Noise Estimation for Video Processing Based on Spatio–Temporal Gradients

Vladimir Zlokolica, *Student Member, IEEE*, Aleksandra Pižurica, *Member, IEEE*, and Wilfried Philips, *Member, IEEE*

***Abstract*—We propose an efficient and accurate wavelet-based noise estimation method for white Gaussian noise in video sequences. The proposed method analyzes the distribution of spatial and temporal gradients in the video sequence in order to estimate the noise variance. The estimate is derived from the most frequent gradient in the two distributions and is compensated for the errors due to the spatio–temporal image sequence content, by a novel correction function. The spatial and temporal gradients are determined from the finest scale of the spatial and temporal wavelet transform, respectively. The main application of the noise estimation algorithm is in wavelet-based video processing. The results show that the proposed method is more accurate than other state-of-the-art noise estimation techniques and less sensitive to varying spatio–temporal content and noise level.**

***Index Terms*—Noise estimation, video analysis and processing, wavelets.**

I. Introduction

VIDEO sequences are often distorted by noise during acquisition or transmission. In many video processing applications, such as video quality enhancement, compression, format conversion, deinterlacing, motion segmentation, etc., accurate knowledge of the noise level present in the input video sequence is of crucial importance for tuning the parameters of the corresponding video processing algorithm. We assume the additive white Gaussian noise model, which is of interest in many video applications [1]. Given a noisy video sequence

$$I_\eta(\mathbf{r},t) = I(\mathbf{r},t) + \eta(\mathbf{r},t) \tag{1}$$

the noise estimation problem is to estimate the standard deviation σ_η of the noise $\eta(\mathbf{r},t)$, i.e., to distinguish noise from the changes due to the spatio-temporal image sequence structure. In (1), $\mathbf{r} = (m,n)$ denotes the discrete spatial (m horizontal and n vertical) coordinate, and t denotes the frame index. Additionally, $I_\eta(\mathbf{r},t)$, and $I(\mathbf{r},t)$ stand for the noisy and original image sequence frame t.

In this letter, we propose a novel gradient-based noise estimator in the wavelet domain, which exploits both the temporal and the spatial correlations (gradient magnitudes) in the sequence. Our initial noise estimate is proportional to the value at which the spatial or temporal gradient-histogram reaches its maximum. The decision of whether to use the spatial or temporal gradient histogram is based on the deviation of the gradient-histogram from the Rayleigh distribution, and so is the correction of the initial estimate. The implementation of these ideas is an efficient scheme suitable for real-time applications. The experimental results show that the proposed method is more accurate than the state-of-the-art techniques and less sensitive to varying noise levels and the presence of spatio–temporal sequence content.

This letter is organized as follows. In Section II, we present an overview of the existing noise estimation techniques. We explain our method for noise estimation in Section III-A and give implementation details in Section III-B. In Section IV, we present experimental results, and we conclude the letter in Section V.

II. Related Work

In the past, a number of different methods have been proposed for noise variance estimation in still images and video, e.g., [1]–[9]. Some of those noise estimation approaches were evaluated and compared in [10], where it was concluded that the *averaging* approach, which is based on first suppressing image structures by prefiltering and then computing the noise variance, provides the most reliable results for a wide range of noise levels and images with different content. Recently, in [8], a block-based noise estimation method was proposed, with a new measure for determining intensity-homogeneous blocks and a structure analyzer for rejecting blocks with structure. They showed to outperform the methods of [4] and [10].

Gradient-based approaches [3], [6] analyze the distribution of the gradient magnitudes in the noisy image. The gradient amplitudes G are determined in terms of horizontal and vertical gradient component values g_x and g_y, where $G = \sqrt{g_x^2 + g_y^2}$. The idea is that in the case of an ideally uniform image with added white Gaussian noise, the two gradient components g_x and g_y are independent white Gaussian processes, thus yielding the Rayleigh distribution for the gradient magnitude G. However, for typical images, which are not ideally uniform, the actual distribution of the gradient magnitudes differs from the Rayleigh distribution, which consequently introduces errors in the noise estimation approach. To our knowledge, no efficient solutions have been proposed for compensating for these errors. In our earlier work [6], we tried to find the optimal correspondence between the gradient value at which the *gradient histogram peaks* (most frequent gradient) and the estimated standard deviation of noise, in the least-square sense, across the training set of sequences and for all noise levels.

Manuscript received September 13, 2005; revised December 6, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Maja Bystrom.

V. Zlokolica and W. Philips are with the Department for Telecommunications and Information Processing (TELIN), Ghent University, B-9000 Ghent, Belgium (e-mail: Vladimir.Zlokolica@telin.UGent.be; philips@telin.UGent.be).

A. Pižurica is with the Department for Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium and also with the Fund for the Scientific Research Flanders (FWO) (e-mail: sanja@telin.UGent.be).

Digital Object Identifier 10.1109/LSP.2006.870481

TABLE I
AVERAGE ERROR $\overline{E}$ AND THE STANDARD DEVIATION σ_E OF THE ERROR FOR EIGHT SEQUENCES AND VARYING INPUT NOISE LEVEL (STANDARD DEVIATION σ_η) AND AVERAGED OVER FIRST 50 FRAMES (FIRST 14 COLUMNS CORRESPOND TO THE SEQUENCES IN THE PROGRESSIVE FORMAT AND LAST TWO COLUMNS TO THE SEQUENCES IN THE INTERLACE FORMAT). HIGHLIGHTED NUMBERS STAND FOR THE BEST RESULTS IN EACH ROW

Noise standard deviation	**New proposed method**		**Spatial gradient method** [6]		**Donoho MAD** [13]		**Moment Matching method**[7]		**CDF method** [7]		**Temporal method** [2]		**Structure Oriented** [8]		**New (interlaced) proposed method**	
σ_η	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E
0	**0.18**	**0.14**	1.65	1.07	3.77	2.02	0.23	1.78	2.56	2.93	2.11	0.92	1.13	0.41	0.22	0.15
5	0.61	0.24	0.87	0.81	2.21	1.54	2.14	1.79	1.84	1.09	1.27	0.63	**0.58**	**0.16**	0.76	0.52
10	0.62	**0.18**	1.49	0.58	1.61	1.56	1.22	1.37	1.14	1.01	0.94	0.66	**0.61**	0.31	0.94	0.53
15	**0.53**	**0.22**	1.27	0.54	1.54	1.62	0.89	0.79	0.93	0.89	0.77	0.61	0.77	0.29	0.95	0.45
20	**0.46**	**0.18**	1.34	0.64	0.99	0.97	0.96	0.53	0.77	0.71	0.86	0.43	1.22	0.23	0.83	0.33
25	**0.52**	**0.25**	1.58	0.36	1.25	0.95	1.44	0.53	0.71	0.47	1.18	0.78	1.79	0.45	0.76	0.34
30	0.78	**0.29**	2.58	0.89	1.20	0.75	3.39	2.31	**0.63**	0.36	2.91	1.36	2.55	0.83	1.07	0.36
average	**0.52**	**0.21**	1.43	0.68	1.75	1.34	1.46	1.18	1.22	1.06	1.44	0.77	1.23	0.38	0.79	0.38

TABLE II
AVERAGE ERROR $\overline{E}$ AND THE STANDARD DEVIATION σ_E OF THE ERROR FOR ALL NOISE LEVELS σ_η PER SEQUENCE AND AVERAGED OVER FIRST 50 FRAMES (FIRST 14 COLUMNS CORRESPOND TO THE SEQUENCES IN THE PROGRESSIVE FORMAT AND LAST TWO COLUMNS TO THE SEQUENCES IN THE INTERLACE FORMAT). HIGHLIGHTED NUMBERS STAND FOR THE BEST RESULTS IN EACH ROW

Image Sequence	**New proposed method**		**Spatial gradient method** [6]		**Donoho MAD** [13]		**Moment Matching method**[7]		**CDF method** [7]		**Temporal method** [2]		**Structure Oriented** [8]		**New (interlaced) proposed method**	
	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E	$\overline{E}$	σ_E
Salesman	**0.50**	**0.29**	1.53	0.92	1.16	1.05	0.97	0.84	0.38	0.28	1.22	0.97	1.41	0.95	0.51	0.31
FlowerGar.	**0.59**	**0.23**	1.08	0.23	1.93	1.12	2.71	1.55	3.21	2.21	1.75	1.33	1.06	0.42	0.62	0.41
Bus	**0.49**	**0.26**	1.42	0.45	0.87	0.83	1.03	0.61	0.51	0.37	1.88	1.05	1.34	0.85	0.56	0.27
Mobile	**0.56**	**0.21**	1.18	0.76	3.41	1.07	1.85	1.43	1.63	1.42	1.08	0.38	1.26	1.17	0.85	0.33
Tennis	**0.67**	**0.25**	2.26	0.39	3.94	2.09	1.35	1.29	2.39	1.37	1.74	1.11	1.35	0.36	1.23	0.42
Football	**0.41**	**0.25**	1.69	1.33	1.01	0.86	1.57	2.75	0.66	0.39	1.54	1.05	1.42	1.01	0.8	0.48
Cargate	**0.46**	**0.28**	1.44	0.97	1.33	0.78	1.71	1.71	0.75	0.51	1.41	1.24	1.05	0.65	0.97	0.49
Renata	0.54	0.36	1.24	0.53	0.74	0.83	0.51	0.31	**0.37**	**0.31**	0.87	0.55	0.98	0.42	0.87	0.49
average	**0.52**	**0.27**	1.43	0.68	1.75	0.92	1.46	1.31	1.22	0.86	1.44	0.96	1.23	0.72	0.79	0.45

Recently, a number of wavelet-based techniques have been developed for video denoising/enhancement and video coding purposes [11], [12]. Accurate and efficient methods for noise estimation *in the wavelet domain* are preferred because all other processing, in wavelet-based denoising, also takes place in the wavelet domain. The most common method for noise estimation in the wavelet domain is a robust median estimator of [13], which computes the noise standard deviation as the median absolute deviation (MAD) of the wavelet coefficients in the highest frequency subband divided by 0.6754. The accuracy of the MAD method of [13] is sensitive to varying noise levels (see Table I), and it also varies for different image sequences (see Table II). Recently, three novel and alternative wavelet-based methods were proposed in [7], where the results were shown to outperform the MAD method of [13].

In contrast to most methods, which are purely intra-frame techniques, the methods in [2] and [9] are inter-frame techniques. The method of [9] uses multiresolution motion estimation in a video coder, in order to estimate noise variance only for the well-motion-compensated macroblocks, which are averaged in each frame. To our knowledge, no spatio–temporal noise estimation techniques that exploit both inter- and intra-frame content have been proposed so far. In this letter, we develop one such estimator that uses both spatial and temporal gradients.

III. NEW WAVELET DOMAIN NOISE ESTIMATION FOR VIDEO

A. Proposed Algorithm

In this letter, we propose a new low-complexity gradient-based noise level estimation method for Gaussian noise that is accurate and insensitive to highly textured image sequences with large moving areas. In contrast to the previously proposed gradient-based method of [6], the new proposed gradient method not only uses information from the spatial gradients but from the temporal gradients as well. Moreover, in a novel way, it corrects the initial estimate of the standard deviation of noise (most frequent gradient), based on the determined deviation of the corresponding (spatial or temporal gradient) distribution to its fitted Raleigh distribution.

In the proposed noise estimation method, the corresponding gradient estimates are the wavelet transform coefficients. Namely, we obtain spatial gradients, in terms of the horizontally $(LH^{(l)}(\mathbf{r},t))$ and vertically $(HL^{(l)}(\mathbf{r},t))$ oriented wavelet bands, of the 2-D wavelet decomposition of the image, and temporal gradients in terms of a 1-D wavelet transform high-pass band $HT^{(l)}(\mathbf{r},t)$. We use wavelet bands from the finest scale ($l=1$), where $l=1,\ldots,M$ denotes the decomposition level (1 denotes the finest scale and M the roughest). In particular, we use here nondecimated wavelet transform [14] with the Haar wavelet.

We define spatial and temporal gradient magnitudes $G_S(\mathbf{r},t)$ and $G_T(\mathbf{r},t)$ for the input sequence frame $I_\eta(\mathbf{r},t)$ as follows:

$$G_S(\mathbf{r},t)=\sqrt{\left(HL(\mathbf{r},t)^{(1)}\right)^2+\left(LH(\mathbf{r},t)^{(1)}\right)^2}$$
$$G_T(\mathbf{r},t)=\sqrt{\left(HT(\mathbf{r},t)^{(1)}\right)^2+\left(HT(\mathbf{r}+\mathbf{q},t)^{(1)}\right)^2} \quad (2)$$

where index $\mathbf{r}+\mathbf{q}$ stands for the randomly chosen spatial neighboring pixel position.

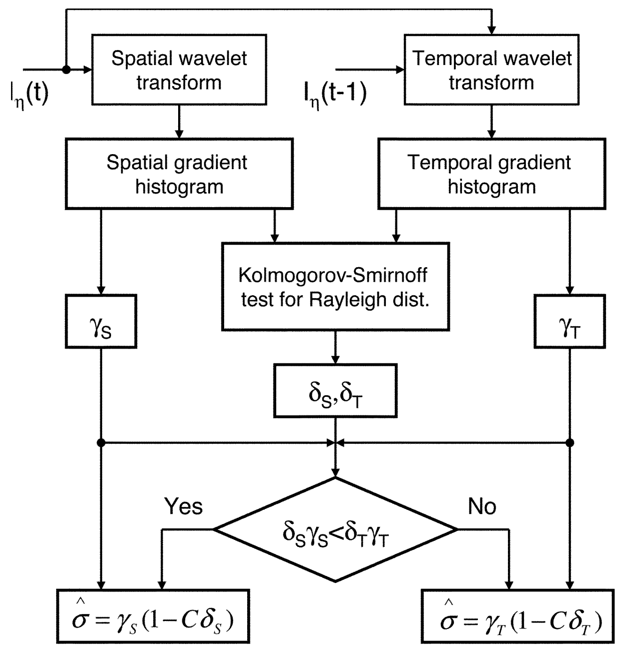

Fig. 1. General block scheme of the proposed noise estimation approach ($I_\eta(t)$ and $I_\eta(t-1)$ stand for the current and previous input noisy frame, respectively).

Let $\mathbf{h}_S(t)$ and $\mathbf{h}_T(t)$ denote the histograms of the spatial and temporal gradient magnitudes $G_S(\mathbf{r},t)$ and $G_T(\mathbf{r},t)$, respectively. Note that in an ideally uniform image sequence with added white Gaussian noise, both $\mathbf{h}_S(t)$ and $\mathbf{h}_T(t)$ follow the Rayleigh distribution. In a typical nonuniform image sequence, these histograms will deviate to some extent from the Rayleigh distribution, depending on the sequence content.

Fig. 1 outlines the proposed algorithm. In the first step for each time instant t, we compute the spatial and temporal gradient histograms. In the second step, we seek the most frequent gradient magnitudes in these histograms, i.e., the abscissa values γ_S and γ_T at which the amplitude gradient histograms, $\mathbf{h}_S(t)$ and $\mathbf{h}_T(t)$, respectively, peak.[1] Specifically, γ_S and γ_T are influenced by both noise and spatio–temporal image sequence structures, from the noisy sequence $I_\eta(\mathbf{r},t)$. We will use either γ_S or γ_T as the initial noise estimate, where the decision about which of the two is used for the initial estimate is based on the closeness of the spatial and temporal gradient histograms to the Rayleigh distribution. In particular, we fit the Rayleigh distribution to the spatial and the temporal magnitude gradient histograms, using the maximum-likelihood approach, and we evaluate the deviation between the fitted Rayleigh distribution and the corresponding histogram using the Kolmogorov–Smirnoff test [15]. The outputs of this test are *deviation errors* δ_S and δ_T for the spatial and temporal magnitude gradient distribution, respectively. Based on the computed errors δ_S and δ_T and using γ_S and γ_T, we make a decision about which magnitude gradient histogram will be used for noise estimation, in the following way. We define the minimum *correction error* as $\Delta = \min(\delta_S\gamma_S, \delta_T\gamma_T)$. If $\Delta = \delta_S\gamma_S$, we choose the spatial most frequent gradient γ_S as the initial estimate, and if $\Delta = \delta_T\gamma_T$, we take the temporal most frequent gradient γ_T as the initial estimate. In the following, we use the notation δ_k and γ_k, where $k = T$ (temporal) if $\delta_T\gamma_T < \delta_S\gamma_S$ and $k = S$ (spatial) if $\delta_S\gamma_S < \delta_T\gamma_T$.

[1]We smooth the spatial and temporal histograms prior to locating the most frequent gradient.

The next and final step is the correction of the initial estimates. The correction is also based on the output of the following Kolmogorov–Smirnoff test. We assume that δ_k measures the noise-free image sequence structures (spatial or temporal). However, there is no one-to-one relationship, because δ_k also depends on noise, when the structure is present. Specifically, δ_k decreases as the noise level increases, i.e., it is inversely proportional to the noise level. Since γ_k increases with the noise level increase, we multiply δ_k by γ_k to compensate for the noise level, in order to obtain approximately constant correction function Δ for a particular image sequence and for all noise levels. Note that this solution is not unique; nevertheless, the experimental results showed good performance of such a model for spatio–temporal image structures present in the image sequence.

For the additive noise model (1), using the previously explained noise-free and noisy image sequence structure description, we define the following noise estimator at time instant t:

$$\hat{\sigma}(t) = \gamma_k(t) - C\Delta(t) = \gamma_k(t)\left(1 - C\delta_k(t)\right) \tag{3}$$

which is essentially a correction (compensation) of the noise estimate based on the gradient peak. The correction factor $(1 - C\delta_k(t))$ in (3) can be viewed as a first-order Taylor series expansion of a more general correction factor $f(\delta_k)$. Nevertheless, the estimated noise variance can still fluctuate from frame to frame in the video sequence, because of the finite (integer) resolution of the evaluated histograms, i.e., because of the histogram binning errors. Consequently, we apply recursive averaging of the estimated $\hat{\sigma}$ in time to compensate for the fluctuations, i.e., smooth changes of σ_f in time, as follows: $\sigma_f(t) = (\sigma_f(t-1) + \hat{\sigma}(t))/2$, where t and $t-1$ correspond to the current and previous frame in the sequence, respectively. Finally, the constant $C = 1.25$ in (3) is determined experimentally so as to minimize the mean-squared error of the estimated noise variance σ_f on ten different images sequences: "Salesman," "Flower Garden," "Tennis," "Deadline," "Mobile," "Football," "Renata," "Cargate," "Bus," and "Uniform," with different contents and resolution and for seven different noise levels ($\sigma_\eta = 0$, 5, 10, 15, 20, 25, and 30).

B. Implementation

In our implementation, we take into consideration only gradients from the spatial positions belonging to luminance values between 16 and 235, in order to avoid the saturation effect, as suggested by ITU-Recommendation CCIR-601 and discussed in [8]. For the computation of the gradients, we use the Haar wavelet because of its low complexity. We have performed several experiments using different types of wavelets, but we have not observed significant change of the performance.

IV. Experimental Results

For the sake of comparison, we have compared the results of our noise estimation method with the well-known *MAD* noise estimator of [13], the spatial gradient method of [6], the structure-oriented method of [8], the wavelet-based moment

TABLE III
COMPUTATION TIME REQUIRED FOR THE PROCESSING

method	required time	method	required time
structure-oriented of [8]	0.45s	proposed spatio-temporal	2.68s
spatial-gradient [6]	0.97s	MAD [13]	2.76s
moment matching [7]	1.96s	temporal-based [2]	3.28s
CDF [7]	2.46s		

matching and CDF method of [7], and the temporal-based noise estimator of [2]. The comparison is made for eight different sequences in progressive format, of which five are in CIF format, namely, "Flower Garden," "Tennis," "Salesman," "Bus," and "Mobile" and three in high-definition format, that is, "Renata," "Football," and "Cargate." The results of the estimated noise standard deviations, averaged over the first 50 consecutive frames in a sequence and for seven different noise levels, are shown in Tables I and II, in terms of average error $\overline{E}$ and its standard deviation σ_E. In the last row of Tables I and II, we show the results for the proposed method applied to the interlaced sequences.

We calculate the absolute difference $E_i = |\sigma_\eta(i) - \sigma_f(i)|$ of the estimated and the true standard deviations, σ_f and σ_η, respectively, for each measurement i, and we tabulate the averaged errors $\overline{E} = \sum_{i=1}^{N} E_i/N$, where N stands for the number of measurements (concerning different noise levels or different test sequences). The standard deviation σ_E is calculated as follows: $\sigma_E = \sqrt{(1/N)\sum_{i=1}^{N}(E_i - \overline{E})^2}$. In Tables I and II, we show that the proposed method provides a smaller error $\overline{E}$ with a smaller standard deviation σ_E, than the algorithms of [2], [6]–[8], and [13]. Also, on average, the new method has a smaller sensitivity to spatio–temporal image sequence content, and the error depends less on the noise level. Finally, the results obtained for the proposed algorithm, applied to interlaced sequences (the last column in Tables I and II) show that the average error is approximately 50% higher than the estimate on the corresponding progressive sequence.

We have compared the complexity of the proposed method to the other compared methods, by evaluating the required time for processing. Under a Pentium 4, 2.66 GHz and Linux operating system (in C++), in case of CIF sequences (size: 352 $\times$ 288) and 50 frames, we have obtained the results shown in Table III. Specifically, in the proposed method, 55% (1.5 s) of computation time is spent for wavelet transform, 27% (0.73 s) for temporal frame memory updating, and 18% (0.45 s) for the rest.

V. CONCLUSION AND FUTURE WORK

In this letter, we have presented a new gradient-based noise estimation method for video sequences. The proposed method uses information from both the spatial and temporal gradients and corrects the initial noise estimate according to the estimated error introduced by the video sequence structures. The algorithm is implemented in the wavelet domain and is intended to be a part of the wavelet-based framework for wavelet-based video denoising techniques. In the future, we aim at extending the algorithm to other types of noise, by modeling the corresponding gradient distribution.

ACKNOWLEDGMENT

The authors would like to thank Prof. A. Amer for kindly providing the executable code of the algorithm [8].

REFERENCES

[1] G. De Haan, *Video Processing for Multimedia Systems*. Eindhoven, The Netherlands: Univ. Press Eindhoven, 2003.

[2] J. O. Drewery, R. Storey, and N. E. Tanton, Video Noise Reduction BBC Research Dept. Rep. BBC RD 1984/7, 1984.

[3] R. Bracho and A. C. Sanderson, "Segmentation of image based on intensity gradient information," in *Proc. Conf. Computer Vision Pattern Recognition*, San Francisco, CA, 1985.

[4] K. Rank, M. Lendi, and R. Ubenhauer, "Estimation of image noise variance," *Proc. Inst. Elect. Eng., Vision, Image, Signal Process.*, vol. 146, no. 2, pp. 80–84, Apr. 1999.

[5] V. Van De Dimitri, M. Nachtegael, D. Van Der Weken, E. E. Kerre, W. Philips, and I. Lemahieu, "Noise reduction by fuzzy image filtering," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 4, pp. 429–434, Aug. 2003.

[6] V. Zlokolica and W. Philips, "Motion-detail adaptive k-nn filter video denoising," in *Proc. SPIE Electronic Imaging 5298-47*, San Jose, CA, Jan. 2004.

[7] A. De Stefano, P. R. White, and W. B. Collis, "Noise level estimation on wavelet components," *J. Appl. Signal Process., Special Issue on Non-Linear Signal and Image Processing—Part 2*, pp. 1947–1956, 2004.

[8] A. Amer and E. Dubois, "Fast and reliable structure-oriented video noise estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 113–118, Jan. 2005.

[9] B. C. Song and K. W. Chun, "Noise power estimation for effective de-noising in a video encoder," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Philadelphia, PA, Apr. 2005, pp. 357–360.

[10] S. I. Olsen, "Estimation of noise in images: An evaluation," *Comput. Vis. Graph. Image Process.*, vol. 55, no. 14, pp. 319–323, 1993.

[11] W. I. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Proc. Wavelet Applications Signal Image Processing (Proc. SPIE 5207)*, San Diego, CA, Aug. 2003.

[12] J. C. Ye and M. Van Der Schaar, "Fully scalable 3-D overcomplete wavelet video coding using adaptive motion compensated temporal filtering," in *Proc. SPIE Video Communications Image Processing*, 2003.

[13] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaption by wavelet shrinkage," *Biometrika*, vol. 8, pp. 425–455, 1994.

[14] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. London, U.K.: Academic, 1999, Oval Road.

[15] S. Kokoska and D. Zwillinger, *Standard Probability and Statistics Tables: Student Edition*. Boca Raton, FL: CRC, 2000.