SPATIO-TEMPORAL APPROACH FOR NOISE ESTIMATION

V. Zlokolica, A. Pižurica, E. Vansteenkiste, W. Philips

Ghent University, Dept. of Telecommunications and Information Processing (TELIN), IPI
Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium
Vladimir.Zlokolica@telin.ugent.be

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 main application of the proposed algorithm is for the estimation of the stationary Gaussian noise in wavelet based video processing, for which we 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. Furthermore, we adapt the algorithm for local noise estimation and test its performance.

1. 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:

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

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 sequence frame t.

In the past a number of different methods have been proposed for noise variance estimation in still images and video, e.g. [1–9]. 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.

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 $G = \sqrt{{g_x}^2 + {g_y}^2}$. 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. A special case of gradient based methods are wavelet based techniques for noise estimation. The most common method for noise estimation in the wavelet domain is a robust median estimator of [10], 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. Recently, three novel wavelet based methods were proposed in [7] and shown to outperform the MAD method of [10].

In contrast to most methods, which are purely intra-frame (spatial) techniques, the methods in [2, 9] are inter-frame techniques which use temporal information exclusively. The method of [9] employs 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 which exploit both inter- and intra-frame content have been proposed so far.

In this paper, we propose a novel gradient-based noise estimator in the wavelet domain, which exploits both the temporal and the spatial correlations 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 the 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 in case of stationary Gaussian noise, 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. Additionally, we develop a modified proposed scheme for the local noise estimation (in a relatively small window), in which case the performance is reduced but still similar or slightly better than the other compared wavelet-based techniques.

The paper is organized as follows: We explain our method for noise estimation in Section 2 and propose the adapted solution for the local noise estimation in Section 2.1. In Section 3 we give implementation details and present experimental results. Finally, we conclude the paper in Section 4.

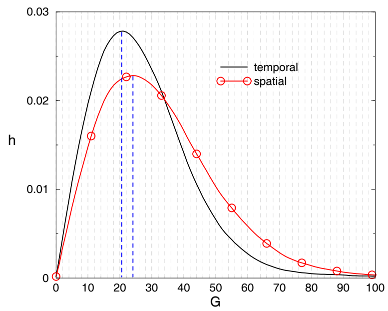

Fig. 1. Noise Estimation: Normalized spatial and temporal gradient magnitude histogram for the 3rd frame of "Tennis" image sequence with Gaussian noise ($\sigma_\eta = 20$).

2. THE PROPOSED WAVELET-BASED NOISE ESTIMATION ALGORITHM FOR VIDEO

In this paper, we propose a new low-complexity gradient-based noise level estimation method for Gaussian noise which is accurate and insensitive to highly textured image sequences with large moving areas. The new proposed method uses information from both the spatial and the temporal gradients. 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 spatial and the temporal gradients are estimated by the corresponding wavelet transform coefficients. Namely, we estimate spatial gradients by the wavelet coefficients in the horizontally ($LH^{(l)}(\mathbf{r}, t)$) and vertically ($HL^{(l)}(\mathbf{r}, t)$) oriented wavelet bands, of the two-dimensional (2D) wavelet decomposition of the image. Analogously, we express the temporal gradients in terms of the one-dimensional (1D) 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.

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:

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

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

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 non-uniform image sequence these histograms will deviate to some extent from the Rayleigh distribution, depending on the sequence content. Fig.1 illustrates the spatial and the temporal gradients histogram for the 3rd frame of the "Tennis" image sequence with noise level $\sigma_\eta = 20$. In this case, the spatial histogram deviates more from the Rayleigh distribution than the temporal one, because much stationary texture is present and motion appears only in a relatively small region. Specifically, the maximum of the spatial gradient distribution is shifted a little to the right and the tail of the distribution is relatively heavier in comparison to the temporal gradient distribution.

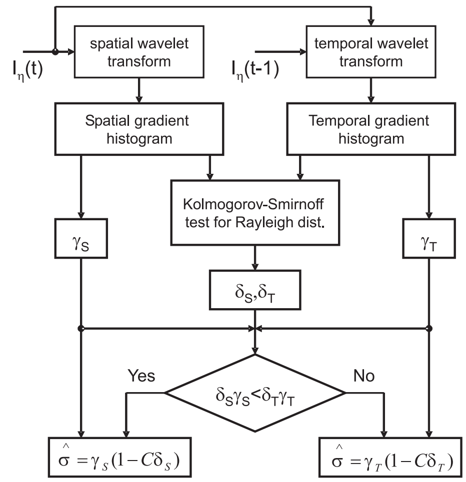

Fig. 2. General block scheme of the proposed noise estimation approach ($I_\eta(t)$ and $I_\eta(t-1)$ stand for the current and previous input noisy frame, respectively).

Fig.2 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 initial noise estimate, where the decision about which of the two is used for the initial estimate is based on the deviation of the corresponding histogram from 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 [11]. The output of this test is the *distribution deviation measure* (DDM) δ. In the following δ_S denotes DDM for the spatial and δ_T DDM for the temporal magnitude gradient distribution. 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.

The next and final step is the correction of the initial estimates. Our correction is also based on the distribution deviation measure (DDM), which is the output of the Kolmogorov-Smirnoff test. We assume that DDM measures the noise-free image sequence structures (spatial or temporal). However, there is no one-to-one relationship, because δ also depends on noise (when the structure is present), i.e., δ decreases as the noise level increases.

On the contrary, the most frequent gradient value γ increases

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

with the noise level increase. Hence, our idea is to compensate for the noise dependence of δ by multiplying it by the corresponding (spatial or temporal) most frequent gradient value γ. Note that this solution is not unique; nevertheless the experimental results showed good performance of such model for spatio-temporal image structures present in the image sequence.

Formally, our noise estimator at time instant t is:

$$\hat{\sigma}(t) = \begin{cases} \gamma_T(1 - C\delta_T) & \gamma_T\delta_T < \gamma_S\delta_S \\ \gamma_S(1 - C\delta_S) & \text{otherwise} \end{cases} \quad (3)$$

which is essentially a correction (compensation) of the noise estimate based on the gradient peak. The constant $C = 1.2$ was determined experimentally so as to minimize mean squared error of the estimated noise variance for the training set of sequences and 7 different noise levels ($\sigma = 0, 5, 10, 15, 20, 25, 30$), consisting of 50 frames. Equation (3) can be viewed as a finite order Taylor series approximation of a more general compensation formula, $\hat{\sigma} \simeq f(\gamma_k\delta)$. Specifically, we have investigated the more general case of Taylor approximation (of δ_k) with order 3, which is as follows:

$$\hat{\sigma}(t) = \gamma_k(t)(C_1 + C_2\delta_k(t) + C_3\delta_k^2(t) + C_4\delta_k^3(t)) \quad (4)$$

where the coefficients $C_1 = 1.0026$, $C_2 = -1.6356$, $C_3 = 3.8242$ and $C_4 = -7.6872$ where obtained by least square fitting of (4), to values of γ_k and δ_k computed on all the images of several training sequences set. For the constant noise level, the estimated noise variance can still fluctuate from frame to frame in the video sequence, because of the finite (integer) resolution of the histogram computation, 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: $\sigma_e(t) = (\sigma_e(t-1) + \hat{\sigma}(t))/2$ where t and $t-1$ correspond to the current and previous frame in the sequence, respectively.

2.1. Local Noise Estimation

Local noise estimation is important in video applications were non-stationary noise is present, i.e. in case when the noise level differs for different spatio-temporal positions in the video sequence. Hence, we extend the proposed algorithm for the stationary Gaussian noise to the non-stationary case. In order to adapt the proposed method to local noise estimation we have simplified the algorithm as follows: for each spatial position we compute only one histogram with both the spatial and the temporal gradients (as defined in (2)). We have experimentally found that a 2D block-window of the neighboring gradients of dimension 32×32 is a good trade off between the complexity and efficiency. Subsequently, we apply the correction function (3) to the obtained most frequent gradient (in this case we have only one distribution). Finally, we average the estimated standard deviations of noise in a 5×5 window in order to determine the final noise estimate for each pixel position.

3. EXPERIMENTAL RESULTS

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]. Further on, in the implementation, we use the non-decimated wavelet transform [12]. Namely, we use orthogonal Haar wavelet transform for 2D and 1D spatial wavelet transform, for its low complexity. Nevertheless, from the experiments done, we have not observed significant change of performance for different wavelet transforms.

Table 3. Local noise estimation with 32×32 window size

	New method	**MAD** [10]	**Matching method** [7]
$\overline{E}_{total}$	1.81	1.85	2.41

For the sake of comparison we have compared the results of our noise estimation method with the well known *MAD* noise estimator of [10], the structure-oriented method of [8], the wavelet-based moment matching and CDF method of [7] and the temporal-based noise estimator of [2]. The comparison is made for 8 different sequences in progressive format, of which 5 are in CIF format, namely, "Flower Garden", "Tennis", "Salesman", "Bus", "Mobile" and three in high definition format, that is, "Renata", "Football" and "Cargate". The results of the estimated noise standard deviations, averaged over first 50 consecutive frames in a sequence and for 7 different noise levels, are shown in Table 1 and Table 2, in terms of average error $\overline{E}$ and its standard deviation σ_E. In the last row of Table 1 and Table 2, 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 $\overline{E} = \frac{\sum_{i=1}^{N} E_i}{N}$, 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: $\sigma_E = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i - \overline{E})^2}$. In Table 1 and Table 2 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, 10]. Also, on average the new method has a smaller sensitivity to the spatio-temporal image sequence content and the error depends less on the noise level.

In the case of local noise estimation we present results in Table 3, where the $\overline{E}_{total}$ stands for the average absolute error $\overline{E}$ averaged over all spatial pixel positions, for the first 10 frames of the "Tennis", "Salesman" and "Bus" sequence. The results show that in case of the local noise estimation the proposed method does not perform as well as in the non-local case, i.e., the accuracy is relatively reduced. However, in comparison to the other methods, in case of the local noise estimation, the proposed modified method performs similarly or slightly better.

4. CONCLUSION AND FUTURE WORK

In this paper 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 presence of spatio-temporal structures in video sequences. In future we aim at extending the algorithm to other types of noise such as Poisson and speckle noise, by modeling the corresponding gradient distribution based on the noise distribution present in the image sequence.

Table 1. The average error $\overline{E}$ and the standard deviation of the error σ_E over all 8 sequences for different input noise levels (standard deviation σ_η); and averaged over first 50 frames.

Noise standard deviation	**New proposed method**		**Donoho MAD** [10]		**Moment Matching method** [7]		**CDF method** [7]		**Temporal method** [2]		**Structure Oriented** [8]	
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
0	0.18	0.14	3.77	2.02	0.23	1.78	2.56	2.93	2.11	0.92	1.13	0.41
5	0.61	0.24	2.21	1.54	2.14	1.79	1.84	1.09	1.27	0.63	0.58	0.16
10	0.62	0.18	1.61	1.56	1.22	1.37	1.14	1.01	0.94	0.66	0.61	0.31
15	0.53	0.22	1.54	1.62	0.89	0.79	0.93	0.89	0.77	0.61	0.77	0.29
20	0.46	0.18	0.99	0.97	0.96	0.53	0.77	0.71	0.86	0.43	1.22	0.23
25	0.52	0.25	1.25	0.95	1.44	0.53	0.71	0.47	1.18	0.78	1.79	0.45
30	0.78	0.29	1.20	0.75	3.39	2.31	0.63	0.36	2.91	1.36	2.55	0.83
average	0.52	0.21	1.75	1.34	1.46	1.18	1.22	1.06	1.44	0.77	1.23	0.38

Table 2. The average error $\overline{E}$ and the standard deviation of the error σ_E over all noise levels σ_η per sequence; and averaged over first 50 frames.

Image Sequence	**New proposed method**		**Donoho MAD** [10]		**Moment Matching method** [7]		**CDF method** [7]		**Temporal method** [2]		**Structure Oriented** [8]	
	$\overline{E}$	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	1.16	1.05	0.97	0.84	0.38	0.28	1.22	0.97	1.41	0.95
FlowerGar.	0.59	0.23	1.93	1.12	2.71	1.55	3.21	2.21	1.75	1.33	1.06	0.42
Bus	0.49	0.26	0.87	0.83	1.03	0.61	0.51	0.37	1.88	1.05	1.34	0.85
Mobile	0.56	0.21	3.41	1.07	1.85	1.43	1.63	1.42	1.08	0.38	1.26	1.17
Tennis	0.67	0.25	3.94	2.09	1.35	1.29	2.39	1.37	1.74	1.11	1.35	0.36
Football	0.41	0.25	1.01	0.86	1.57	2.75	0.66	0.39	1.54	1.05	1.42	1.01
Cargate	0.46	0.28	1.33	0.78	1.71	1.71	0.75	0.51	1.41	1.24	1.05	0.65
Renata	0.54	0.36	0.74	0.83	0.51	0.31	0.37	0.31	0.87	0.55	0.98	0.42
average	0.52	0.27	1.75	0.92	1.46	1.31	1.22	0.86	1.44	0.96	1.23	0.72

5. REFERENCES

[1] G. De Haan, *Video Processing for multimedia systems*, University Press Eindhoven, Eindhoven, The Nederlands, 2003.

[2] J.O. Drewery, R. Storey, and N.E. Tanton, "Video noise reduction," in *BBC Research Departement Report "BBC RD 1984/7"*, 1984.

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

[4] S. I. Olsen, "Estimation of noise in images: An evaluation," *Comput. Vision Graphics Image Processing*, vol. 55, no. 14, pp. 319–323, 1993.

[5] Dimitri Van De V., M. Nachtegael, D. Van Der Weken, E.E. Kerre, W. Philips, and I. Lemahieu, "Noise reduction by fuzzy image filtering," *IEEE Trans. on Fuzzy Systems*, 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, California, Jan. 2004.

[7] A. De Stefano, P.R. White, and W.B. Collis, "Noise level estimation on wavelet components," *Journal on Applied Signal Processing, 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. on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 113–118, 2005.

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

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

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

[12] S. Mallat, *A wavelet tour of signal processing (2nd ed.)*, Academic Press, Oval Road, London, 1999.