NOISE POWER ESTIMATION FOR EFFECTIVE DE-NOISING IN A VIDEO ENCODER

Byung Cheol Song and Kang Wook Chun

Digital Media R&D Center, Samsung Electronics Co., Ltd.
416 Maetan-3dong, Yeongtong-gu, Suwon, 443-742, Republic of Korea

ABTRACT

This paper presents a noise estimation algorithm using multi-resolution motion estimation in a video encoder. Firstly, the motion estimator finds minimum block-matching errors at the finest resolution and the middle resolution for each macroblock. Secondly, if the minimum block-matching error at the finest resolution of a certain macroblock is less than a particular threshold, the variance of the inter-macroblock is computed. And then, we employ the square of the minimum block-matching error at the middle resolution as a predictor of the variance of the desired inter-macroblock without noise. The noise variance in the macroblock can be estimated by subtracting the predictor from the variance of the inter-macroblock. Finally, the noise variances estimated only for the well-motion-compensated macroblocks are averaged in each frame. Experimental results show that the proposed noise estimation is very accurate with negligible computational cost.

1. INTRODUCTION

Input video sequences of a video encoder can be often degraded due to various noise sources. A main source of contamination of video sequences is the additive noise introduced by the channel in the conventional analogue transmission systems. The noisy video sequences, which are not only visually annoying, but they are also hard to be encoded efficiently owing to uncorrelated nature of noise, need to be sometimes digitally encoded in the receiver side. The main goal of pre-filtering is to remove as much high-frequency information as possible without compromising visual quality.

A number of pre-filtering schemes have been developed [1-3]. The conventional pre-filtering schemes have been devised in the light of noise reduction itself rather than optimal combination of pre-filtering with a video encoder. Since the noise reduction operation has been thought of as a process independent of video encoding, the pre-filtering schemes have been usually cascaded with video encoders. However, in the cascaded structure, a video encoder becomes computationally heavier due to the additional pre-filtering complexity.

Kim and Ra proposed a DCT-domain noise reduction scheme (DCTNR), which is gracefully embedded with a video coder [4]. They first applied the concept of the generalized Wiener filter [5] to a video encoder. The DCTNR accomplishes fast pre-filtering because all the processing is operated in the DCT domain simply by scaling the DCT coefficients. Simulation results show that the DCTNR noticeably outperforms the other pre-filtering schemes such as a spatial-domain adaptive Wiener filter [2] in terms of noise reduction and coding efficiency [4].

Most pre-filtering algorithms as well as the DCTNR assume that noise power, i.e., noise variance is already known prior to encoding. However, actual noise variance is practically impossible to be known in advance. Since an inaccurate noise variance different from the true noise variance deteriorates coding performance as well as noise reduction, a precise noise estimation algorithm is demanding.

The organization of this paper is as follows. Section 2 introduces previous algorithms. Section 3 presents the proposed algorithm. Section 4 gives simulation results. Finally, Section 5 provides conclusion.

2. PREVIOUS WORKS

There are two kinds of noise estimation: intra-image estimation and inter-image estimation [6-8]. Recently, research on noise estimation has been mainly focused on intra-image noise estimation because inter-image noise estimation requires more memory and is usually computationally heavier. Intra-frame noise estimation methods can be classified as smoothing-based methods and block-based methods [6-7]. In smoothing-based methods, the image is first smoothed, i.e., low-pass-filtered, and then the difference between the noisy and the filtered image is assumed to be the noise; noise is then estimated at each pixel where the gradient is less than a certain 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.

Amer *et al.* proposed a block-based method that estimates the noise variance from the variances of a set of blocks classified as the most homogeneous blocks in images with smooth and textured areas [6]. The method selects intensity-homogeneous blocks in an image by rejecting inhomogeneous blocks using a new structure analyzer, which is based on high-pass operators and special masks for corners to allow implicit detection of structure and to stabilize the homogeneity estimation. However, it is still difficult to detect homogeneous blocks accurately in a noisy environment, and a false selection of the reference noise variance may often bring out a significant estimation error.

On the other hand, Natarajan introduced an interesting noise estimation method for reducing additive random noise from signals using data compression [7]. However, his scheme needs a huge amount of computation due to a number of compression/ decompressions, and it does not provide the precise noise power of each image. Moreover, Natarajan's scheme cannot be merged into the popular video encoders such as MPEG and H.26x encoders. To our knowledge, there are no schemes that estimate noise power exactly inside a video encoder, with small amount of computation.

Song and Chun proposed an efficient noise estimation algorithm, based on motion-compensated block differences obtained from the motion estimator in a video encoder, while having minor computational cost [8]. However, the algorithm still does not provide a solid noise prediction model proper for various kinds of video sequences.

This paper presents an advanced noise estimation algorithm that utilizes some relevant information from the multi-resolution motion estimator in a video encoder. Firstly, minimum block-matching errors at two different resolution levels, i.e., the finest resolution and the middle resolution, are extracted from the multi-resolution motion estimator. Secondly, for a well-motion-compensated macroblock (MB), the difference between the variance of the inter-MB and the square of the minimum matching error at the middle resolution is calculated. Since noise and a desired block-matching error signal are generally uncorrelated, the variance of the inter-MB is assumed to be the sum of the noise variance and the variance of the desired inter-MB without noise. We adopt the square of the block-matching error at the middle resolution as a predictor of the variance of the desired inter-MB. So, the difference between the variance of the inter-MB and the predictor can be the noise variance estimate. Finally, the differences of the MB's, i.e., the estimated noise variances are averaged in each frame. The average is determined as the noise variance estimate of the frame.

3. THE PROPOSED ALGORITHM

In this paper, the proposed noise estimator is implemented in a video encoder where the DCTNR is embedded for noise reduction (see Fig. 1). Note that the DCTNR is known as one of good solutions of noise reduction.

3.1. DCTNR

In Fig. 1, we can find that the whole operation for DCTNR is equivalent to a unified scaling operation with a scaling matrix **F** in the DCT domain, which is described as follows [4]:

$$F(k,l)=\frac{1+S(k,l)\cdot\dfrac{\sigma_n^2}{\sigma^2}\cdot\dfrac{1}{\gamma(k,l)}}{1+\dfrac{\sigma_n^2}{\sigma^2}\cdot\dfrac{1}{\gamma(k,l)}}. \tag{1}$$

In (1), if (k,l) is (0,0), $S(k,l)$ is 0. Otherwise, $S(k,l)$ is 1. So, $F(k,l)$, i.e., (k,l)-th filter coefficient is determined depending on the signal-to-noise level and $\gamma(k,l)$. $\gamma(k,l)$ is a normalized element on a diagonalized matrix of DCT-ed covariance, and realistic covariance estimates for mean-subtracted intra and inter blocks can be found by using some training sequences [4]. σ^2 denotes the variance of the desired data without noise, and it is usually estimated by subtracting the noise variance (σ_n^2) from the variance of an input block, i.e., σ_e^2. Hence,

$$\sigma^2=\max\{\sigma_e^2-\sigma_n^2,0\}. \tag{2}$$

σ_e^2 can be computed easily. So, if σ_n^2 is known, $F(k,l)$ is determined. As a result, a (k,l)-th coefficient of the DCT-ed input block **E,** $E(k,l)$ is filtered simply as follows:

$$\hat{E}(k,l)=E(k,l)\cdot F(k,l). \tag{3}$$

More detailed explanation about (1) is described in [4].

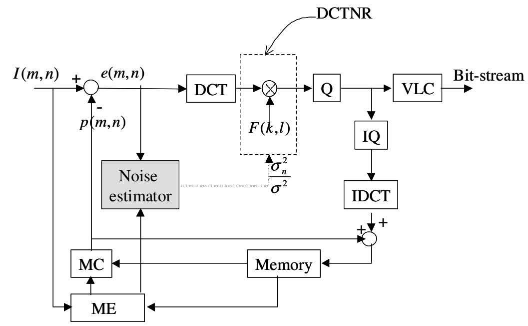

Fig. 1. A video encoder employing the proposed noise estimator.

3.2. Proposed Algorithm

We propose a motion-compensated noise estimation algorithm based on multi-resolution motion estimation in a video encoder. We presented a multi-resolution block matching algorithm (MRBMA) that provides high computational speed and high estimation performance concurrently [9].

3.2.1. MRBMA

MRBMA consists of three levels. Prior to motion estimation processing, the coarser level frames of a current frame as well as a reference frame should be produced. The coarser level frames are constructed by a 2-dimensional low-pass-filter (LPF) and sub-sampling. In this paper, a 2×2 average filter is used for the 2-dimensional LPF. So, the MB size becomes 16×16, 8×8, and 4×4 at level 0, 1, and 2, respectively. At level 2, two motion vector (MV) candidates are obtained on the basis of minimum matching error for the next level search. At level 1, the two candidates selected at level 2 and the other one based on spatial MV correlation at level 0 are used as the center points for local searches, and a single MV candidate is chosen for the next level search. Then, at level 0, the final MV is obtained from a local search around the single candidate. MRBMA outperforms the other multi-resolution motion estimation algorithms in terms of estimation accuracy and computational complexity. So, we employ the MRBMA for motion estimation in a video encoder of Fig. 1.

3.2.2. Proposed noise estimation

MAD (Mean of Absolute Difference) is normally used as a measure of block-matching error. At level 1, MRBMA finds a MV ($MV^{(1)}$) with minimum *MAD* for each MB. The minimum *MAD* corresponding to $MV^{(1)}$, i.e., $(p^{(1)}, q^{(1)})$ is described by

$$MAD_{\min}^{(1)}=\min_{(p,q)\in\Omega^{(1)}} MAD^{(1)}(p,q), \tag{4}$$

where $\Omega^{(1)}$ is the search ranges corresponding to three local searches at level 1, and $MAD^{(1)}(p,q)$ is defined as follows:

$$MAD^{(1)}(p,q)=\frac{1}{64}\sum_{x=m/2}^{m/2+7}\sum_{y=n/2}^{n/2+7}\left|I_i^{(1)}(x,y)-I_{i-1}^{(1)}(x+p,y+q)\right|. \tag{5}$$

$I_i^{(l)}(m,n)$ represents the intensity value at the position (m,n) of the i-th frame at level l.

At level 0, MRBMA locates a MV with minimum *MAD* within a local search range around $2\cdot MV^{(1)}$, i.e., $\Omega^{(0)}$. The minimum *MAD* at level 0 is depicted as

$$MAD_{\min}^{(0)} = \min_{(p,q)\in\Omega^{(0)}} MAD^{(0)}(p,q), \quad (6)$$

where $MAD^{(0)}(p,q)$ is the MAD corresponding to (p,q) at level 0. Considering additive noise, the i-th input frame is re-defined as follows:

$$I_i^{(0)}(m,n) = I_{i,org}^{(0)}(m,n) + \eta^{(0)}(m,n), \quad (7)$$

where $I_{i,org}^{(0)}(m,n)$ denotes an original desired frame with no noise, and $\eta^{(0)}(m,n)$ is a noise component at the same location. Assume that the (i-1)-th frame is a reference frame for motion estimation. In this paper, we employ the (i-1)-th reconstructed frame for motion estimation at each level.

From (6) and (8),

$$MAD_{\min}^{(0)} = \frac{1}{256}\sum_{x=m}^{m+15}\sum_{y=n}^{n+15}|\Delta I^{(0)} + \Delta\eta^{(0)}|, \quad (8)$$

where $\Delta I^{(0)} = I_{i,org}^{(0)}(x,y) - I_{i-1,rec}^{(0)}(x+p^{(0)}, y+q^{(0)})$ and $\Delta\eta^{(0)} = \eta^{(0)}(x,y) - \eta^{(0)}(x+p^{(0)}, y+q^{(0)})$. In (8), it is assumed that $\Delta\eta^{(0)} \cong \eta^{(0)}(x,y)$ because additive noise in the (i-1)-th reconstructed frame is already removed by the DCTNR in a video encoder. So, (8) can be approximated as

$$MAD_{\min}^{(0)} \cong \frac{1}{256}\sum_{x=m}^{m+15}\sum_{y=n}^{n+15}|\Delta I^{(0)} + \eta^{(0)}(x,y)|. \quad (9)$$

On the other hand, if signal and noise are uncorrelated, the inter-MB corresponding to $MAD_{\min}^{(0)}$ has the following variance:

$$\sigma_{\min}^2 = \sigma_{\Delta I^{(0)}}^2 + \sigma_n^2, \quad (10)$$

where $\sigma_{\Delta I^{(0)}}^2$ and σ_n^2 are the variances corresponding to $\Delta I^{(0)}$ and $\eta^{(0)}$ of the MB, respectively. Assuming perfect motion compensation, i.e., $\Delta I^{(1)} = \Delta I^{(0)} = 0$, (9) is re-written:

$$MAD_{\min}^{(0)} \cong \frac{1}{256}\sum_{x=m}^{m+15}\sum_{y=n}^{n+15}|\eta^{(0)}(x,y)|. \quad (11)$$

In this ideal case, since $\sigma_{\Delta I^{(0)}}^2$ is zero, $\sigma_{\min}^2$ is equal to σ_n^2. However, perfect motion compensation is generally difficult. So, if we predict $\sigma_{\Delta I^{(0)}}^2$ well, σ_n^2 can be derived. To prove this, we applied MRBMA to several video sequences without noise. The same experiment setup as Section 4 is employed. In this experiment, we examined the relation between $\sigma_{\Delta I^{(0)}}^2$ and $\left(MAD_{\min,org}^{(1)}\right)^2$ (see Fig. 2). In Fig. 2, the horizontal axis indicates $\left|\sigma_{\Delta I^{(0)}}^2 - \left(MAD_{\min,org}^{(1)}\right)^2\right|$ and the vertical axis indicates it histogram. Statistically, the probability that $\sigma_{\Delta I^{(0)}}^2$ is very close to $\left(MAD_{\min,org}^{(1)}\right)^2$. So,

$$\sigma_{\Delta I^{(0)}}^2 \approx \left(MAD_{\min,org}^{(1)}\right)^2. \quad (12)$$

On the other hand, $MAD_{\min}^{(1)}$ can be described as follows:

$$MAD_{\min}^{(1)} \cong MAD_{\min,org}^{(1)} + \alpha. \quad (13)$$

Since $\eta^{(1)}(.)$ is a 2-dimensional LPF-ed version of $\eta^{(0)}(.)$, α may be small in general. So,

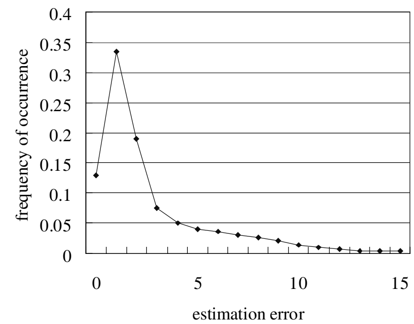

Fig. 2. Histogram of the estimation error.

$$\sigma_{\Delta I^{(0)}}^2 \approx \left(MAD_{\min}^{(1)} - \alpha\right)^2. \quad (14)$$

Thus, since $\sigma_n^2 = \sigma_{\min}^2 - \sigma_{\Delta I^{(0)}}^2$ from (11), the estimated noise variance ($\tilde{\sigma}_n^2$) of the current MB is obtained as follows:

$$\tilde{\sigma}_n^2 = \max\left(\sigma_{\min}^2 - \left(MAD_{\min}^{(1)} - \alpha\right)^2, 0\right). \quad (15)$$

The noise variance estimate of a frame can be obtained by averaging $\tilde{\sigma}_n^2$'s of all MB's in the frame according to (15). In order to minimize the effect of the motion error on the noise estimation, we need to consider the only MB whose $MAD_{\min}^{(0)}$ is less than a specific threshold T, i.e., the well-motion-compensated MB. T is adaptively determined for each frame as follows:

$$T = \min_k MAD_{\min}^{(0)}[k] + \varepsilon, \quad (16)$$

where $MAD_{\min}^{(0)}[k]$ denotes the $MAD_{\min}^{(0)}$ of the k-th MB and ε is a proper offset. T computed in the current frame is used for the next frame. ε is set to 2 empirically. Therefore, we can obtain the reliable average of $\tilde{\sigma}_n^2$'s which projects noise power of the input frame.

The proposed algorithm is summarized as follows (T is set to 10 initially):

1) Set k, num, and sum to 0. Set MIN to an infinity.
2) For the k-th MB, $MAD_{\min}^{(0)}[k]$ is compared with T. If $MAD_{\min}^{(0)}[k]$ is less than T,
$sum = sum + \max\left(\sigma_{\min}^2 - \left(MAD_{\min}^{(1)} - \alpha\right)^2, 0\right)$ and $num = num + 1$.
3) If $MAD_{\min}^{(0)}[k]$ is less than MIN, $MIN = MAD_{\min}^{(0)}[k]$.
4) If the k-th MB is a final MB, go to step 5). Otherwise, go to step 2) with $k=k+1$.
5) T is updated to $MIN + \varepsilon$. The noise variance estimate is also updated to sum/num. Then, go to step 1) for the next frame.

4. SIMULATION RESULTS

We have used an MPEG-2 video encoder merged with the DCTNR as in Fig. 1. We select the following parameters for MPEG-2 encoding; GOP (group of pictures) size = 12 frames, distance between P-frames = 1 frames, and target bit rate = 5Mbps. The horizontal/vertical search ranges of P-frames are set to ±63/±31. Well-known three MPEG-2 video sequences having various motion types are used; "football (*foot*)," "flower garden

Table I. Average values of $\tilde{\sigma}_n^2$, μ_{E_n}, and σ_{E_n}.

Video Sequence	Algorithms	σ_n^2=0			σ_n^2=25			σ_n^2=49			σ_n^2=64		
		$\tilde{\sigma}_n^2$	μ_{E_n}	σ_{E_n}	$\tilde{\sigma}_n^2$	μ_{E_n}	σ_{E_n}	$\tilde{\sigma}_n^2$	μ_{E_n}	σ_{E_n}	$\tilde{\sigma}_n^2$	μ_{E_n}	σ_{E_n}
foot	Proposed	5.04	5.04	0.72	31.4	6.44	1.56	51.5	2.73	1.83	62.9	2.06	1.58
	SONEA	0.13	0.13	0.02	26.2	7.51	7.94	53.1	13.7	11.9	65.4	16.1	13.6
fg	Proposed	3.29	3.29	0.35	24.4	0.95	0.55	42.3	6.75	1.55	52.9	11.1	2.25
	SONEA	0.001	0.001	0.004	6.7	18.1	4.41	13.1	35.9	9.03	18.4	45.7	14.3
m&c	Proposed	8.93	8.93	1.7	34.2	9.2	1.49	53.7	4.7	1.62	65.5	2.0	1.42
	SONEA	2.27	2.27	0.3	24.6	5.6	7.3	45.5	10.6	12.3	56.7	12.2	8.7

(*fg*)," and "mobile & calendar (*m&c*)." Each frame has a resolution of 720×480. By deliberately adding additive white Gaussian noise (AWGN) to the original MPEG-2 test sequences, we produced noisy video sequences whose σ_n^2's are set to 0, 25, 49, and 64.

To evaluate the performance of the proposed algorithm, the estimation error $E_n = \left|\sigma_n^2 - \tilde{\sigma}_n^2\right|$ is first calculated. E_n is the difference between the true noise variance and the estimated noise variance on a single frame. The average μ_{E_n} and the standard deviation σ_{E_n} of the estimation error are then computed as follows:

$$\mu_{E_n} = \frac{\sum_{i=1}^{N} E_n(i)}{N}, \quad \sigma_{E_n}^2 = \frac{\sum_{i=1}^{N} \left(E_n(i) - \mu_{E_n}\right)^2}{N}. \tag{17}$$

Here, N is the number of frames for each sequence.

Table I gives the evaluation results. In this experiment, α is set to 0. We could find that α of 0 rarely takes an effect on the overall estimation performance. In Table I, the proposed algorithm is compared with the structure-oriented noise estimation algorithm (SONEA) presented in [6]. The proposed algorithm is reliable for both high and low noise levels. Except when σ_n^2 is 0, the proposed algorithm is always superior to SONEA in terms of $\tilde{\sigma}_n^2$, μ_{E_n} and σ_{E_n}. It is notable that the proposed algorithm provides more accurate estimation performance as σ_n^2 becomes larger. Also, the proposed algorithm produces very small σ_{E_n}'s in comparison to SONEA. For all the test sequences, the proposed algorithm provides much more steady and accurate estimation performance than SONEA, independently of variations of motion and texture.

Table II shows that the combination of the DCTNR and the proposed noise estimation improves coding performance in an MPEG-2 video encoder. The peak signal-to-noise ratio (PSNR) is employed as an evaluation measure. NoNR stands for pure MPEG-2 TM5 without DCTNR. Ideal DCTNR assumes that actual σ_n^2's are known beforehand. Note that Ideal and the proposed algorithm provide almost the same visual quality as well as the same PSNR performance.

4. CONCLUSIONS

We propose an accurate noise estimation algorithm using multi-resolution motion estimation in a video encoder. The proposed algorithm is based on the property that the variance of an inter-MB is equivalent to the sum of the noise variance and the variance of the desired inter-MB because noise and a desired matching error signal are uncorrelated. Experimental results show that the estimated noise variances are very close to actual noise variances. Therefore, the proposed noise estimation is very useful for effective pre-filtering such as DCTNR.

Table II. PSNR performance for two noise variances [dB].

Video Sequence	Algorithms		σ_n^2=0	σ_n^2=64
foot	No NR		31.5	29.0
	DCTNR	Proposed	31.5	29.9
		SONEA	32.5	29.9
		Ideal	31.5	29.9
fg	No NR		28.5	26.5
	DCTNR	Proposed	28.5	27.4
		SONEA	28.5	26.6
		Ideal	28.5	27.5
m&c	No NR		26.3	24.9
	DCTNR	Proposed	26.3	25.6
		SONEA	26.3	25.5
		Ideal	26.3	25.6

REFERENCES

[1] O. K. Al-Shaykh and R. M. Mersereau, "Lossy compression of noisy images," *IEEE Trans. Image Processing*, vol. 7, no. 12, pp. 1641-1652, Dec. 1998.

[2] J. S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall, 1990.

[3] K. J. Boo and N. K. Bose, "A motion-compensated spatio-temporal filter for image sequences with signal-dependent noise," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 3, pp. 287-298, June 1998.

[4] S. D. Kim and J. B. Ra, "Efficient block-based video encoder embedding a Wiener filter for noisy video sequences," *Journal of Visual Communication and Image Representation*, vol. 14, no.1, pp. 22-40, March 2003.

[5] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989.

[6] A. Amer, E. Dubois, and A. Mitiche, "Reliable and fast structure-oriented video noise estimation," *IEEE International Conference on Image Processing*, vol. 1, pp. 840-843, Rochester, USA, Sept. 2002.

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

[8] B. C. Song and K. W. Chun, "Motion-compensated noise estimation for efficient pre-filtering in a video encoder," *IEEE International Conference on Image Processing*, pp. 211-214, Sept. 2003.

[9] B. C. Song and K. W. Chun, "Multi-resolution block matching algorithm and its VLSI architecture for fast motion estimation in an MPEG-2 video encoder," to be appeared in *IEEE Trans. Circ. and Syst. for Video Technol.,* April 2004.