RELIABLE AND FAST STRUCTURE-ORIENTED VIDEO NOISE ESTIMATION

Aishy Amer *Amar Mitiche* *Eric Dubois*

INRS-Télécommunications
Université du Québec, Montréal, Canada

School of Information Technology and Eng.
University of Ottawa, Ottawa, Canada

amer@inrs-telecom.uquebec.ca

ABSTRACT

The purpose of this paper is to introduce a fast automated white-noise estimation method which gives reliable estimates in images with smooth and textured areas. This method is a block-based method that takes image structure into account and uses a measure other than the variance to determine if a block is homogeneous. It uses no thresholds and automates the way that block-based methods stop the averaging of block variances. The proposed method selects intensity-homogeneous blocks in an image by rejecting blocks of structure using a new structure analyzer. The analyzer used is based on high-pass operators and special masks for corners to allow implicit detection of structure and to stabilize the homogeneity estimation. For typical image quality (PSNR of 20-40 dB) the proposed method outperforms other methods significantly and the worst-case estimation error is 3 dB which is suitable for real applications such as video surveillance or broadcasts. The method performs well even in images with few smooth areas and in highly noisy images.

1. INTRODUCTION AND RELATED WORK

The effectiveness of video processing methods can be significantly reduced in the presence of noise. For example, intensity variation due to noise may introduce motion estimation errors. When information about the noise becomes available, processing can be adapted to the amount of noise to provide stable processing methods. For instance, image segmentation [1] and smoothing [2, 3] can be significantly improved when the noise variance is known. In current TV receivers the noise is typically estimated in the black lines of the TV signal [4]. In other applications, the noise estimate is provided by the user and no automated methods have been proposed that work well for various noise levels.

Noise can be estimated within an image (intra-image estimation) or between two or more successive images (inter-image estimation). Inter-image estimation requires more memory and is, in general, more computationally demanding. Intra-image noise estimation methods can be classified as smoothing-based or block-based. In smoothing-based methods the image is first smoothed, for example using an averaging filter, and then the difference between the noisy and enhanced image is assumed to be the noise; noise is then estimated at each pixel where the gradient is smaller than a given threshold. In block-based methods, the variance over a set of blocks of the image is calculated and the average of the smallest variances is taken as an estimate.

This work was supported, in part, by the the Natural Sciences and Engineering Research Council of Canada under Strategic Grant SRT224122 and Research Grant OGP0004234.

In [5], an evaluation of noise estimation methods is given: Many noise estimation methods have difficulties estimating noise in highly noisy images and in textured images. no techniques were found to perform best for various noise levels and images. Some methods use thresholds, for example, to decide whether an edge is given at a particular image position. Smoothing-based methods were found to perform well with high-noise levels but they require large computations and fine tuning for various images. Smoothing-based methods have difficulties in images with fine texture and they tend to overestimate the noise variance. Block-based methods are, in general, less complex than smoothing-based methods. They tend, in general, to overestimate the noise in good quality images and underestimate it in highly noisy images. In some cases, no estimate is even possible.

Recently, a new interesting smoothing-based noise estimation method has been proposed [6]. Its main difficulty is the heavy computational cost. Its success seems to depend heavily on many parameters to fix, for example, on the number of process iterations, or the shape of the fade-out cosine function to evaluate the variance histogram (Eqs. 9 and 10 in [6]). Other noise estimation methods determine the noise within the larger context of a video processing system where they are adapted to specific needs of the system (e.g., in the context of coding [7] and image segmentation [1]).

2. STRUCTURE-ORIENTED NOISE ESTIMATION

The proposed method is a block-based method that estimates the noise variance σ_n^2 from the variances of a set of blocks classified as the most homogeneous blocks in an image $I(n)$, i.e., blocks with the lowest structure variation. In [8] a comparison of structure detectors is given. There, it is shown that precise structure and edge detection is computationally expensive and computationally less expensive detectors either need some manual tuning, are designed for specific edge models, or are not precise enough.

The proposed method uses a low-complexity automated homogeneity measure ξ_{B_h} to determine if an image region has uniform intensities, where uniformity is equated to piece-wise constant gray-level pixels. This novel noise estimation operates 1) without a prior knowledge of the image or noise, 2) without context, i.e., it is designed to work for different video processing systems, and 3) without user interactions. The only underlying assumption is that in an image there exist neighborhoods (usually chosen as a 2-dimensional rectangular window) with smooth intensities (i.e., the proposed homogeneity measure $\xi_{B_h} \simeq 0$). This assumption is realistic since real-world images have well-defined regions of distinct properties, one of which is smoothness. The proposed noise estimation operates as follows:

1. Detecting intensity-homogeneous blocks The pixels in an intensity-homogeneous block $B_h = \{I(i,j)\}_{(i,j)\in W_{ij}}$ are assumed to be independent and identically-distributed (iid) but not necessarily zero-mean. W_{ij} denotes the rectangular window of size $W \times W$. These uniform samples $\{I(i,j)\}$ of the image have variance $\sigma^2_{B_h}$, which is assumed to represent the local variance of the noise. The signal in a homogeneous block is approximately constant and variation is due to noise. With the iid property their empirical mean and variance are defined as

$$\mu_{B_h} = \frac{\sum_{(i,j)\in W_{ij}} I(i,j)}{W \times W}; \sigma^2_{B_h} = \frac{\sum_{(i,j)\in W_{ij}} (I(i,j) - \mu_{B_h})^2}{W \times W}. \tag{1}$$

With $l = W \times W$ and by the law of large numbers

$$\lim_{l\to\infty} \sigma^2_{B_h} = \sigma^2_n. \tag{2}$$

2. Averaging To estimate the global image noise variance, σ^2_n, the local variances of the m most homogeneous blocks, $\{B_h\}$, are averaged to $\sigma^2_n = \mu_{\sigma^2_{B_h}} = \frac{\sum_{h=1}^{m} \sigma^2_{B_h}}{m}$. Since the noise is assumed to be stationary, the average of the variances of the m most homogeneous regions can be taken as a representative for the noise in the whole image. To achieve faster noise estimation, ξ_{B_h} is calculated for a subset of the image pixels by skipping each s^{th} pixel of an image row. Simulations are carried using different skipping steps where a good compromise between efficiency (computational costs) and effectiveness (solution quality) is obtained with $s = 5$ (the choice of s depends on W).

3. Adaptive averaging Since the most homogeneous blocks could show strongly variable homogeneities and hence highly variable variances, only blocks which show similar homogeneities and thus similar variances $\sigma^2_{B_h}$ to a *reference representative variance* $\sigma^2_{B_r}$ are included in the averaging process. To decide whether the reference and a current variance are similar, a threshold t_σ is used, i.e., $\sigma^2_{B_h}$ is similar to $\sigma^2_{B_r}$ if $|\sigma^2_{B_r} - \sigma^2_{B_h}| < t_\sigma$. This stabilizes the averaging process and no threshold is needed to stop the averaging process. The threshold t_σ is relatively easy to define and does not depend on the input image. It can be seen as the maximal affordable difference (i.e., error) between the true variance and the estimated variance. For example, in noise reduction in TV receivers a t_σ between 3 and 5 is common [2, 4]. In the simulations of this study, t_σ is set to 3.

2.1. Detecting homogeneous blocks

The image is first divided into blocks $\{B_h\}$ of the size $W \times W$. In each block B_h a homogeneity measure ξ_{B_h} is then computed using a local image analyzer based on high-pass operators that are able to measure homogeneity in eight different directions as shown in Fig. 1 where special masks for corners are also considered that stabilize the homogeneity estimation. In this local uniformity analyzer, high-pass operators with coefficients {-1 -1 ... (W-1) -1 -1} are applied along *all* directions for each pixel of the image (e.g., if $W = 3$ the coefficients are {-1, 2, -1 }). If in one direction the image intensities are uniform then the result of the high-pass operator is close to 0. To calculate the homogeneity measure for all eight directions, the absolute value of all eight quantities are added to give a measure, ξ_{B_h}, for homogeneity.

The detection of homogeneity can be expressed as a second-order operator on the image function I for $W = 3$. The following example illustrates this in the horizontal direction:

$$\begin{aligned} I_o(i) &= -I(i-1) + 2 \cdot I(i) - I(i+1) \\ &= -I'(i) + I'(i-1) \\ &= -(I'(i) - I'(i-1)). \end{aligned} \tag{3}$$

Therefore, $I_o(i)$ is a second-order finite-difference operator which acts as a high-pass operator. Note that the detection of homogeneity is done along edges and not across edges. Various simulations (see Sec. 3) show that this proposed homogeneity measure performs better than using the variance to detect uniform intensities. A variance-based homogeneity measure fails in the presence of fine structures and textures. Note that the proposed high-pass masks respond well for line, step, and shoulders of ramp edges but have difficulties at centers of ramp edges. However, the special corner masks used compensates this in blocks with ramp edges. In addition, experimental comparisons suggest that no definite advantage in using first-order edge detector can be achieved (see Fig. 4). Note further that many first-order edge detectors are either designed for specific edge models or need manual tuning.

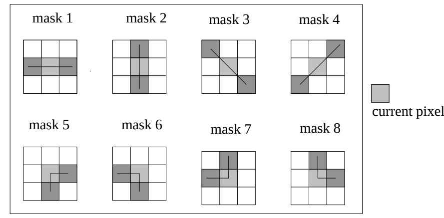


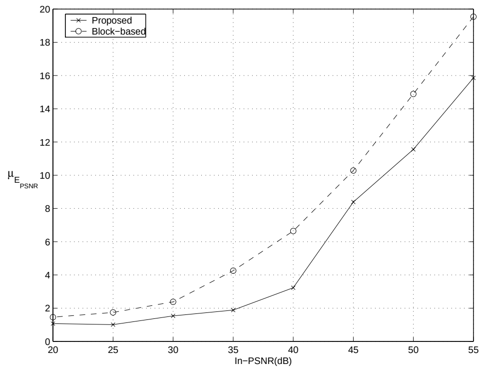
Fig. 1. Directions of the homogeneity analyzer.

2.2. Defining a reference variance $\sigma^2_{B_r}$

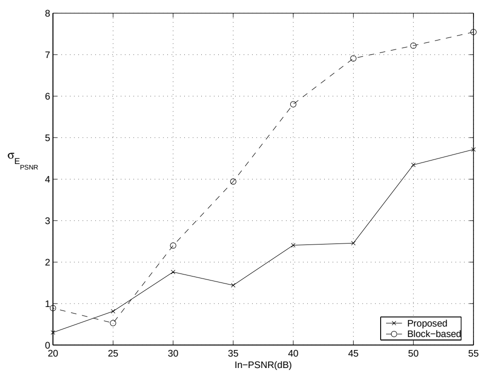
To stabilize the averaging of local variances, the reference variance is chosen as the median of the variances of the first three most homogeneous blocks (i.e., the blocks with the smallest sum). The first three values are taken because they are most representative of the noise variance since they are calculated from the three most homogeneous blocks. Higher-order median operators can be also used but attention should be paid to large deviations between the variances included in the operation. Instead of the median, the mean can be used to reduce computation. Simulations show that better estimation is achieved using the 3-tap median operator. In some cases, the difference between the first three variances can be large and a median filter would result in a good estimate of the true reference variance. Further investigation can determine the best order of the median filter, or examine if there are cases where the mean operator would give better results.

3. ANALYSIS AND EVALUATION

The new estimator has been tested with nine commonly used images from the image processing literature (see the web at [9] and [5]). White additive noise is the most common form of noise in images and it has been used in the tests. To test the reliability of the proposed method, noise giving a PSNR between 20 and 50 dB is added to the nine images. Typical PSNR values in real-world images range between 20 and 40 dB. Noise is also estimated in the original images where noise is usually present in unknown amount



(a) Average of E_{PSNR}.



(b) Standard deviation of E_{PSNR}.

Fig. 2. Comparison of the block-based ($W = 7$, see [5]) and the proposed method ($W = 5$).

(thus the actual noise variance would be the sum of this unknown noise variance and the added noise variance).

Due to the limited range of intensities ($[0, 255]$), saturation effects result in a Gaussian noise not having exactly zero-mean, especially with large noise variances. In this paper, therefore, attention is paid to this saturation or clipping effect. This has been done according to the ITU-R Recommendation CCIR-601.1 for the YCrCb video standard. In this recommendation, the reference black is represented by 16 and the reference white by 235 for the 8-bit range $[0, 255]$. Thus, noise is estimated solely in regions of these ranges so that clipping effects are excluded from the estimation process. This, however, could limit the performance of the algorithm where the homogeneous regions lay outside these ranges. As the evaluation of the proposed method below show, the proposed method gives reliable results despite this range limitation.

To evaluate the performance of the algorithm, the estimation error $E_n = |\sigma_n^2 - \sigma_e^2|$ is first calculated. E_n is the difference between the true and the estimated noise variance. The average μ_{E_n} and the standard deviation σ_{E_n} of the estimation error are then computed from all the measures as follows:

$$\mu_{E_n} = \frac{\sum_{i=1}^{N} E_n(i)}{N}; \sigma_{E_n}^2 = \frac{\sum_{i=1}^{N}(E_n(i) - \mu_{E_n})^2}{N} \quad (4)$$

where N is the number of tested images and E_n is the estimation error for a particular variance σ_n^2 on a single image. The reliability of an estimation method can be thus measured by σ_{E_n} and/or μ_{E_n}.

Evaluation results are given in Table 1. As shown, the proposed method is reliable for both high and low noise levels and the estimation errors remain reliable even in the worst-case when deviation is around 1.81. For example, in high-end noise reduction techniques, the adjustment is done in an interval of 2-5 dB [9, 2, 4]. In [5], an evaluation of noise estimation methods is given. When our results are compared to those of Table 1 in [5], the comparison suggests that the proposed method outperforms the block-variance-based method, which has been found in [5] to be a good compromise between efficiency and effectiveness. Moreover, the proposed method adapts thresholds whereas the block-based method requires manual tuning for improved performance.

PSNR	55 (org.)	50	45	40	35	30	25	20	
σ_n	0	0.80	1.43	2.55	4.53	8.06	14.33	25.50	ave.
μ_{E_n}	1.85	1.99	1.78	1.32	1.04	0.90	1.45	2.55	1.61
σ_{E_n}	1.73	1.81	1.40	1.16	0.71	1.40	1.37	1.27	1.36

Table 1. The average μ_{E_n} and the standard deviation σ_{E_n} of the estimation error as a function of σ_n for $W = 5$ and $s = 5$.

Fig. 2(a) reveals that the estimation error using the proposed method is lower than that of the block-variance method for *all* noise variances. More interestingly, the standard deviation of the error using the proposed method is significantly less (Fig. 2(b)). Note that the highest error for PSNRs of 20-40 dB using the proposed method is 3 dB. (In Figs. 2-5, $E_{\text{PSNR}} = |\text{PSNR}_{true} - \text{PSNR}_{est}|$) is used instead of E_n since dB values are more illustrative).

We have evaluated the proposed method using different window sizes $W = 3, 5, 7, 9, 11$. As shown in Fig. 3, using a window size of 3×3 results in a better estimation in less noisy images (PSNR>40 dB), whereas using a window size of 5×5 gives better results in noisy images. This is reasonable since, in noisy images, larger samples are needed to calculate the noise accurately. The choice of the window size can be oriented to some image information if available. Since reliable estimation is more required in heavy-noisy images and as a compromise between efficiency and effectiveness, a window size of 5×5 is used. If a reduction in computation cost is required, the proposed noise estimation can be carried out only in the horizontal direction, i.e., along one line, for example, using 3×1 or 5×1 window size.

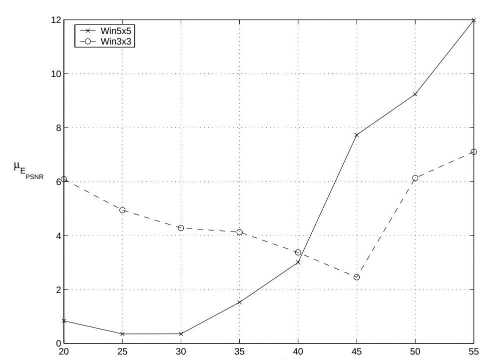


Fig. 3. Estimation ($\mu_{E_{\text{PSNR}}}$) comparison of the proposed method using different window sizes.

We have examined the use of low-complexity first-order edge detectors to calculate ξ_{B_h} (Sec. 2.1). Preliminary results suggest that there may not be definite advantages in using these (Fig. 4).

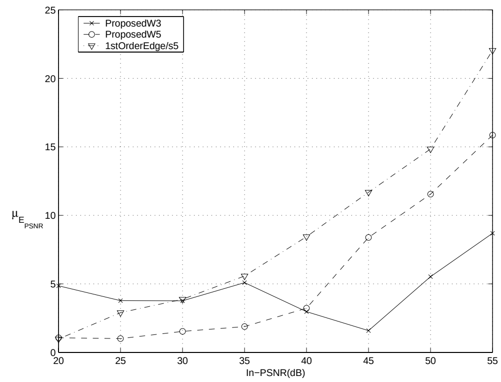


Fig. 4. Comparison of $\mu_{E_{\text{PSNR}}}$ using first-order edge detection and the proposed higher-order structure detection ($s = 5$).

The stability of the proposed method is further confirmed when applied to sequences with and without global motion. For example, the sequences 'Prlcar', 'Flowergarden', and 'Train' (see the web at [9]) are overlaid with 30 dB PSNR noise and the PSNR estimates are: 'Prlcar' 29.19-30.65 dB; 'Flowergarden' 27.51-31.23 dB; and 'Train' 29.40-31.18 dB. The proposed method is thus temporally stable for noise-adaptive video applications (e.g., [2, 4]).

Table 2 summarizes the performance of the proposed, block-based, and average methods [5]. As shown, the proposed method gives significantly lower error than the reference methods. Table 2) also show that the proposed method is four times faster than the block-based method which has been found to be the most computationally efficient among tested noise estimation methods [5].

	AVE	BLC	Proposed
average(μ_{E_n})	2.22	4.45	1.61
average(σ_{E_n})	2.51	3.25	1.36
T_c	6×slower than BLC		4×faster than BLC

Table 2. Effectiveness and complexity comparison between the proposed method, averaging method (AVE), and block-based method (BLC). T_c is the computational time. See also [5].

4. CONCLUSION

This paper contributes a reliable fast automated method for estimating the variance of white noise. This method requires a 5×5 mask and averages noise variances of blocks with lowest structure and similar variances. It uses eight high-pass operators (with special corner masks) to measures the high-frequency image components. The mask used is separable and can be implemented with simple FIR-filters. The operators compensates for the noise along eight directions and stabilizes the selection of homogeneous blocks. Experimentally, no advantage using first-order edge detectors was confirmed. The proposed method performs well even in textured and heavy noisy images. As shown in Fig. 5, for a typical image quality of PSNR between 20 and 40 dB the proposed method outperforms other methods significantly and the highest estimation error is approximately 3 dB which is suitable for real video applications such as surveillance or TV signal broadcasts.

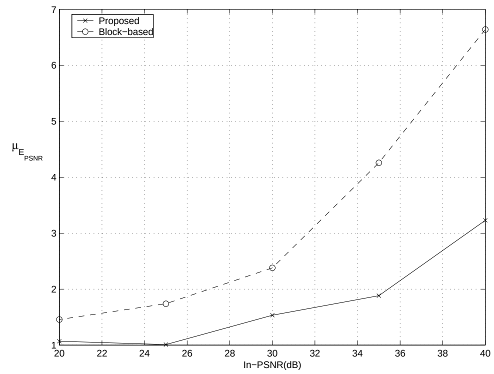


Fig. 5. Performance ($\mu_{E_{\text{PSNR}}}, W = 5$) comparison of the proposed and block-based [5] method in typical PSNR range.

Our future work includes 1) a comparative study of the proposed and first-order structure detection masks and their combination, 2) an evaluation of our method for non-white and non-stationary noise, and 3) a study of the proposed method in (e.g., MPEG-2) compressed images and after applying a noise filter.

5. REFERENCES

[1] M. Spann and R. Wilson, "A quad-tree approach to image segmentation which combines statistical and spatial information," *Pattern Recognit.*, vol. 18, no. 3/4, pp. 257–269, 1985.

[2] H. Blume, A. Amer, and H. Schröder, "Vector-based postprocessing of MPEG-2 signals for digital TV-receivers," in *Proc. SPIE Visual Communications and Image Process.*, San Jose, CA, Feb. 1997, vol. 3024, pp. 1176–1187.

[3] P. Meer, R.H. Park, and K. Cho, "Multiresolution adaptive image smoothing," *Graphical Models and Image Process.*, vol. 44, no. 2, pp. 140–148, Mar. 1994.

[4] G. de Haan, T. Kwaaitaal-Spassova, and O.A. Ojo, "Automatic 2-D and 3-D noise filtering for high-quality television receivers," in *Proc. Int. Workshop on Signal Process. and HDTV*, Turin, Italy, 1996, vol. VI, pp. 221–230.

[5] S.I. Olsen, "Estimation of noise in images: An evaluation," *Graphical Models and Image Process.*, vol. 55, no. 4, pp. 319–323, July 1993.

[6] K. Rank, M. Lendl, and R. Unbehauen, "Estimation of image noise variance," *IEE Proc. Vis. Image Signal Process.*, vol. 146, no. 2, pp. 80–84, Apr. 1999.

[7] K. Konstantinides, B. Natarajan, and G.S. Yovanof, "Noise estimation and filtering using block-based singular-value decomposition," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 479–483, Mar. 1997.

[8] E.P. Lyvers and O.R. Mitchell, "Precision edge contrast and orientation estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 6, pp. 927–937, November 1988.

[9] A. Amer, *Object and Event Extraction for Video Processing and Representation in On-Line Video Applications*, Ph.D. thesis, INRS-Télécommunications, Dec. 2001, www.inrs-telecom.uquebec.ca/users/amer/phd/.