching with smoothness constraint: the above schemes do 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 smoothness constraint and perform BM by solving the optimization problem

$$\arg \min_{\mathbf{d}} \left\{ \sum_{i,j} \left[\epsilon(i, j, k, \mathbf{d}(i, j, k)) + \gamma \cdot \left(\sum_{(p,q) \in N_b(i,j,k)} |d_x(i, j, k) - d_x(i + p, j + q, k)| + |d_y(i, j, k) - d_y(i + p, j + q, k)| \right) \right] \right\}, \quad (12)$$

where $N_b(i, j, k)$ is the neighborhood set of the element (i, j, k) and γ controls the tradeoff between frame difference and smoothness.

For simplicity, we assume a first-order neighborhood for $N_b(i, j, k)$, often used in MRF models for image processing [9, 15]. It is difficult to derive the optimal (in the mean-squared error sense) value of γ because of the high complexity of motion in natural video sequences. However, we find experimentally that PSNR is not sensitive to γ when $0.004 < \gamma < 0.02$, as shown in Figure 2, so we have chosen $\gamma = 0.01$. Also, to keep the algorithm complexity low, we use the iterated conditional mode (ICM) method of Besag [18] to solve the optimization problem in (12). Although ICM cannot guarantee a global minimum, we find its results (Section 4) are satisfactory in the sense of both PSNR and subjective evaluation.

Experimentally, we have found approach 4 to be the most robust to noise and yield reasonable motion estimates. An experimental comparison of all four methods follows in Section 4.

Spatial smoothing

To effectively take advantage of spatial correlations while preserving spatial details, adaptive 2D wavelet shrinkage is applied when the motion estimates are unreliable. As has been done by others [19, 20], we classify the 2D wavelet coefficients into significant and insignificant ones, where the significant coefficients are left untouched to avoid spatial blurring.¹ Motivated by the clustering and persistence properties of wavelet transforms, we define significant coefficients as those which have large local activity:

$$\mathcal{A}^l(i, j) = \sum_{(i,j) \in \Xi^l} |y_H^l(i, j)| \cdot \sum_{(i,j) \in \Xi^{l+1}} |y_H^{l+1}(i, j)|, \quad (13)$$

¹ To minimize MSE, both significant and insignificant wavelet coefficients should be shrunk, as in [19, 20] for image denoising. However, for natural images, shrinking *significant* coefficients often generates denoising artifacts, which we hope to avoid. Thus we choose to denoise significant coefficients *only* in the temporal domain when motion estimation is robust.

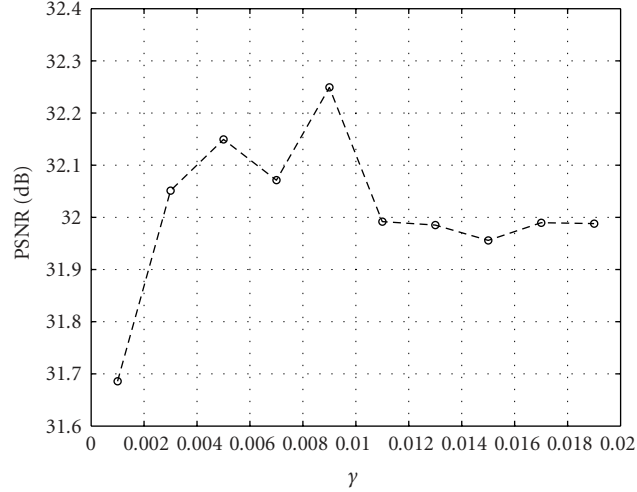


FIGURE 2: Averaged PSNR versus γ curve. PSNR is not sensitive to γ when $0.004 < \gamma < 0.02$.

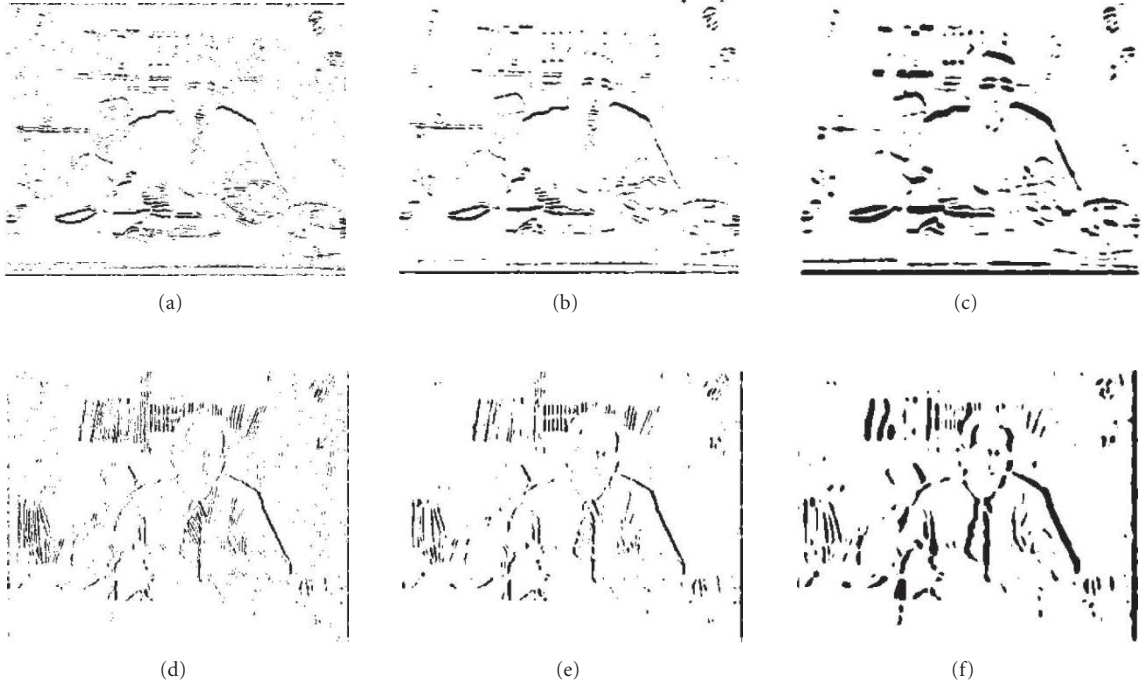


FIGURE 3: Significance maps for a three-level wavelet transform used by the adaptive wavelet shrinkage filter to preserve spatial details. These significance maps are estimated from a noisy image version: (a) level 1 (horizontal); (b) level 2 (horizontal); (c) level 3 (horizontal); (d) level 1 (vertical); (e) level 2 (vertical); (f) level 3 (vertical).

where Ξ^l is the neighborhood structure of level l . In contrast to [19, 20], in (13) we used the local energy of the parent, instead of just using the parent itself, to minimize the potential negative effects of the phase shifts of wavelet filters. The wavelet significance is found by comparing the activity with a level-dependant threshold T^l :

$$\mathcal{S}^l(i, j) = \begin{cases} 1 & \text{if } \mathcal{A}^l(i, j) > T^l, \\ 0 & \text{if } \mathcal{A}^l(i, j) \leq T^l. \end{cases} \quad (14)$$

The thresholds are level-adaptive, set to identify as significant 5% of the coefficients on the two finest scales and 10% on coarser scales. Figure 3 shows the significance maps for the wavelet coefficients in the first three levels of the image sequence *Salesman*, clearly identifying the high-activity (detail) areas, not to be blurred in the 2D wavelet shrinkage.

Given appropriately chosen thresholds T^l , we model the insignificant wavelet coefficients, dominated by noise, as independent zero-mean Gaussian [8, 19] with *spatially varying* variances. Motivated by Table 1, processing the wavelet

coefficients independently leads to relatively slight increases in MSE, in which case the appropriate shrinkage is the linear-Bayes-Wiener

$$\hat{x}_H^l(i, j) = \frac{(\hat{\sigma}_{x_H}^l(i, j))^2}{(\hat{\sigma}_{x_H}^l(i, j))^2 + (\sigma_{v_H}^l)^2} \cdot y_H^l(i, j), \quad (15)$$

where the measurement noise variance $(\sigma_{v_H}^l)^2$ is given, or may be robustly estimated [21]. All that remains is the inference of the process variance $(\sigma_{x_H}^l)^2$, which we find as a spatial sample variance over a 7×7 local window of insignificant coefficients:

$$\begin{aligned} & (\hat{\sigma}_{x_H}^l(i, j))^2 \\ &= \max \left(0, \frac{\sum_{p, q \in S_0^l} (y_H^l)^2(i+p, j+q)}{\sum_{p, q \in S_0^l} 1} - (\sigma_{v_H}^l)^2 \right), \end{aligned} \quad (16)$$

where $S_0^l = \{(p, q) : S^l(p, q) = 0\}$.

Wavelet-based recursive filtering

As was illustrated in Section 2, filtering the wavelet coefficients independently, a particularly simple and computationally efficient approach, gives good results in the sense of MSE. For video processing, we further develop this idea and perform temporal Kalman filtering in the wavelet domain, achieving simple *scalar* filtering close to optimal Kalman filtering.

Because motion estimation is an ill-posed problem, there often exist serious estimation errors, for example around motion boundaries, in which case the temporal dynamic model (7) is invalid. To adapt to motion estimation errors, we perform hypothesis testing on (7) to establish validity based on the observations y_H . Specifically, when the motion information is unambiguous,

$$|y_H^l(i, j, k) - y_H^l(i + d_x(i, j), j + d_y(i, j), k - 1)| < \beta \sigma_{v_H}^l, \quad (17)$$

only temporal Kalman filtering is used, whereas when the motion estimates are poor,

$$|y_H^l(i, j, k) - y_H^l(i + d_x(i, j), j + d_y(i, j), k - 1)| \geq \beta \sigma_{v_H}^l, \quad (18)$$

we perform only 2D wavelet shrinkage (15) on the *insignificant* wavelet coefficients, leaving *significant* coefficients untouched. The threshold $\beta = 2\sqrt{2}$ is set to preserve temporal matches for most ($\sim 95\%$) correctly matched pixels.

The resulting Kalman filter is particularly simple because of the deterministic form of (7); that is, the standard Kalman filter [14] reduces to a dynamic temporal averaging filter.

4. EXPERIMENTAL RESULTS

The proposed denoising approach has been tested using the standard image sequences *Miss America*, *Salesman*, and *Paris*, using a three-level wavelet decomposition. First, Figure 4

compares our regularized (12) and nonregularized (11) MRBM approaches with standard MRBM [16] and standard MRBM with median filtering [17]. Since the true motion field is unknown, we evaluate the performance of noisy motion estimation by comparing with the motion field estimated from noise-free images (Figure 4(b)), and by comparing the corresponding denoising results. The unregularized approaches do not exploit any smoothness or prior knowledge, and therefore perform poorly in the presence of noise (Figures 4(c), 4(d), 4(e)). In comparison, our proposed approach gives far superior results (Figure 4(f)). Although our MRBM approach introduces one new parameter γ , experimentally we found PSNR to be weakly dependent on γ , as illustrated in Figure 2, and in all of the following tests, we fix $\gamma = 0.01$.

Next, we compare our proposed denoising approach with three recently published methods: two wavelet-based video denoising schemes [5, 12] and one non-wavelet nonlinear approach [22]. Selesnick and Li [12] generalized the ideas of many well-developed 2D wavelet-based image denoising methods and used a complex-valued 3D wavelet transform for video denoising. Pizurica et al. [5] combined a temporal recursive filter with sophisticated wavelet-domain image denoising, but without motion estimation. Zlokolica and Philips [22] used multiple-class averaging to suppress noise, which performs better than the traditional nonlinear methods, such as the α -trimmed mean filter [23] and the rational filter [24]. Table 2 compares the PSNRs averaged from frames 10 to 30 of the sequence *Salesman* for different noise levels. Our approach yields higher PSNRs than those in [12, 22], and is comparable to Pizurica's results which use a sophisticated image denoising scheme in the wavelet domain. However, the similar PSNRs between the results of our proposed method and that of Pizurica et al. [5] obscure the significant differences, as made very clear in Figures 5 and 6. In particular, we perform less spatial smoothing, shrinking only insignificant coefficients, but rely more heavily upon temporal averaging. Thus, our results have very little spatial blurring, preserving subtle textures and fine details, such as the desk-top and bookshelf in Figure 5 and the plant in Figure 6.

Table 3 compares the overcomplete and orthogonal Daubechies-4 wavelet transforms for video denoising. For the orthogonal Daubechies-4 wavelet transform, we perform motion estimation and recursive filtering for each scale separately. We see that the overcomplete wavelet outperforms the Daubechies wavelet by more than 1dB in PSNR. As discussed in the introduction, this advantage of the overcomplete wavelet is expected, stemming from its shift-invariance, whereas the orthogonal Daubechies wavelets are highly shift-sensitive.

5. CONCLUSIONS

We have proposed a new approach to video denoising, combining the power of the spatial wavelet transform and temporal filtering. Most significantly, motion estimation/compensation, temporal filtering, and spatial smoothing are *all* undertaken in the wavelet domain. We

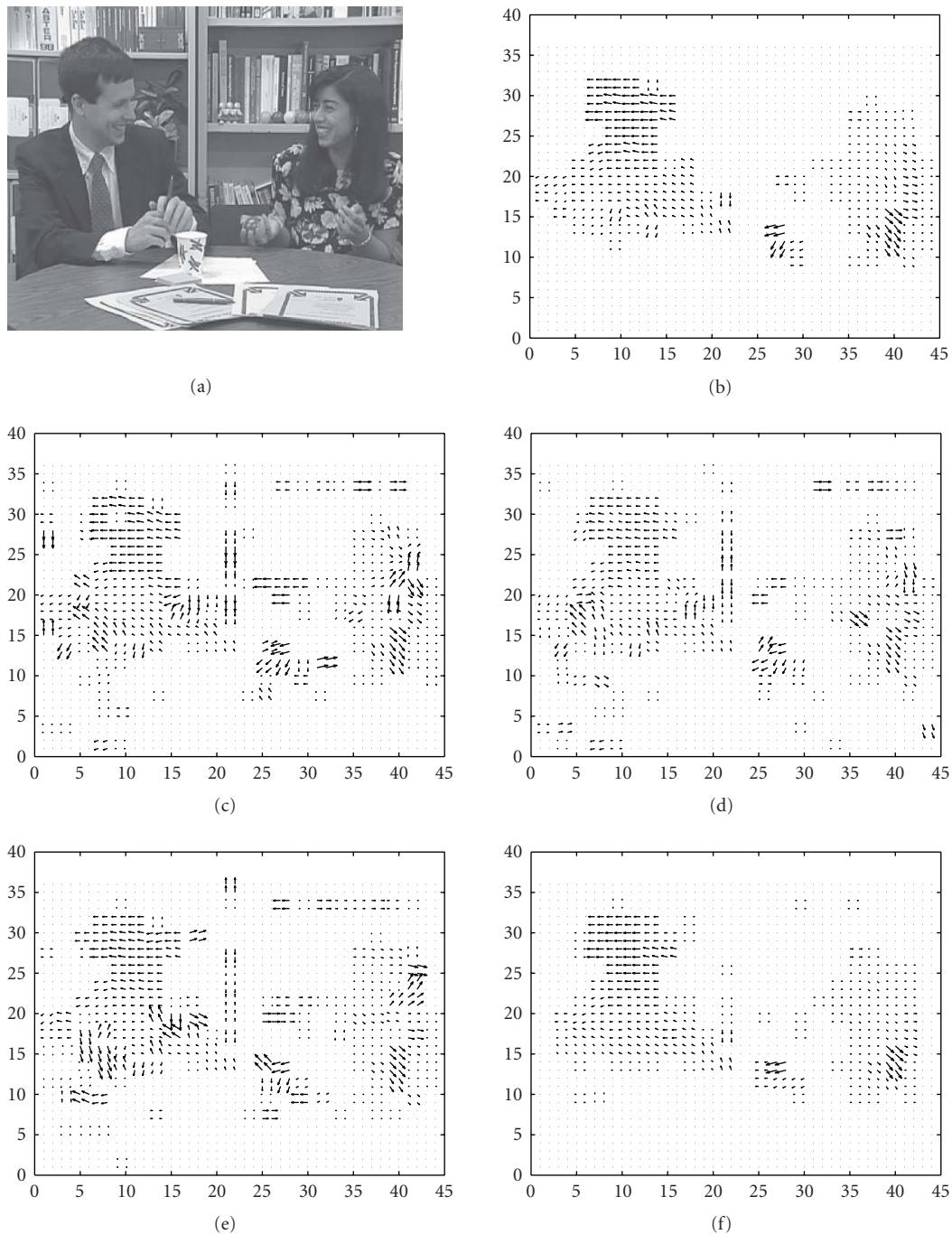


FIGURE 4: A comparison of four methods of motion estimation applied to the Paris sequence (a) with added noise. The three methods of (c) standard MRBM [16], (d) standard MRBM with median filtering [17], and (e) our unregularized approach (proposed approach 3) do not exploit any smoothness or prior knowledge of the motion and perform poorly in the presence of noise. In contrast, our proposed approach (f), smoothness-constrained MRBM with $\gamma^3 = 0.01$, compares very closely with the noise-free estimates in (b).

TABLE 2: Comparison of PSNR (dB) of the proposed method and several other video denoising approaches for the *Salesman* sequence.

PSNR (original)	28.2	24.6	22.1
PSNR (proposed method)	33.9	31.6	30.5
PSNR (3D complex DWT [12])	32.1	30.5	29.3
PSNR (Pizurica et al. [5])	33.7	31.7	30.5
PSNR (Zlokolica and Philips [22])	32.5	30.8	29.7

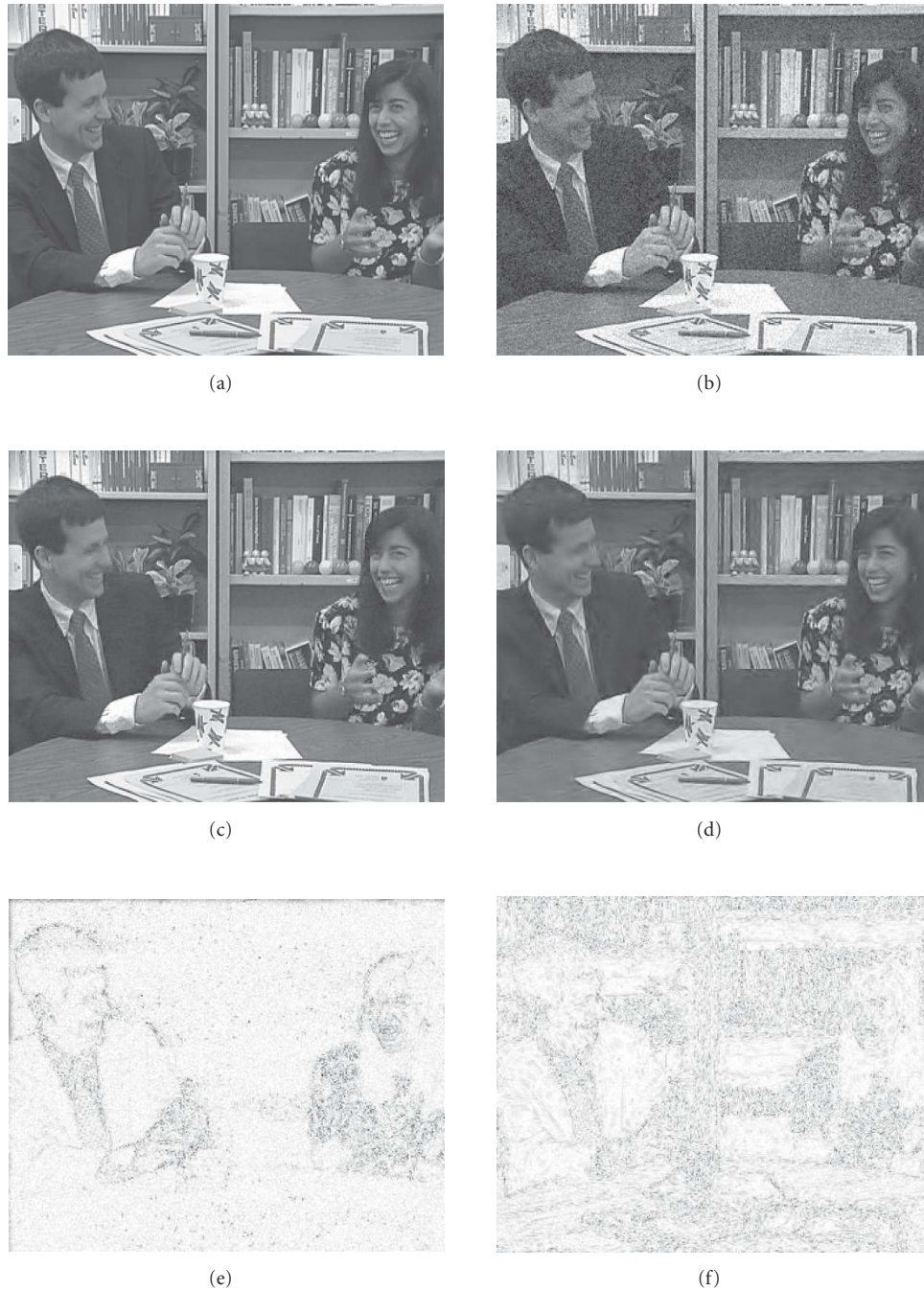


FIGURE 5: Comparison of (c) denoising of our proposed approach and (d) denoising by Pizurica's approach [5] ($\sigma_{vH} = 15$). (a) Represents the original image and (b) the noisy image. Our approach can better preserve spatial details, such as textures on the desktop as made clear in the difference images in (e) that represents absolute difference between (a) and (c), and in (f) that represents absolute difference between (a) and (d).

also avoid spatial blurring by restricting to temporal filtering when motion estimates are reliable, and spatially shrinking only insignificant coefficients when the motion is unreliable. Tests on standard video sequences show that our results yield comparable PSNR to the state-of-the-art methods in the

literature, but with considerably improved preservation of fine spatial details. Future improvements may include more sophisticated approaches to spatial filtering, such as that in [5], and more flexible temporal models to better represent image dynamics.

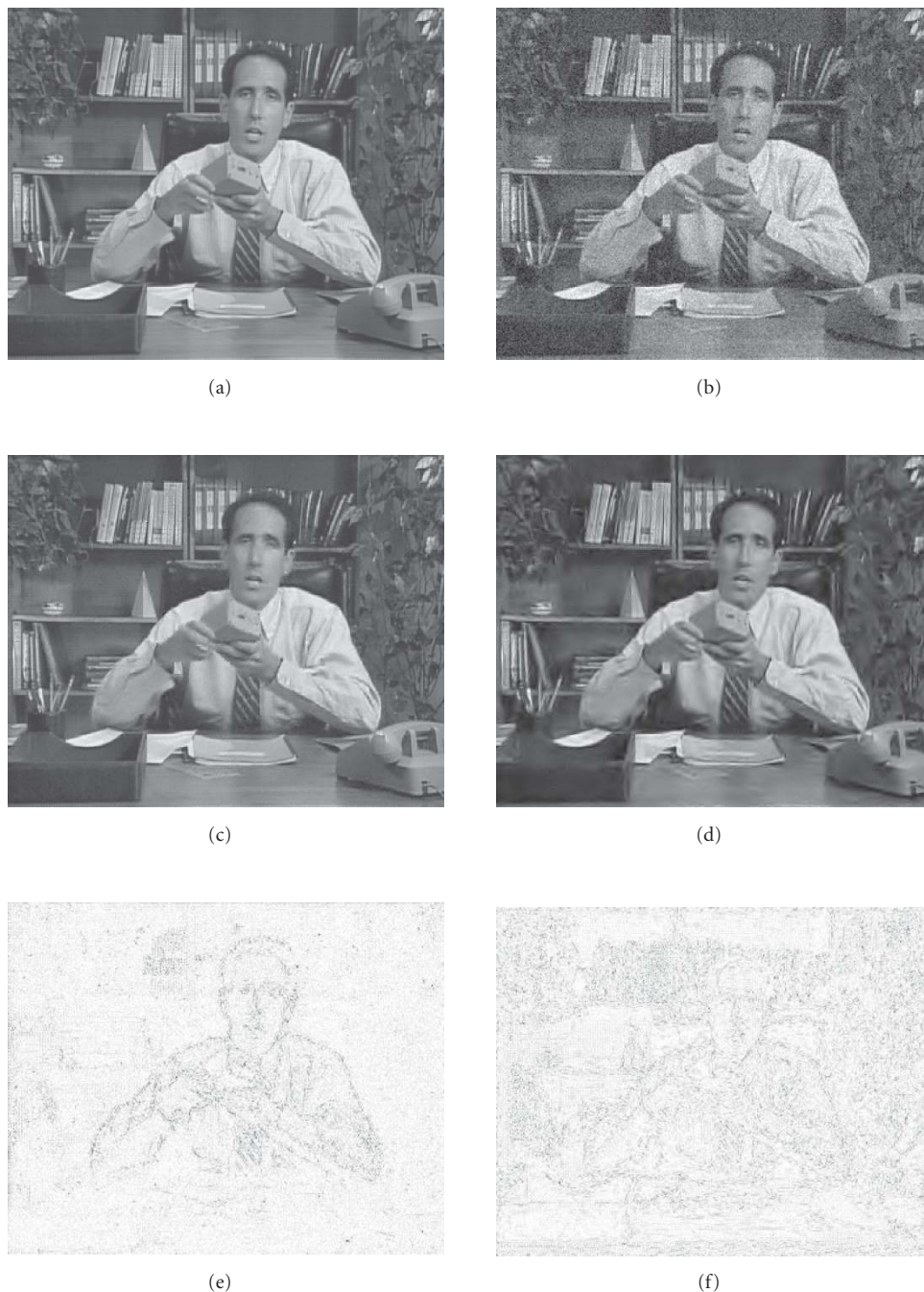


FIGURE 6: Denoising results for Salesman. Note in particular the textures of the plants, well preserved in our results in (c) that represents denoising by the proposed approach, but obviously blurred in (d) that represents denoising by Pizurica's approach [5]. (a) Original image, (b) noisy image, (e) absolute difference between (a) and (c), and (f) represents the absolute difference between (a) and (d).

TABLE 3: Comparison of PSNR (dB) of the overcomplete and the orthogonal length-4 Daubechies wavelet for the *Salesman* sequence. Due to shift-invariance, the overcomplete wavelet yields much better results than the orthogonal length-4 Daubechies wavelet.

PSNR (original)	28.2	24.6	22.1
PSNR (overcomplete wavelet)	33.9	31.6	30.5
PSNR (orthogonal length-4 Daubechies wavelet)	32.4	30.3	29.5

REFERENCES

- [1] J. C. Brailan, R. P. Kleihorst, S. Efstratiadis, A. K. Katsaggelos, and R. L. Lagendijk, "Noise reduction filters for dynamic image sequences: a review," *Proceedings of IEEE*, vol. 83, no. 9, pp. 1272–1292, 1995.
- [2] G. R. Arce, "Multistage order statistic filters for image sequence processing," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1146–1163, 1991.
- [3] J. Kim and J. W. Woods, "Spatio-temporal adaptive 3-D Kalman filter for video," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 414–424, 1997.
- [4] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, 2000.
- [5] A. Pizurica, V. Zlokolica, and W. Philips, "Combined wavelet domain and temporal video denoising," in *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS '03)*, pp. 334–341, Miami, Fla, USA, July 2003.
- [6] M. K. Ozkan, M. I. Sezan, and A. M. Tekalp, "Adaptive motion-compensated filtering of noisy image sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 4, pp. 277–290, 1993.
- [7] J. C. Brailan and A. K. Katsaggelos, "Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences," *IEEE Transactions on Image Processing*, vol. 4, no. 9, pp. 1236–1251, 1995.
- [8] M. Kivanc Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, 1999.
- [9] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.
- [10] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [11] P. M. B. van Roosmalen, S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Noise reduction for image sequences using an oriented pyramid thresholding technique," in *Proceedings of IEEE International Conference on Image Processing (ICIP '96)*, vol. 1, pp. 375–378, Lausanne, Switzerland, September 1996.
- [12] I. W. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Wavelets: Applications in Signal and Image Processing X*, vol. 5207 of *Proceedings of SPIE*, pp. 607–618, San Diego, Calif, USA, August 2003.
- [13] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, 1992.
- [14] A. Rosenfeld and A. Kak, *Digital Picture Processing*, Academic Press, New York, NY, USA, 1982.
- [15] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Transactions on Image Processing*, vol. 6, no. 4, pp. 549–565, 1997.
- [16] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, no. 3, pp. 285–296, 1992.
- [17] J. Zan, M. O. Ahmad, and M. N. S. Swamy, "New techniques for multi-resolution motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 793–802, 2002.
- [18] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.
- [19] J. Liu and P. Moulin, "Image denoising based on scale-space mixture modeling of wavelet coefficients," in *Proceedings of IEEE International Conference on Image Processing (ICIP '99)*, vol. 1, pp. 386–390, Kobe, Japan, October 1999.
- [20] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 323–331, 2003.
- [21] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [22] V. Zlokolica and W. Philips, "Motion- and detail-adaptive denoising of video," in *IS&T/SPIE 16th Annual Symposium on Electronic Imaging: Image Processing: Algorithms and Systems III*, vol. 5298 of *Proceedings of SPIE*, pp. 403–412, San Jose, Calif, USA, January 2004.
- [23] J. Bednar and T. Watt, "Alpha-trimmed means and their relationship to median filters," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 32, no. 1, pp. 145–153, 1984.
- [24] F. Cocchia, S. Carrato, and G. Ramponi, "Design and real-time implementation of a 3-D rational filter for edge preserving smoothing," *IEEE Transactions on Consumer Electronics*, vol. 43, no. 4, pp. 1291–1300, 1997.

Fu Jin received the B.S. and M.S. degrees from the Department of Electrical Engineering, Changsha Institute of Technology, China, in 1989 and 1991, respectively, and Ph.D. degree from the Department of Systems Design Engineering, University of Waterloo, in 2004. His research interests include signal processing, image/video processing, and statistical modeling. He is now a Senior R&D Engineer with VIXS Company in Toronto, Canada, working on video compression and processing.



Paul Fieguth received the B.A.Sc. degree from the University of Waterloo, Ontario, Canada, in 1991 and the Ph.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, in 1995, both degrees in electrical engineering. He joined the faculty at the University of Waterloo in 1996, where he is currently an Associate Professor in Systems Design Engineering. He has held visiting appointments at the Cambridge Research Laboratory, at Oxford University, and the Rutherford Appleton Laboratory in England, and at INRIA/Sophia in France, with postdoctoral positions in the Department of Computer Science at the University of Toronto and in the Department of Information and Decision Systems at MIT. His research interests include statistical signal and image processing, hierarchical algorithms, data fusion, and the interdisciplinary applications of such methods, particularly to remote sensing.



Lowell Winger received the Ph.D. degree in electrical and computer engineering from University of Toronto, Canada, in 1998, and M.A.Sc. and B.A.Sc. degrees in systems design engineering from University of Waterloo in 1996 and 1994. Since 2001, he has actively contributed to the development of international video standards: ITU-T/ISO VCEG/MPEG/JVT (H.264/MPEG4-AVC), SMPTE (VC-1), DVD Forum WG-1 (HD-DVD), ATSC-S6, and DVB-AVC. In 2002, with the acquisition of VideoLocus where he was CTO, he joined LSI Logic as Principal Engineer of advanced video codec algorithms. Before cofounding VideoLocus, which developed the first real-time SD H.264 encode platform, he worked as a Senior Research and Design Engineer on video DSP algorithms, multipass encoding, and video processing platforms at PixStream Inc., which was acquired by Cisco in 2001. Prior to Pixstream/Cisco, he was an Assistant Professor at the University of Ottawa, teaching digital signal processing, pattern recognition, and image and video courses, and authoring several of his over 60 refereed journal articles, standards contributions, and conference presentations. He holds one issued US patent, has three US patents allowed, and has several pending. He has also held positions with HP, Omron Japan, Atomic Energy of Canada, University of Waterloo, McMaster University Hospital, and Raytheon.

