Combined Spatial and Temporal Domain Wavelet Shrinkage Algorithm for Video Denoising

Eric J. Balster, *Member, IEEE*, Yuan F. Zheng, *Fellow, IEEE*, and Robert L. Ewing, *Senior Member, IEEE*

***Abstract*—A combined spatial- and temporal-domain wavelet shrinkage algorithm for video denoising is presented in this paper. The spatial-domain denoising technique is a selective wavelet shrinkage method which uses a two-threshold criteria to exploit the geometry of the wavelet subbands of each video frame, and each frame of the image sequence is spatially denoised independently of one another. The temporal-domain denoising technique is a selective wavelet shrinkage method which estimates the level of noise corruption as well as the amount of motion in the image sequence. The amount of noise is estimated to determine how much filtering is needed in the temporal-domain, and the amount of motion is taken into consideration to determine the degree of similarity between consecutive frames. The similarity affects how much noise removal is possible using temporal-domain processing. Using motion and noise level estimates, a video denoising technique is established which is robust to various levels of noise corruption and various levels of motion.**

***Index Terms*—Combined spatial- and temporal-domain processing, motion estimation, selective wavelet shrinkage, video denoising.**

I. Introduction

THE recent advance in multimedia technology has promoted a large amount of research in the area of image and video processing. Included in many image and video processing algorithms such as compression, enhancement, and target recognition are preprocessing functions for noise removal. Noise removal is one of the most common and important processing steps in many image and video systems.

Because of the commonality of noise removal functions in most image and video systems, there has been an large amount of research dedicated to the subject of image denoising over the past several decades, and many different mathematical tools have been proposed. Various established denoising methods using variable coefficient linear filters [5], [18], [23], [30], adaptive nonlinear filters [10], [17], [20], [32], discrete cosine transform (DCT)-based solutions [12], cluster filtering [29], genetic algorithms [28], and fuzzy logic [13], [25], etc., have all been proposed in the literature.

Perona and Malik use a series of Gaussian filters and a method referred to as anisotripic diffusion to estimate edge locations and both smooth regions and enhance edges in imagery [19].

Manuscript received March 24, 2004; revised July 3, 2005. This paper was recommended for publication by Associate Editor E. Izquierdo.

E. J. Balster and R. L. Ewing are with the Information Directorate/IFTA, Wright-Patterson AFB, OH 45433-7334 USA (e-mail: eric.balster@wpafb.af.mil; robert.ewing@wpafb.af.mil).

Y. F. Zheng is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: zheng@ece.osu.edu).

Digital Object Identifier 10.1109/TCSVT.2005.857816

The wavelet transform has also been used to suppress noise in digital images. It has been shown that the reduction in absolute value of wavelet coefficients is successful in signal restoration [16]. This process is known as wavelet shrinkage. Other denoising techniques select or reject wavelet coefficients based on their predicted contribution to reconstructed image quality. This process is known as *selective* wavelet shrinkage, and many works have used it as the preferred method of image denoising [1], [4], [6], [7], [9], [11], [14]–[16], [27]. Notably, Pizurica *et al.* approximate the significance of each coefficient in the wavelet transform by modeling them as Marcov random sequence and applying a Bayesian probabilistic formulation. The measure of significance is then used to determine the reduction amount of each coefficient [21]. Also, Zhang uses a thresholding neural network to determine the significance of wavelet coefficients [31].

However, until recently, the removal of noise in video signals has not been studied seriously. Cocchia *et al.* developed a three-dimensional (3-D) rational filter for noise removal in video signals [3]. The 3-D rational filter not only removes noise, but also preserves important edge information. Also, the 3-D rational filter uses a motion estimation technique. Where there is no motion detected, the 3-D rational filter is applied in the temporal domain. Otherwise, only spatial-domain processing is applied.

Later, Zlokolica *et al.* uses two new techniques for noise removal in image sequences [32]. Both these new techniques show improved results upon the method of [3]. The first method is an alpha-trimmed mean filter of [2] extended to video signals, and the second is the K nearest neighbors (KNN) filter. Both alpha-trimmed and KNN denoising methods are based on ordering the pixel values in the neighborhood of the location to be filtered, and averaging a portion of those spatially contiguous pixels. Each of these methods attempts to average values which are close in value, and avoid averaging values which are largely dissimilar in value. Thus, the image sequence is smoothed without blurring edges, or smearing motion.

However, because the success of the wavelet transform over other mathematical tools in denoising images, some researchers believe that wavelets may be successful in the removal of noise in video signals as well. Pizurica *et al.* uses a wavelet-based image denoising method to remove noise from each individual frame in an image sequence and then applies a temporal filtering process for temporal-domain noise removal [22]. The combination of wavelet image denoising and temporal filtering outperforms both wavelet based image denoising techniques [1], [15], [16], [21] and spatial-temporal filtering techniques [2], [3], [32].

The temporal-domain filtering technique described in [22] is a linear IIR filter which will continue to filter until it reaches a large temporal discontinuity. It will not filter the locations of large temporal discontinuity where the absolute difference in neighboring pixel values is greater than a threshold T thus preserving motion while removing noise.

Although temporal processing aids in the quality of the original image denoising method, the parameter T varies with differing video signals for improved performance. That is, proper selection of T may be large in sequences where there is little motion for improved noise removal, i.e., there is more redundancy between consecutive frames. Thus, the redundancy may be exploited by a large T to improve video quality. However, in image sequences where there exists a large amount of motion, consecutive frames are more independent and there exists little to no redundancy to exploit. Thus, the parameter T must be small to achieve optimal performance.

In the case of video denoising, it has been fairly well documented that the amount of noise removal achievable from temporal-domain processing, while preserving overall quality, is dependent on the amount of motion in the original video signal [3], [22]. Thus, a robust, high-quality video denoising algorithm is required to not only be scalable to differing levels of noise corruption, but also scalable to differing amounts of motion in the original signal. Unfortunately, this principle has not been seriously considered in video denoising.

In this paper, we develop a noise removal algorithm for video signals. This algorithm uses selective wavelet shrinkage in all three dimensions of the image sequence and proves to outperform the few video denoising algorithms given in the relevant literature. First, the individual frames of the sequence are denoised by the method of [1], which we had developed earlier. Then a new selective wavelet shrinkage method is used for temporal-domain processing.

Also, a motion estimation algorithm is developed to determine the amount of temporal-domain processing to be performed. Several motion estimators have been proposed [3], [22], but few are robust to noise corruption. The proposed motion estimation algorithm is robust to noise corruption and an improvement over the motion estimation method of [3]. The proposed denoising algorithm, including the proposed motion estimation method, is experimentally determined to be an improvement over the methods of [3], [22], [32].

Following the introduction, Section II gives a brief description of the image denoising method of [1], used as the spatial denoising method in the proposed video denoising algorithm. Section III describes the temporal-domain wavelet shrinkage method and explores the proper order of temporal and spatial-domain processing functions. Section IV provides the proposed motion estimation index used in the temporal-domain processing and compares it with the motion estimation method of [3]. Section V develops the parameters for temporal-domain processing, and Section VI gives the experimental results of the proposed method as well as other established methods. Section VII concludes the paper.

II. Spatial Domain Denoising Technique

The proposed video denoising technique uses the selective wavelet shrinkage algorithm of [1] for denoising of the individual frames of the image sequence. A brief review of the algorithm is included in this section for completeness.

A. Coefficient Selection Method

First, we will review the proposed coefficient selection method of [1]. The coefficient selection method is based on a two-threshold criteria, selecting wavelet coefficients with large magnitude and spatial regularity.

Assume that an image signal is corrupted with additive noise, i.e.,

$$\widetilde{f}(l) = f(l) + \eta(l) \tag{1}$$

where $f(l)$ is the noiseless image pixel of position l, $\eta(l)$ is a random noise function, and $\widetilde{f}(l)$ is the corresponding corrupted signal.

The wavelet shrinkage algorithm takes the nondecimated two-dimensional (2-D) wavelet transform of $\hat{f}(l)$, and selects the wavelet coefficients for denoising. The first step for selecting the wavelet coefficient is to find a binary label for each coefficient which collectively forms a binary map. The binary map is then used to determine whether or not a particular wavelet coefficient is included in a regular spatial feature. The nondecimated, 2-D wavelet transform of $\widetilde{f}(l)$ generates coefficients $\widetilde{\lambda}_{m,k}[l]$ of spatial location l, resolution k, and subband $m \in \{lh, hl, hh\}$. The subband designation m denotes the low–high (lh), high–low (hl), and high–high (hh) subbands. For example, the lh subband is produced by convolving the input function with the low-pass scaling filter $h[\cdot]$ in the horizontal dimension then convolving the result with the high-pass wavelet filter $g[\cdot]$ in the vertical dimension. $\widetilde{\lambda}_{m,k}[l]$ is used to create the preliminary binary label $I_{m,k}[l]$.

$$I_{m,k}[l] = \begin{cases} 1, & \text{when } \left|\widetilde{\lambda}_{m,k}[l]\right| > \tau \\ 0, & \text{else} \end{cases} \tag{2}$$

where τ is a threshold for selecting *valid coefficients* in the construction of the binary coefficient map. A *valid coefficient* is defined as a coefficient $\widetilde{\lambda}_{m,k}[l]$ which results in $I_{m,k}[l] = 1$; hence the coefficient has been selected due to its magnitude. After coefficients are selected by magnitude, spatial regularity is used to further examine the role of the *valid coefficient*: whether it is isolated noise or part of a spatial feature. The number of supporting binary values around a particular nonzero value $I_{m,k}[l]$ is used to make the judgement. The support value $S_{m,k}[l]$ is the sum of all $I_{m,k}[l]$ which support the current binary value $I_{m,k}[l]$; that is, the total number of all *valid coefficients* which are spatially connected to $I_{m,k}[l]$.

A coefficient is spatially connected to another if there exists a continuous path of *valid coefficients* between the two. Fig. 1 gives a generic coefficient map. The *valid coefficients* are highlighted in gray. From Fig. 1 it can be shown that coefficients A, B, C, and H do not support any other *valid coefficients* in the coefficient map. However, coefficients D and F support each other, coefficients E and G support each other, and N and O support each other. Also, coefficients I, J, K, L, M, P, Q, and R all support one another. Fig. 2 gives the value of $S_{m,k}[l]$ for each of the *valid coefficients* given in Fig. 1. A method of computing

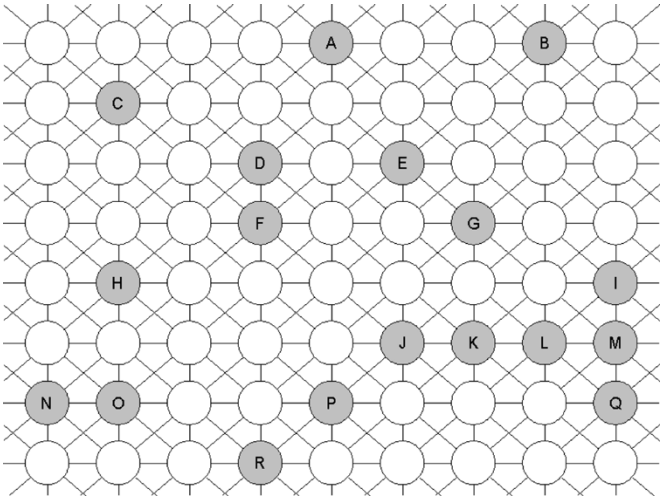

Fig. 1. Generic coefficient array.

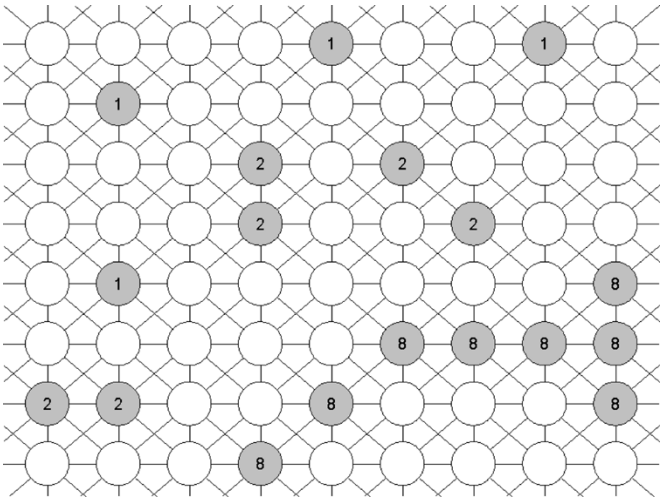

Fig. 2. Generic coefficient array, with corresponding $S_{\cdot,k}$ values.

$S_{m,k}[l]$ is given in [1]. $S_{m,k}[l]$ is used to refine the original binary map $I_{m,k}[l]$ by

$$J_{\cdot,k}[x,y] = \begin{cases} 1, & \text{when } S_{\cdot,k}[x,y] > s, \\ & \quad \text{or } J_{\cdot,k+1}[x,y] I_{\cdot,k}[x,y] = 1 \\ 0, & \text{else} \end{cases} \tag{3}$$

where $J_{m,k}[l]$ is the refined coefficient map, and s is the necessary number of support coefficients for selection. $J_{\cdot,\cdot}[\cdot]$ is calculated recursively, starting from the highest multiresolution level, and progressing downward.

Equation (3) is equal to one when there exists enough wavelet coefficients of large magnitude around the current coefficient. However, it also retains coefficients in which the magnitude of the coefficient is effectively large $(I_{m,k}[l] = 1)$ but not locally supported $(J_{m,k}[l] = 0)$ only if the coefficient of the larger scale is large and locally supported $(J_{m,k+1}[l] = 1)$. The decision to use this criterion is in the somewhat rare case when a useful coefficient is not locally supported. In the general case, wavelet coefficients of images are clustered together, but rarely they are isolated. In [16], wavelet coefficients are modified *only* by their evolution across scales. Regular signal features contain wavelet coefficients which increase with increasing scale. Thus, if there exists a useful coefficient which is isolated in an image, it is reasonable that a coefficient in the same spatial location of an increase in scale will be sufficiently large and spatially supported. Thus, the coefficient selection method provided by (3) selects coefficients which are sufficiently large and locally supported as well as isolated coefficients which are sufficiently large and supported by scale.

This type of scale-selection is consistent with the findings of Said and Pearlman [24], who developed an image codec based on a "spatial self-symmetry" between differing scales in wavelet transformed images. They discovered that most of an images energy is concentrated in the low-frequency subbands of the wavelet transform. And because of the self-symmetry properties of wavelet transformed images, if a coefficient value is insignificant (i.e., of small value or zero), then it can be assumed that the coefficients of higher spatial frequency and same spatial location will be insignificant. In our application, however, we are looking for significance rather than insignificance, so we look to the significance of lower frequency coefficients to determine significance of the current coefficient. In this way, the preliminary binary map is refined by both spatial and scalar support, given by (3).

The final coefficients retained for reconstruction are given by

$$L_{\cdot,k}[x,y] = \begin{cases} \widetilde{\lambda}_{\cdot,k}[x,y], & \text{when } J_{\cdot,k}[x,y] = 1 \\ 0, & \text{else.} \end{cases} \tag{4}$$

The denoised image is reconstructed by synthesizing the supported wavelet coefficients, $L_{m,k}[l]$ using the nondecimated inverse wavelet transform.

In general, natural and synthetic imagery can be compactly represented in few wavelet coefficients of large magnitude. These coefficients are in general spatially clustered. Thus, it is useful to obtain selection methods based on magnitude and spatial regularity to distinguish between useful coefficients which are representative of the image and useless coefficients representative of noise. The two-threshold criteria for the rejection of noisy wavelet coefficients is a computationally simple test for magnitude and spatial regularity which can effectively distinguish between useful and useless coefficients.

B. Determining the τ and s Thresholds

In determining the optimal threshold values, it is found that both thresholds are a function of the noise standard deviation σ_n [1]. Therefore

$$\tau = a_\tau \widetilde{\sigma_n} + b_\tau \tag{5}$$

and

$$s = \lfloor a_s \widetilde{\sigma_n} + b_s \rfloor \tag{6}$$

where $\widetilde{\sigma_n}$ is an estimate of the noise $a_\tau = 2.12$, $b_\tau = 0.80$, $a_s = 0.26$, and $b_s = 2.81$. The estimate of the noise is taken from that of [21] and is given by

$$\widetilde{\sigma_n} = \frac{\text{Median}\left(\left|\widetilde{\lambda}_{hh,0}[\cdot]\right|\right)}{0.6745} \tag{7}$$

where $\widetilde{\lambda}_{hh,0}[\cdot]$ are the noisy wavelet coefficients of the 0^{th} level and hh subband. For a more detailed treatment of the proposed spatial denoising method, refer to [1].

III. Temporal Denoising and Order of Operations

In this section, we develop the principal algorithm for video denoising. Additional mechanisms required by this algorithm will be discussed in later sections.

A. Temporal Domain Denoising

Let us define f_l^z as a pixel of spatial location l and frame z in a given image sequence. The nondecimated wavelet transform applied in the temporal domain is given by

$$\lambda_{k+1}^{3-D}[l,z] = \sum_p g[p]\alpha_k^{3-D}[l, 2^{k+1}p - z] \tag{8}$$

and

$$\alpha_{k+1}^{3-D}[l,z] = \sum_p h[p]\alpha_k^{3-D}[l, 2^{k+1}p - z] \tag{9}$$

where

$$\alpha_{-1}^{3-D}[l,z] = f_l^z \tag{10}$$

where $\lambda_k^{3-D}[l,z]$ is the high-frequency wavelet coefficient of spatial location l, frame z, and scale k. Also, $\alpha_k^{3-D}[l,z]$ is the low-frequency scaling coefficient of spatial location l, frame z, and scale k. Thus, multiple resolutions of wavelet coefficients may be generated from iterative calculation of (8) and (9).

The wavelet function used in the temporal-domain denoising process is the Haar wavelet given by

$$h[n] = \begin{cases} \frac{1}{\sqrt{2}}, & n = 0,1 \\ 0, & \text{else} \end{cases} \qquad g[n] = \begin{cases} \frac{-1}{\sqrt{2}}, & n = 0 \\ \frac{1}{\sqrt{2}}, & n = 1 \\ 0, & \text{else.} \end{cases} \tag{11}$$

The decision to use the Haar wavelet is based on experimentation with several other wavelet functions and finding the greatest results with the Haar. The compact support of the Haar wavelet makes it a suitable function for denoising applications. Because of it's compact support, the Haar coefficients represent least number of original pixels in comparison to other types of wavelets. Thus, when a coefficient is removed because of its insignificance, the result affects the smallest area of the original signal in the reconstruction.

Significant wavelet coefficients are selected by their magnitude with a threshold operation

$$L_k^{3-D}[l,z] = \begin{cases} \lambda_k^{3-D}[l,z], & \text{when } \left|\lambda_k^{3-D}[l,z]\right| > \tau_z[l] \\ 0, & \text{else} \end{cases} \tag{12}$$

where $L_k^{3-D}[l,z]$ are the thresholded wavelet coefficients used in signal reconstruction, and $\tau_z[\cdot]$ is the threshold value. The resulting denoised video signal is computed via the inverse nondecimated wavelet transform

$$\hat{\alpha}_k^{3-D}[l,z] = \frac{1}{2}\sum_p h[p]\hat{\alpha}_{k+1}^{3-D}[l, z - 2^{k+1}p] + \frac{1}{2}\sum_p g[p]L_{k+1}^{3-D}[l, z - 2^{k+1}p] \tag{13}$$

which leads to

$$\hat{f}_l^{z,3-D} = \hat{\alpha}_{-1}^{3-D}[l,z] \tag{14}$$

where $\hat{f}_l^{z,3-D}$ is the temporally denoised video signal.

B. Order of Operations

With a spatial denoising technique and a temporal denoising technique established in Sections II and above, respectively, there still remains the question of the order of operations. The highest quality may occur with temporal-domain denoising followed by spatial-domain (TFS) denoising, or spatial denoising followed by temporal (SFT) denoising.

Theoretically, is it not possible to prove and determine which operation is better because the description of the noise is not known. However, it is our hypothesis that SFT denoising can more aptly determine noise from signal information. The reasoning behind this hypothesis is that removing noise in the spatial domain is a well known process, and any noise removal prior to temporal-domain processing is helpful in discriminating between the residual noise and motion in the image sequence. However, a validation of this hypothesis is determined heuristically.

Thus, a simple test is conducted with two test video signals. The first video signal is one which contains little motion, and the other contains a great deal of motion. The selected image sequences are the "CLAIRE" sequence from frame #104-167 and the "FOOTBALL" sequence from frame #33-96.

Both of the image sequences are denoised with τ and τ_z ranging from 0–30 for both TFS and SFT denoising operations. Note that in the test, τ_z is a single value and spatially independent, unlike the temporal threshold used in the final denoising algorithms $\tau_z[\cdot]$ which is dependent upon spatial position. Also, the s parameter for feature selection in the image denoising method described in Section II is calculated by taking (5) and (6) and solving for s. The parameter s is given by

$$s = \left\lfloor \frac{a_s}{a_\tau}(\tau - b_\tau) + b_s \right\rfloor. \tag{15}$$

Also, the number of resolutions of the nondecimated wavelet transform used in both the spatial and temporal denoising methods is $k = 1, \ldots, 5$. The average peak signal-to-noise ratio (PSNR) of each trial is recorded. The PSNR of an image is given by

$$\text{PSNR}_z = 20\log_{10}\left(\frac{255}{\sqrt{\text{mse}_z}}\right) \tag{16}$$

where

$$\text{mse}_z = \frac{1}{L}\sum_l \left(\hat{f}_l^z - f_l^z\right)^2 \tag{17}$$

where L is the size of the image, $\hat{f}_l^z$ is the denoised pixel of spatial location l and frame z, and f_l^z is the corresponding pixel of the original signal.

Fig. 3 gives the results of testing. As shown in Fig. 3, the highest average PSNR is achieved by SFT denoising; first spatially denoising each frame of the sequence followed by temporal-domain denoising. Thus, for the proposed denoising

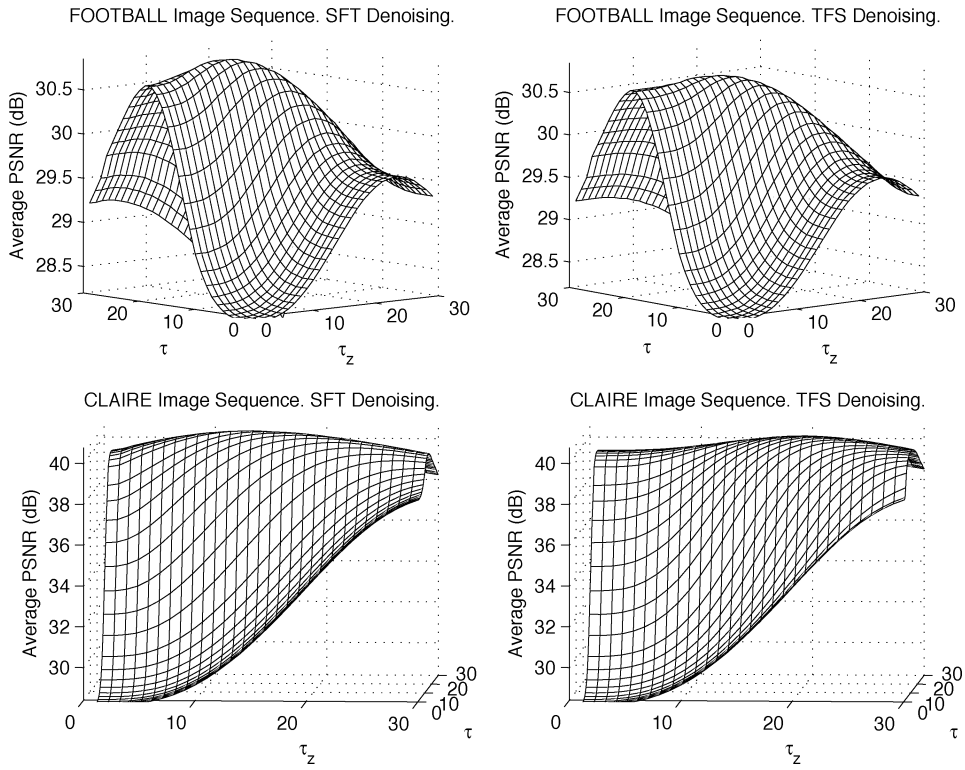

Fig. 3. Test results of both TFS and SFT denoising methods. Upper left: FOOTBALL image sequence, SFT denoising, max. PSNR $= 30.85$, $\tau = 18$, and $\tau_z = 12$. Upper right: FOOTBALL image sequence, TFS denoising, max. PSNR $= 30.71$, $\tau = 18$, and $\tau_z = 12$. Lower left: CLAIRE image sequence, SFT denoising, max. PSNR $= 40.77$, $\tau = 19$, and $\tau_z = 15$. Lower right: CLAIRE image sequence, TFS denoising, max. PSNR $= 40.69$, $\tau = 15$,and $\tau_z = 21$.

method, spatial-domain denoising occurs prior to temporal-domain denoising, exclusively.

In addition to a higher average PSNR, there is another benefit to SFT denoising. The level of motion in an image sequence is known to be crucial in determining the amount of noise reduction possible from temporal-domain processing, and a motion index calculation is inevitably done by comparing consecutive frames to one another. Thus, let us define a noisy image sequence where $\hat{f}_l^z$ is a corrupted pixel in spatial position l and frame z and is defined by

$$\hat{f}_l^z = f_l^z + \eta_l^z \tag{18}$$

where f_l^z is the noiseless pixel value, and η_l^z is the noise function. We can compare consecutive frames by taking the difference as in [3] and [22] to find

$$\hat{f}_l^z - \hat{f}_l^{z+1} = \left(f_l^z - f_l^{z+1}\right) + \left(\eta_l^z - \eta_l^{z+1}\right) = \Delta f_l^z + \Delta\eta_l^z. \tag{19}$$

Thus, by taking the difference between frames to find the level of motion, the noise function is subtracted from itself, in effect doubling the amount of noise corruption [26]. Therefore, by applying spatial denoising prior to motion index calculation we can reduce the value of $\Delta\eta_l^z$ and provide a more precise calculation of the motion given in the image sequence.

IV. Proposed Motion Index

A motion index is important in the success of a video denoising method in order to discriminate between large temporal variances in the video signal which are caused by noise and large temporal variances which are caused by motion in the original (noiseless) signal. A motion index is able to aid temporal denoising algorithms to eliminate the large temporal variances caused by noise while preserving the temporal variances caused by motion in the original image sequence, creating a higher quality video signal. That is, the motion index is used to determine $\tau_z[\cdot]$.

A. Motion Index Calculation

Several works have developed a motion estimation index to determine the amount of temporal-domain processing to perform, i.e., the amount of information that can be removed from the original signal to improve the overall quality [3], [22]. However, none of these proposed indices are robust to noise corruption, which is an important feature in a motion index. There are a few characteristics that a motion index must possess. One, a motion index should be a localized value. The reasoning behind a localized motion index is because the amount of motion may vary in different spatial portions of an image sequence. Thus, the motion index should be able to identify those differences. Two, a motion index needs to be unaffected by the amount of noise corruption in a given video signal. A motion index should be robust to noise corruption to aptly determine the proper amount of temporal-domain processing.

Thus, a localized motion index is developed which is relatively unaffected by the level of noise corruption in the original image sequence. A spatially averaged temporal standard deviation (SATSD) is used as the index of motion. Spatial averaging is used to remove the noise inherent in the signal, and the temporal standard deviation is used to detect the amount of activity in the temporal domain.

Fig. 4. Spatial positions of motion estimation test points. Left: FOOTBALL image sequence, frame #96. Right: CLAIRE image sequence, frame #167.

Let us define $\hat{f}_l^{z,2-D}$ as pixel value in the spatial location l of the zth frame of an image sequence already processed by the 2-D denoising method of [1]. The spatial averaging of the spatially denoised signal is given by

$$A_l^z = \frac{1}{B^2} \sum_{i \in I} \hat{f}_i^{z,2-D} \tag{20}$$

where I is the set of spatial locations which form a square area centered around spatial location l, and B^2 is the number of spatial locations contained in I; typically, $B = 15$. The value of B must be an odd value to allow for the square area to set centrally around spatial location l. This average is used to find the standard deviation in the temporal domain

$$\mu_l = \frac{1}{F} \sum_{i=1}^{F} A_l^i \tag{21}$$

and

$$M_l = \sqrt{\frac{1}{F} \sum_{i=1}^{F} \left(A_l^i - \mu_l\right)^2} \tag{22}$$

where M_l is the localized motion index, F is the number of frames in the image sequence, and μ_l is the temporal mean of the spatial average at location l.

B. Motion Index Testing

The FOOTBALL and CLAIRE image sequences are used once more to test the proposed motion index as well as the motion index given in [3], and two specific spatial locations are selected from each sequence: a location where there is little to no motion present, and a location where motion is present. A frame from each of the two image sequences is given in Fig. 4, and the four spatial locations for evaluation of the proposed motion index are highlighted.

The two sequences are corrupted with various levels of noise, and the motion is estimated at each of the four spatial locations selected with both the proposed motion index and that of [3]. The results of the motion index used in [3] is given in Fig. 5. As shown in Fig. 5, the motion index of [3] is not robust to noise corruption. That is, the motion calculation from the same spatial location increases with an increase in noise. Also, the motion index shows the FOOTBALL image sequence

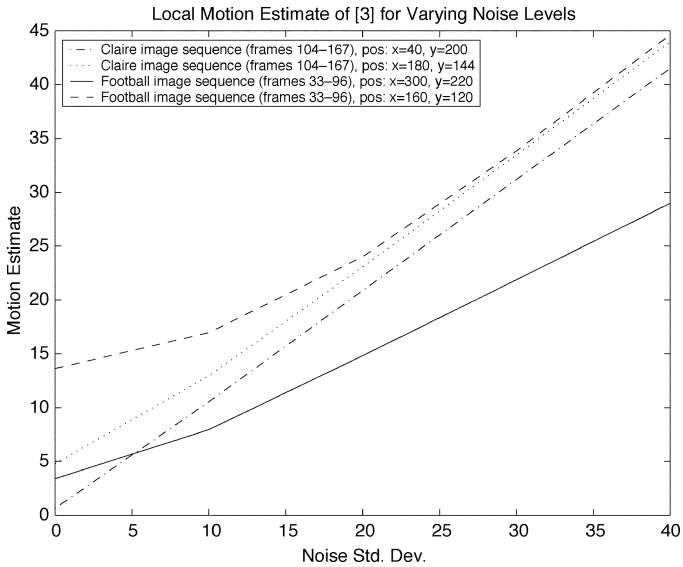

Fig. 5. Motion estimate given in [3] of image sequences, CLAIRE and FOOTBALL.

$(x = 300, y = 220)$ as having a higher motion index than the CLAIRE image sequence $(x = 40, y = 200)$ with zero noise corruption. However, the motion index shows the opposite results with higher levels of noise. Thus, the motion index gives conflicting results with the introduction of noise.

The results of the proposed SATSD motion index are given in Fig. 6. As shown in Fig. 6, the proposed motion index is much more robust to varying noise levels, and the order of locations from highest to lowest motion is what one would believe is correct. The location with the lowest motion index is in the CLAIRE image sequence where there is no camera motion, and there are no moving objects in that spatial location. The next lowest motion location is in the FOOTBALL image sequence in the spatial location where there are no moving objects. However, there is some slight camera motion in the sequence, so the motion index is slightly higher than in the CLAIRE image sequence. The location with the next highest motion index is the center of the CLAIRE image sequence, where there is some motion due to movement of the head, and the location with the highest motion index is the FOOTBALL image sequence in the spatial location where many objects cross.

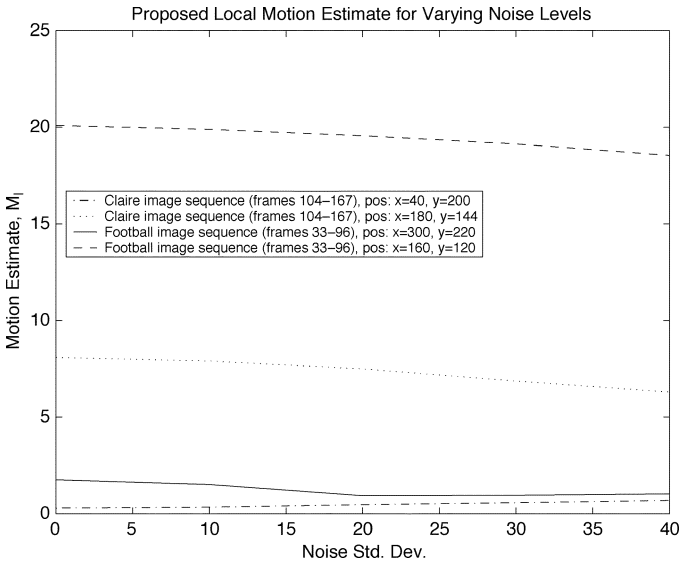

Fig. 6. Proposed motion estimate of image sequences, CLAIRE and FOOTBALL.

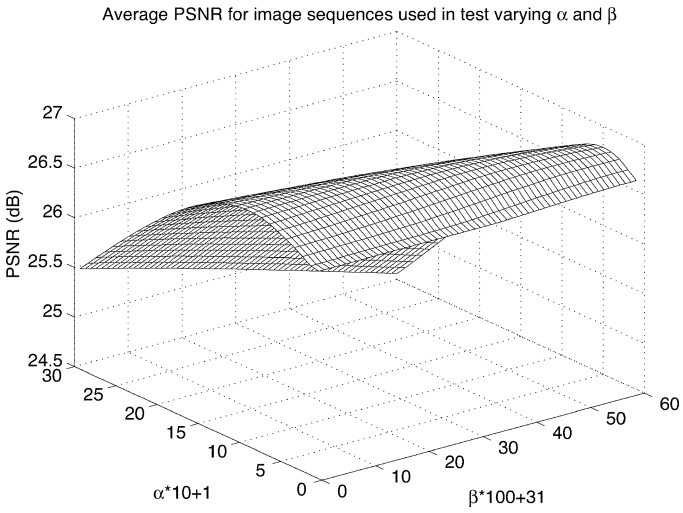

Fig. 7. α and β parameter testing for temporal-domain denoising.

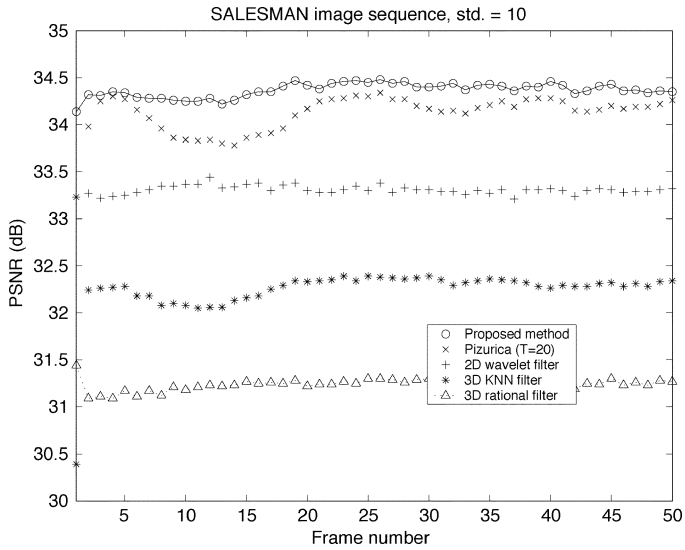

Fig. 8. Denoising methods applied to the SALESMAN image sequence, std = 10.

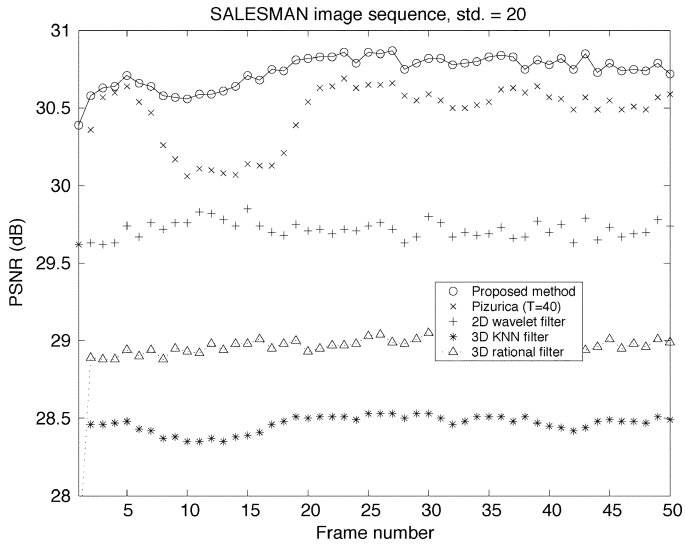

Fig. 9. Denoising methods applied to the SALESMAN image sequence, std = 20.

V. Temporal-Domain Parameter Selection

The amount of temporal denoising which is beneficial to an image sequence is dependent upon the amount of noise corruption as well as the amount of motion. Thus, the threshold $\tau_z[l]$ is given by

$$\tau_z[l] = \alpha\widetilde{\sigma_n} + \beta M_l \tag{23}$$

where M_l is the motion index of spatial position l, and $\widetilde{\sigma_n}$ is the estimated noise standard deviation of the image sequence. The two parameters α and β are determined experimentally using test image sequences.

In the proposed coefficient selection method, we use a *training sample* approach. The approach starts with a series of test image sequences serving as training samples to derive the functions which determine the optimal set of the values for α and β. Theoretically, we may represent each training sample as a vector V_i, $i = 1, n$. Those training samples should span a space which covers more corrupted image sequences than the training samples

$$S = \text{Span}\,\{V_i; i = 1, \ldots n\}. \tag{24}$$

The original data and the statistical distribution of the noise are given for each of the training samples which are corrupted. The optimal set of parameters can then be determined for the training samples using the approach described earlier. Ideally, the space spanned by the training samples contains the type of the corrupted image sequences which are to be denoised. As a result, the same set can generate an optimal or close to optimal performance for the corrupted image sequences of same type. It is clear that more training samples will generate parameters suitable for more types of image sequences, while a space of fewer training samples is suitable for fewer types of image sequences.

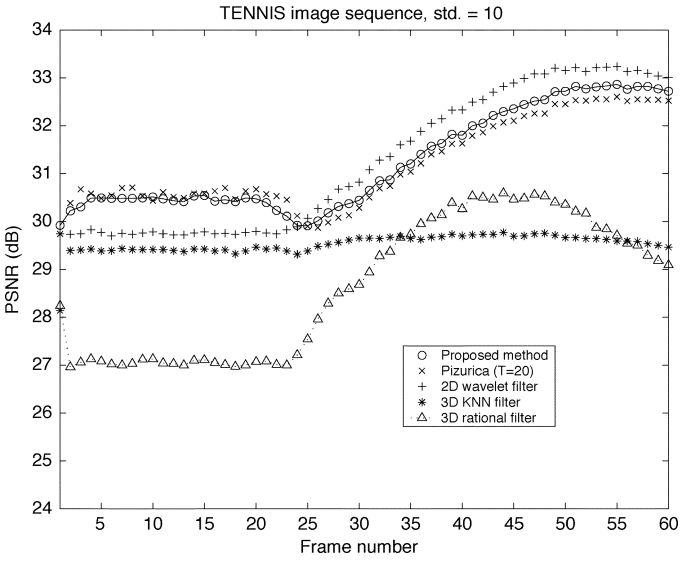

Fig. 10. Denoising methods applied to the TENNIS image sequence, std $= 10$.

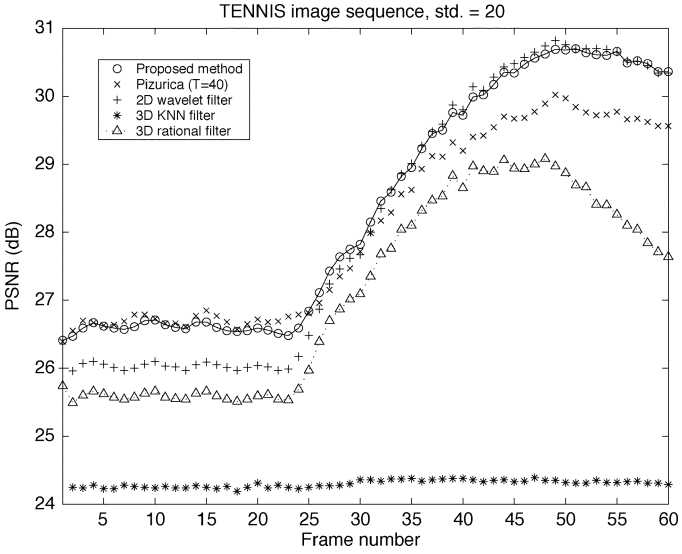

Fig. 11. Denoising methods applied to the TENNIS image sequence, std $= 20$.

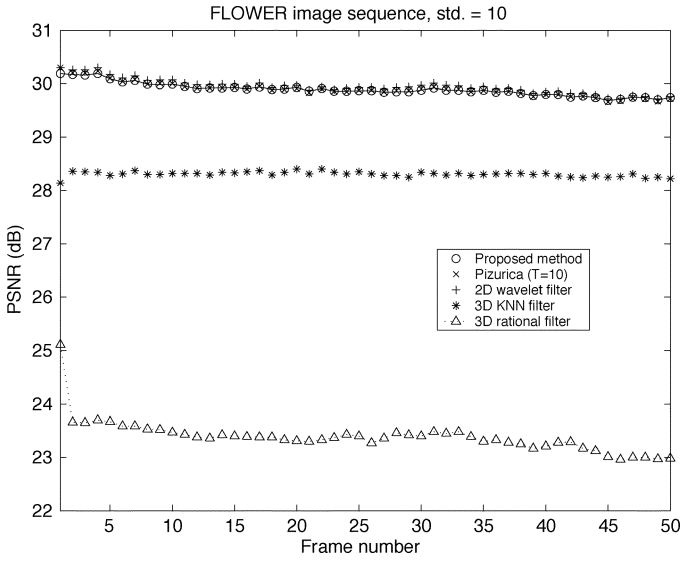

Fig. 12. Denoising methods applied to the FLOWER image sequence, std $= 10$.

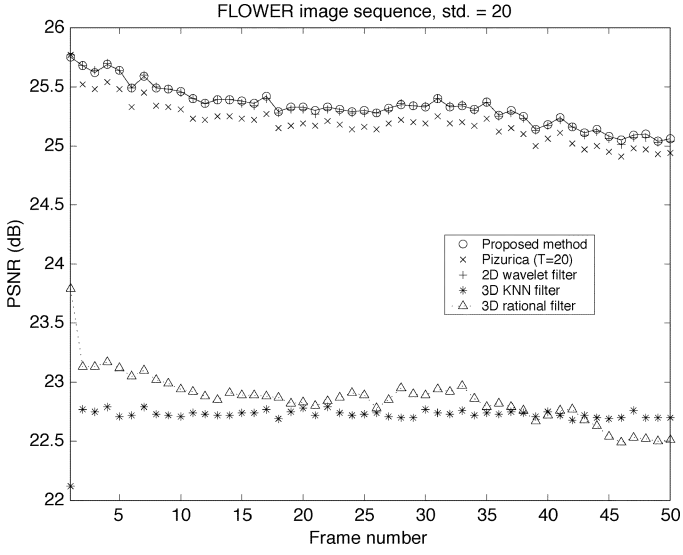

Fig. 13. Denoising methods applied to the FLOWER image sequence, std $= 20$.

In order to obtain an estimate of the noise level $\widetilde{\sigma_n}$ an average is taken from the noise estimates of each frame in the image sequence, given by (7). It is reasonable to assume an independent, identically distributed (IID) model for the level of noise for each pixel position since noise in each pixel position is generated by individual sensing units of the image sensor such as charge-coupled devices (CCDs) [8] which are independent. As a result, the estimate of the standard deviation of the noise (σ_n) in each image also represents the standard deviation of the noise in the temporal domain. Therefore, we can use the estimate of the noise in the spatial domain to estimate that in the temporal domain.

It should be pointed out that after the denoising has occurred in the spatial domain using the SFT method, the standard deviation of the noise is significantly reduced. That reduction is statistically equal to each frame. As a result, the estimated noise in the spatial domain can still be nominally used for noise reduction in the temporal domain as the reduction of σ_n can be automatically absorbed by α.

The sequences CLAIRE, FOOTBALL, and TREVOR are used for α and β selection. Each of the image sequences are corrupted with differing levels of noise corruption ($\sigma_n = 10$,20) and denoised with the SFT denoising method where (23) is used as the temporal-domain threshold. Values of α and β are used ranging from $\alpha = 0$ to 3.0 and $\beta = -0.3$ to 0.3. The results of this testing is given in Fig. 7. As shown in Fig. 7 the maximum average PSNR is achieved when $\alpha = 0.9$ and $\beta = -0.11$. The result is reasonable, of course, because as the motion increases in an image sequence the redundancy between frames decreases, and the benefits of temporal-domain processing decrease. Thus, as the testing has shown, the temporal-domain threshold decreases as the motion increases. Fig. 7 can be

deceiving, however, because it seems that the average PSNR is more greatly affected by changes in α than β. However, (23)

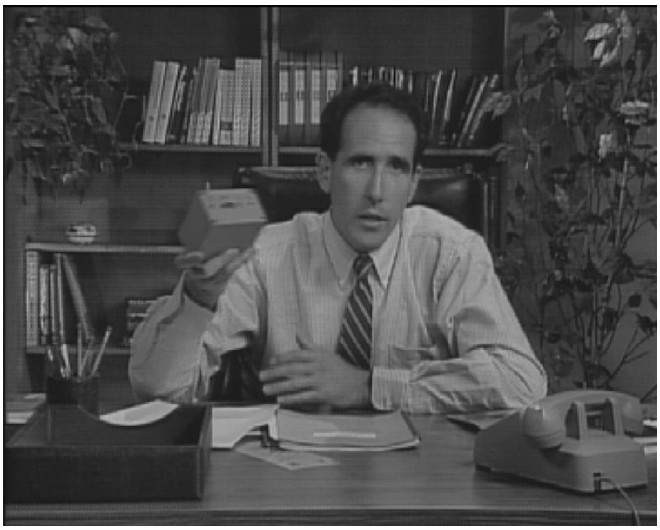

Fig. 14. Original frame #7 of the SALESMAN image sequence.

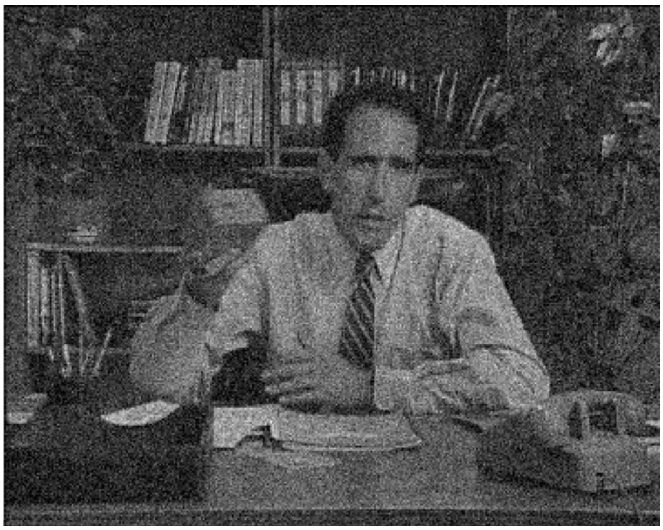

Fig. 15. SALESMAN image sequence corrupted std $= 20$ and PSNR $= 22.10$.

shows that the PSNR should be equally effected by changes in both α and β, assuming that on average noise variance estimates and motion estimates carry similar value. The reason that α seems to be a greater factor in determining the PSNR is that on this particular test, the range of α far exceeds the range of β. These ranges were experimentally chosen to find the peak average PSNR.

VI. Experimental Results

The proposed video denoising algorithm first is applied to each of the video frames individually and independently. The method of [1] was developed earlier by our previous research to denoise images, and is used as the spatial denoising portion of the wavelet-based video denoising algorithms.

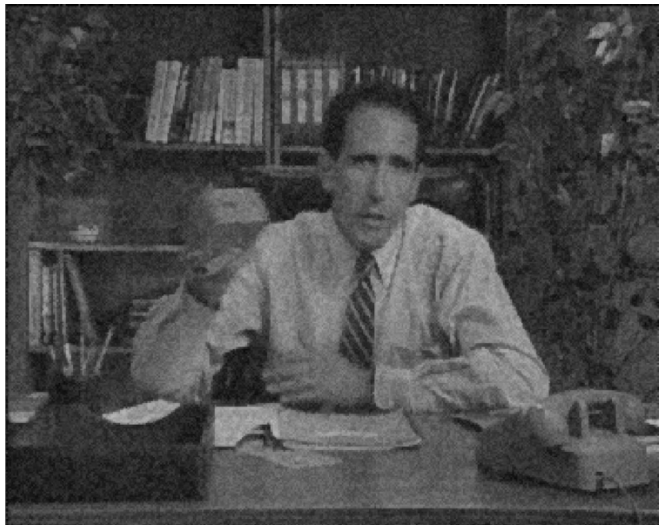

Fig. 16. Results of the 3-D K-nearest neighbors filter [32], PSNR $= 28.42$.

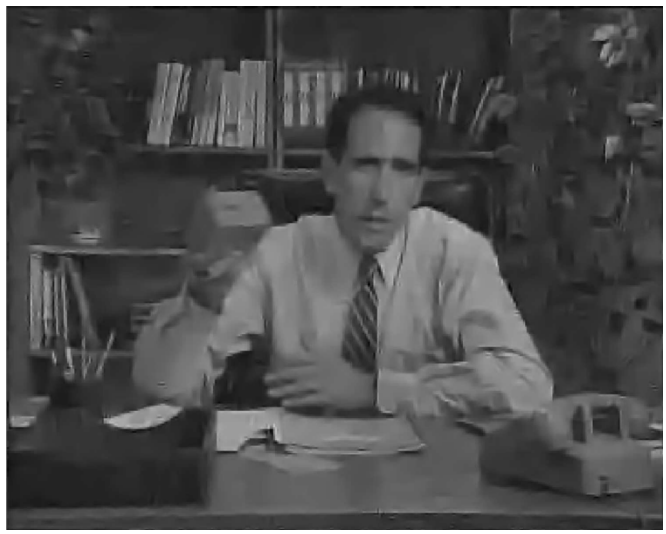

Fig. 17. Results of the 2-D wavelet denoising filter [1], PSNR $= 29.76$.

The video signal is then denoised in the temporal domain by the method developed in Sections III and V. The temporal denoising algorithm is a selective shrinkage algorithm which uses a proposed motion estimation index to determine the temporal threshold, $\tau_z[\cdot]$. The temporal threshold is modified by the motion index to effectively eliminate temporal-domain noise while preserving important motion information.

Three image sequences are used to determine the effectiveness of the proposed video denoising method. They are the SALESMAN image sequence, the TENNIS image sequence, and the FLOWER image sequence. These three sequences are all corrupted with various levels of noise and denoised with the methods of [1], [3], [22], [32] as well as the proposed method. Please note that only the temporal-domain denoising algorithm of [22] is being tested. The spatial-domain denoising methods of [1] is used for all the wavelet-based video denoising methods. The results are given in Figs. 8–13. As shown in Figs. 8–13, the proposed method consistently outperforms the other methods presented. In all cases, the proposed denoising method has a higher average PSNR then all other denoising

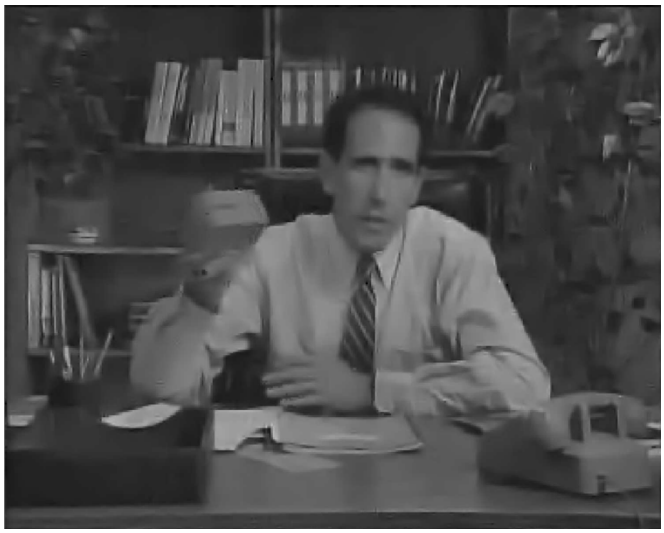

Fig. 18. Results of the 2-D wavelet filtering with linear temporal filtering, [22], PSNR $= 30.47$.

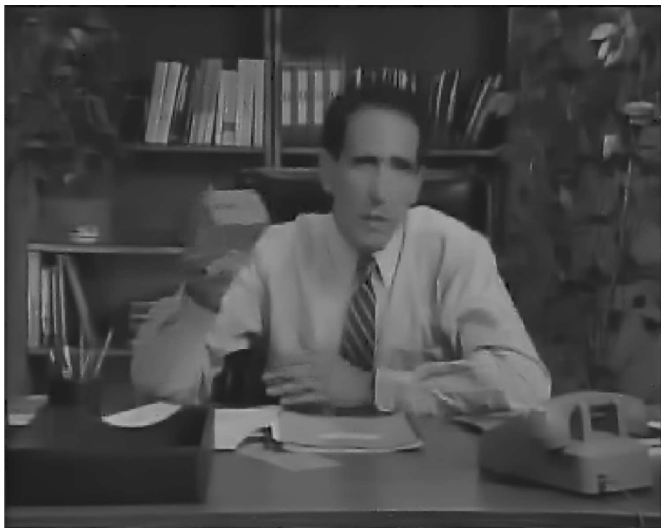

Fig. 19. Results of the proposed denoising method PSNR $= 30.66$.

methods tested. Also, note that in the method of [22], the threshold T changes due to video content and noise level to obtain the highest average PSNR using that particular method. In the proposed method, the temporal-domain threshold is automatically calculated due to estimates of the noise level and motion.

Figs. 14–19 give an example of the effectiveness of each of the denoising methods. Fig. 14 gives the original frame #7 of the SALESMAN image sequence, and Fig. 15 gives frame #7 corrupted with noise. Figs. 16–19 give frame #7 denoised by each of the methods mentioned in the section.

VII. Conclusion

In this paper, a new combined spatial and temporal-domain wavelet shrinkage method is developed for the removal of noise in video signals. The proposed method uses a geometrical approach to spatial-domain denoising to preserve edge information, and a newly developed motion estimation index for selective wavelet shrinkage in the temporal domain.

The spatial denoising technique is a selective wavelet shrinkage algorithm developed in [1] and is shown to outperform other wavelet shrinkage denoising algorithms given in the literature in denoised image quality.

The temporal denoising algorithm is also a selective wavelet shrinkage algorithm which uses a motion estimation index to determine the level of thresholding in the temporal domain.

The proposed motion index is experimentally determined to be more robust to noise corruption than other methods, and is able to help determine the threshold value for selective wavelet shrinkage in the temporal domain. With the motion index and temporal-domain wavelet shrinkage, the proposed video denoising method is experimentally proven to outperform other methods given in the literature for various levels of noise corruption applied to video signals with varying amounts of motion.

References

[1] E. J. Balster, Y. F. Zheng, and R. L. Ewing, "Feature-based wavelet shrinkage algorithm for image denoising," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2024–2039, Dec. 2005.

[2] J. B. Bednar and T. L. Wat, "Alpha-trimmed means and their relationship to median filters," *IEEE Trans. Acoust., Speech, Signal Proces.*, vol. ASSP-32, no. 2, pp. 145–153, Feb. 1984.

[3] F. Cocchia, S. Carrato, and G. Ramponi, "Design and real-time implementation of a 3-D rational filter for edge preserving smoothing," *IEEE Trans. Consum. Electron.*, vol. 43, no. 11, pp. 1291–1300, Nov. 1997.

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

[5] R. Dugad and N. Ahuja, "Video denoising by combining Kalman and Wiener estimates," in *Proc. IEEE Int. Conf. Image Processing*, vol. 4, 1999, pp. 152–156.

[6] F. Faghih and M. Smith, "Combining spatial and scale-space techniques for edge detection to provide a spatially adaptive wavelet-based noise filtering algorithm," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1062–1071, Sep. 2002.

[7] M. Ghazel, G. H. Freeman, and E. R. Vrscay, "Fractal-wavelet image denoising," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, 2002, pp. I836–I839.

[8] G. Healey and R. Kondepudy, "CCD camera calibration and noise estimation," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognition*, vol. 1, Jun. 1992, p. 90.

[9] T. C. Hsung, D. P.-K. Lun, and W. C. Siu, "Denoising by singularity detection," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3139–3144, Nov. 1999.

[10] S. J. Huang, "Adaptive noise reduction and image sharpening for digital video compression," in *Proc. IEEE Int. Conf. Computational Cybernetics Simulation*, vol. 4, 1997, pp. 3142–3147.

[11] C. R. Jung and J. Scharcanski, "Adaptive image denoising in scale-space using the wavelet transform," in *Proc. XIV Brazilian Symp. Computer Graphics and Image Process.*, 2001, pp. 172–178.

[12] S. D. Kim, S. K. Jang, M. J. Kim, and J. B. Ra, "Efficient block-based coding of noise images by combining pre-filtering and DCT," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 4, 1999, pp. 37–40.

[13] W. Ling and P. K. S. Tam, "Video denoising using fuzzy-connectedness principles," in *Proc. 2001 Int. Symp. Intelligent Multimedia, Video, Speech Process.*, 2001, pp. 531–534.

[14] W. S. Lu, "Wavelet approaches to still image denoising," in *Proc. Alisomar Conf. Signals, Syst., Comput.*, vol. 2, 1998, pp. 1705–1709.

[15] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Trans. Image Process.*, vol. 6, no. 4, pp. 549–565, Apr. 1997.

[16] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inf. Theory*, vol. 38, no. , pp. 617–623, Mar. 1992.

[17] M. Meguro, A. Taguchi, and N. Hamada, "Data-dependent weighted median filtering with robust motion information for image sequence restoration," *IEICE Trans. Fundamentals*, vol. 2, pp. 424–428, 2001.

[18] O. Ojo and T. Kwaaitaal-Spassova, "An algorithm for integrated noise reduction and sharpness enhancement," *IEEE Trans. Consum. Electron.*, vol. 46, pp. 474–480, May 2000.

[19] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, July 1990.
[20] R. A. Peters, "A new algorithm for image noise reduction using mathematical morphology," *IEEE Trans. Image Process.*, vol. 4, no. 5, pp. 554–568, May 1995.
[21] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for bayesian wavelet based image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 5, pp. 545–557, May 2002.
[22] A. Pizurica, V. Zlokolica, and W. Philips, "Combined wavelet domain and temporal video denoising," in *Proc. IEEE Int. Conf. Advanced Video Signal Based Surveillance*, vol. 1, Jul. 2003, pp. 334–341.
[23] P. Rieder and G. Scheffler, "New concepts on denoising and sharpening of video signals," *IEEE Trans. Consum. Electron.*, vol. 47, no. 8, pp. 666–671, Aug. 2001.
[24] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, Jun. 1996.
[25] L. Shutao, W. Yaonan, Z. Changfan, and M. Jianxu, "Fuzzy filter based on neural network and its applications to image restoration," in *Proc. IEEE Int. Conf. Signal Process.*, vol. 2, 2000, pp. 1133–1138.
[26] H. Stark and J. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*: Prentice-Hall, Englewood Cliffs, NJ, 1994.
[27] A. De Stefano, P. R. White, and W. B. Collis, "An innovative approach for spatial video noise reduction using a wavelet based frequency decomposition," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, 2000, pp. 281–284.
[28] C. Vertan, C. I. Vertan, and V. Buzuloiu, "Reduced computation genetic algorithm for noise removal," in *Proc. IEEE Conf. Image Process. Its Applications*, vol. 1, Jul. 1997, pp. 313–316.
[29] Y. F. Wong, E. Viscito, and E. Linzer, "PreProcessing of video signals for MPEG coding by clustering filter," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, 1995, pp. 2129–2133.
[30] Y. I. Wong, "Nonlinear scale-space filtering and multiresolution system," *IEEE Trans. Image Process.*, vol. 4, pp. 774–786, Jun. 1995.
[31] X.-P. Zhang, "Thresholding neural network for adaptive noise reduction," *IEEE Trans. Neural Netw.*, vol. 12, pp. 567–584, May 2001.
[32] V. Zlokolica, W. Philips, and D. Van De Ville, "A new nonlinear filter for video processing," in *Proc. IEEE Benelux Signal Processing Symp.*, vol. 2, 2002, pp. 221–224.

Eric J. Balster (S'02–M'04) received the B.S. and M.S. degrees from the University of Dayton, Dayton, OH, and the Ph.D. degree from The Ohio State University, Columbus, all in electrical engineering in 1998, 2000, and 2004, respectively.

In 2002, he joined the Air Force Research Laboratory's Information Directorate, Wright-Patterson Air Force Base, Dayton. Currently, he is continuing his research in the areas of image and video compression, and noise removal in image and video signals.

Yuan F. Zheng (S'82–M'86–SM'90–F'97) received the B.S. degree from Tsinghua University, Beijing, China, in 1970, and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 1980 and 1984, respectively.

From 1984 to 1989, he was with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. Since August 1989, he has been with The Ohio State University, where he is currently Winbigler Professor and was the Chairman, Electrical and Computer Engineering Department between 1993 and 2004. His research interests include two aspects. One is wavelet transform for image and video compression for internet and satellite communications. His current efforts focus on content-based compression, 3-D wavelet transformation, video object tracking, and content-based retransmission in Internet communications. The other is in robotics, which includes robots for biological applications, multiple robots coordination, legged robots, human–robot coordination, and personal robotics. He is currently on the Editorial Boards of *International Journal of Multimedia Tools and Applications*, *Autonomous Robots*, *International Journal of Intelligent Control and Systems*, and *International Journal of Control, Automation, and Systems*, and is an Associate Editor of the *International Journal of Intelligent Automation and Soft Computing*

Dr. Zheng was Vice-President for Technical Affairs of the IEEE Robotics and Automation Society from 1996 to 1999. He was an Associate Editor of the IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION between 1995 and 1997. He was the Program Chair of the 1999 IEEE International Conference on Robotics and Automation, held in Detroit, MI, on May 10–15, 1999, and received the Presidential Young Investigator Award in 1986.

Robert L. Ewing (S'77–M'87–SM'02) received the B.S.E.E. degree and the M.S. degree in physics from the University of Cincinnati, Cincinnati, OH, and the Ph.D. degree in electrical engineering from the University of Dayton, Dayton, OH.

He began his career in the Propulsion Laboratory at Wright Patterson AFB during the early 1970s with the development of jet engine control systems and the initial control system used on the F-15 aircraft. He was with the University of Cincinnati's medical school in the area of electronic control and regeneration of peripheral (sciatic) nerves used in walking. From 1977 to 1982, he was a Medical Research Scientist at the Aerospace Medical Research Laboratory, Biodynamic Effects Division. He worked to develop the pilot's analog and digital flight control interfacing and aircraft ejection systems for low-level, high-speed flight. In 1982, he became an Instructor for the Army, at the Air Force Institute of Technology (AFIT) and an adjunct instructor at Wright State University. During his work at the AFIT, he developed many of the early short courses and classes in robotics, digital control, artificial intelligence, neural nets, database systems, low observables (radar), navigation and guidance systems, microprocessor design and microelectromechanical devices (MEMS). In 1993, he was with the Wright Laboratory's Solid State Electronic Devices Directorate in the area of hardware description language (VHDL) for very-large scale integrated synthesis. Currently, he is the Technical Advisor for the Information Directorate's Embedded Information Systems Branch, and also the Director of the Computer Engineering Research Consortium (CERC) of local universities in the area of embedded system design (1996 to present). He is working toward the development and use of bio-inspired intelligent information processing and data compression for embedded systems and synthesis (IIP). He has been a registered Professional Engineer with the state of Ohio since 1984, and is currently an adjunct professor at AFIT and Wright State University.