MOTION-COMPENSATED TEMPORAL PRE-FILTERING FOR NOISE REDUCTION IN A VIDEO ENCODER

Byung Cheol Song and Kang Wook Chun

Digital Media R&D Center, Samsung Electronics Co., Ltd.
416 Maetan-3dong, Yeungtong-gu, Suwon-City, Republic of Korea
E-mail: bcsong@samsung.com

ABTRACT

For coding efficiency as well as noise reduction, efficient pre-filtering needs to be performed prior to video encoding. This paper presents a DCT-domain temporal filtering scheme in an MPEG video encoder. It is proven that the multiplication of every DCT coefficient in an inter block with a proper weight is equivalent to motion compensated temporal filtering in the spatial domain for the inter block. Also, we propose to determine properly the weight by using effective noise estimation scheme. Experimental results show that the proposed pre-filtering scheme provides outstanding coding efficiency as well as noise reduction in an MPEG video encoder.

1. INTRODUCTION

Generally, input video sequences of a video encoder can be deteriorated due to various noise sources. A main source of contamination for video sequences is the additive noise introduced by the capture device, e.g., photonic noise in the case of charge-coupled devices (CCD) camera. Analogue video sequences, which are often degraded due to channel noise in the conventional analogue transmission systems, needs to be sometimes digitally encoded in the receiver side. Noisy video sequences are not only visually annoying, but they are also hard to be encoded efficiently owing to uncorrelated nature of noise. Therefore, although pre-filtering is not widely addressed in the literature up to now, it is commonly utilized in video coding systems. The main goal of pre-filtering is to remove as much high-frequency information as possible without compromising visual quality.

Noise removal schemes for video sequences have been widely studied [1-6]. Among them, a spatial-domain adaptive Wiener filtering scheme [3] and a motion-compensated spatio-temporal filtering scheme [5] have good de-noising performance. 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, which is gracefully embedded with a video coder [6]. They first applied the concept of the generalized Wiener filter [7] to a video encoder. This transform-domain Wiener filtering (TDWF) 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 TDWF noticeably outperforms the other pre-filtering schemes such as a spatial-domain adaptive Wiener filter [3] in terms of noise reduction and coding efficiency [6].

Most pre-filtering algorithms including the TDWF assume that the noise strength such as noise variance is already known prior to encoding. Precise noise estimation is required for effective pre-filtering, but actual noise strength is practically impossible to be known in advance. We proposed the very accurate noise estimation algorithm, which keeps negligible computational complexity in a video encoder [8]. Note that the noise variance predicted by this noise estimator is useful for pre-filtering.

This paper presents a transform-domain temporal filtering (TDTF) scheme in a video encoder. We show that the multiplication of every DCT coefficient in an inter block with a proper weight is equivalent to the motion-compensated temporal filtering in the spatial domain. The weight is a single value for every DCT coefficient. It is very important to choose an appropriate weight for each DCT block in terms of coding performance. The weight value depends on the noise variance and the motion-compensated error. We propose to determine properly the weight by using the noise variance predicted from the

0-7803-8554-3/04/$20.00 ©2004 IEEE.

above-mentioned noise estimator [8]. Experimental results show that the proposed pre-filtering scheme provides outstanding coding efficiency as well as noise reduction in an MPEG encoder while requiring negligible computational complexity.

The organization of this paper is as follows. Section 2 presents the spatial-domain motion-compensated temporal filtering scheme. Section 3 describes the proposed algorithm. Section 4 gives simulation results. Finally, Section 5 provides conclusion.

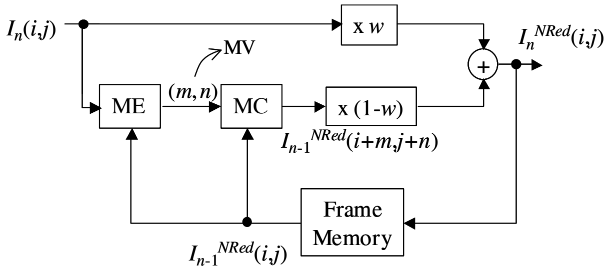

Fig. 1. The motion-compensated temporal filtering in the spatial domain.

2. MOTION-COMPENSATED TEMPORAL FILTERING IN THE SPATIAL DOMAIN

Fig. 1 describes the motion-compensated temporal filtering in the spatial domain. Here, $I_n(i,j)$ denotes the pixel value located at (i,j) in the current frame, i.e., the n-th frame. $I_n^{NRed}(i,j)$ denotes its filtered output pixel value, which is stored into a frame memory and is used for the pre-filtering of the next frame.

Motion vector (MV) corresponding to the current pixel $I_n(i,j)$, i.e., (m,n) is predicted by motion estimation for the previous pixels within a proper search range. And then, the motion-compensated pixel of $I_n(i,j)$, that is, $I_{n-1}^{NRed}(i+m,j+n)$ is found.

Temporal filtering is performed by the weighted averaging of the corresponding two pixels: $I_n(i,j)$ and $I_{n-1}^{NRed}(i+m,j+n)$, which is depicted as follows:

$$I_n^{NRed}(i,j) = w \cdot I_n(i,j) + (1-w)\cdot I_{n-1}^{NRed}(i+m,j+n), \qquad (1)$$

where the weight w can be adaptively varied from 0 to 1. For example, if each pixel is motion-compensated ideally, all the noise components can be removed statistically by this recursive filtering [3, 5, 7].

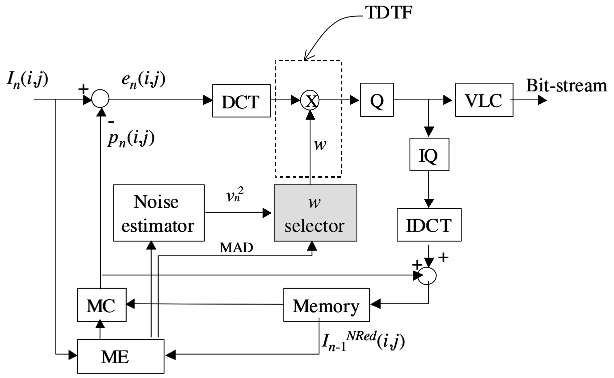

Fig. 2. A video encoder employing the TDTF and the noise estimator [8].

3. THE PROPOSED SCHEME

In this paper, the proposed TDTF is implemented in an MPEG video encoder for noise reduction (see Fig. 2). This video encoder framework employs a motion-compensated noise estimation algorithm for effective pre-filtering [8]. The noise estimator is based on a multi-resolution block-matching algorithm (MRBMA) that provides high computational speed and high estimation performance concurrently [9]. Since MRBMA is used for motion estimation (ME) in this encoder framework, we describe MRBMA briefly and propose TDTF.

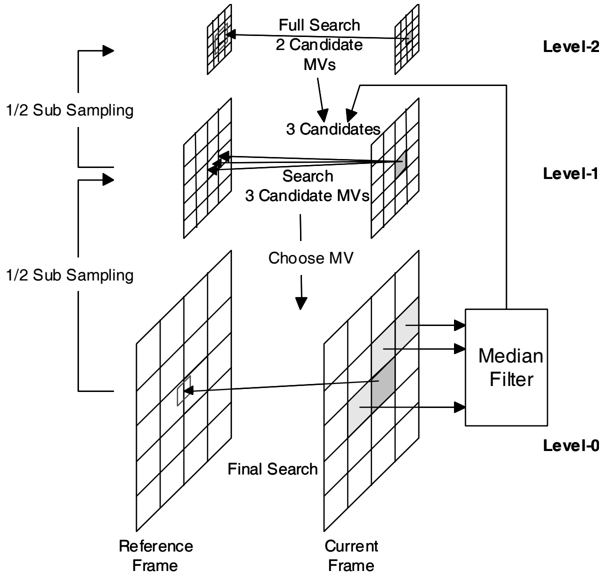

Fig. 3. The overall scheme of MRBMA.

3.1. MRBMA

MRBMA consists of three levels (see Fig. 3). Prior to ME processing, the coarser level frames of current frame as well as reference frame should be produced.

The coarser level frames are constructed by a certain low-pass-filter (LPF) and sub-sampling as in Fig. 3. So, the MB size becomes 16x16, 8x8, and 4x4 at level 0, 1, and 2, respectively. At level 2, two 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 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 local search around the single candidate. MRBMA outperforms the other multi-resolution block-matching algorithms in terms of estimation accuracy and computational complexity. So, we employ the MRBMA for ME in a video encoder of Fig. 2.

3.2. The Proposed Algorithm

Let $\mathbf{I}_n$ and $\mathbf{p}_n$ be the current macroblock (MB) and its motion-compensated MB (see Fig. 2). $I_n(i,j)$ and $p_n(i,j)$ denote the pixel included in $\mathbf{I}_n$ and $\mathbf{p}_n$, respectively. If the MV of $\mathbf{I}_n$ is (m,n), $p_n(i,j)$ is equal to $I_{n-1}^{NRed}(i+m,j+n)$. Assume that the reference frame is the $(n-1)$-th frame. From Eq. (1),

$$I_n^{NRed}(i,j) = w \cdot \{I_n(i,j) - p_n(i,j)\} + p_n(i,j). \quad (2)$$

Since $e_n(i,j) = I_n(i,j) - p_n(i,j)$,

$$I_n^{NRed}(i,j) = w \cdot e_n(i,j) + p_n(i,j). \quad (3)$$

Finally, if the pre-filtered current MB is an inter MB, $I_n^{NRed}(i,j) - p_n(i,j)$ should be DCT-ed. So, from Eq. (3),

$$DCT\{I_n^{NRed}(i,j) - p_n(i,j)\} = DCT\{w \cdot e_n(i,j)\}. \quad (4)$$

Since w is a constant,

$$DCT\{I_n^{NRed}(i,j) - p_n(i,j)\} = w \cdot DCT\{e_n(i,j)\}. \quad (5)$$

Therefore, if the pre-filtered current MB is an inter MB, the motion-compensated temporal filtering is achieved by multiplying a specific weight w with every coefficient of the DCT-ed inter block. Note that the proposed TDTF is a sort of block-based temporal filtering unlike spatial domain temporal filtering.

Another merit of the TDTF has no computational complexity for filtering itself because the filtering is gracefully merged with the quantization procedure (Q). Assuming that the limited number of w values is permitted for implementation, the quantization table corresponding to each w can be prepared in advance. So, if w is determined, we can just choose a proper quantization table for the w without additional computation for pre-filtering. Note that every DCT coefficient corresponds to a different weight in the TDWF. So, TDTF is computationally simpler than TDWF.

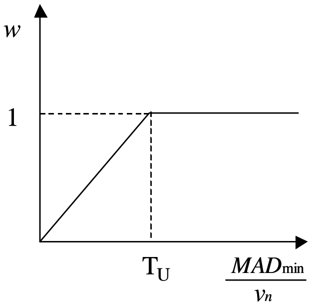

Fig. 4. Decision of w.

3.2.1 Decision of *w*

As mentioned previously, the performance of the TDTF depends on w. So, we have to carefully apply a proper w to each DCT block. Assume that the noise variance of the current frame, i.e., v_n^2 can be predicted by the noise estimator [8]. Minimum MAD (Mean of Absolute Difference) of each MB is also obtained from the motion estimator. Let $\mathrm{MAD}_{\min}$ be the minimum MAD. In general, strong filtering needs to be applied to the well-motion-compensated block. On the contrary, the weak filtering should be applied to the poorly motion-compensated block. On the other hand, the strength of filtering is proportional to v_n without the loss of generality. Thus, w is proportional to $MAD_{\min}/v_n$. So, we derive the graph for determining w like Fig. 4. The threshold $\mathrm{T_U}$ can be emprically determined. If w is too small, the ghost artifact often happens. To prevent this phenomenon, the lower bound of w is required. Thus,

$$w = \max\{w, B_L\}. \quad (6)$$

A threshold B_L is also emprically determined.

4. SIMULATION RESULTS

We have used an MPEG-2 video encoder merged with the TDTF as in Fig. 2. 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 search ranges of P-frames are set to [-64, +64]. The vertical search range is a half of the horizontal counterpart. The first 100 frames of four MPEG-2 video sequences having various motion types are used; “football (*foot*),” “mobile and calendar (*mob*),” “flower garden (*fg*),” and “table tennis (*table*).” Each frame has a resolution of 720x480. By deliberately adding additive white Gaussian noise (AWGN) to the original MPEG-2 test sequences, we produced noisy video sequences whose v_n^2’s are set to 0, 25, and 64. T_U and B_L are set to 2 and 0.125, respectively. The peak signal-to-noise ratio (PSNR) is employed as an evaluation measure.

Table I. Coding performance [dB].

	v_n^2	NoNR	TDWF	TDTF
foot	64	29.29	30.21	30.08
	25	30.72	31.28	31.20
	0	31.93	32.15	31.86
fg	64	26.80	27.92	28.11
	25	27.92	28.65	28.78
	0	28.86	29.38	29.52
mob	64	24.89	25.54	25.80
	25	25.51	26.09	26.30
	0	26.04	26.49	26.70
table	64	27.93	29.26	29.61
	25	29.26	30.44	30.74
	0	30.76	31.70	31.63

Table I shows that the TDTF noticeably improves coding performance in an MPEG-2 video encoder. Coding performance of the proposed TDTF is compared with NoNR and the TDWF [6]. NoNR stands for pure MPEG-2 TM5 without any pre-filtering. PSNR's of the TDTF are much higher than those of NoNR. For example, the TDTF outperforms NoNR by about 2dB for "table tennis" sequence in case of $v_n^2=64$. Note that the TDTF is comparable with the TDWF in terms of PSNR performance. We can find that the TDTF is usually better than the TDWF, while the computational complexity of the TDTF is much lower than that of TDWF as mentioned in Section 3.

In addition, Fig. 5 compares PSNRs of the TDTF with those of the TDWF and the NoNR for various bit-rates, i.e., 3, 4, 5, and 6Mbps. Note that for the same visual quality, e.g., PSNR of about 28 dB, the TDTF as well as the TDWF requires only a half of the bit-rate of NoNR. It is also shown that the TDTF can improve the compression ratio up to 10 percent in comparison to the TDWF.

5. CONCLUSIONS

For coding efficiency as well as noise reduction, efficient pre-filtering needs to be performed prior to video encoding. This paper proposed a transform-domain temporal filtering scheme (TDTF) in an MPEG video encoder. We showed that the multiplication of every DCT coefficient in an inter block with a proper weight is equivalent to motion-compensated temporal filtering in the spatial domain. Also, we proposed to determine properly the weight by using effective noise estimation scheme for effective filtering. Experimental results show that the proposed pre-filtering scheme provides outstanding coding efficiency as well as noise reduction in an MPEG encoder.

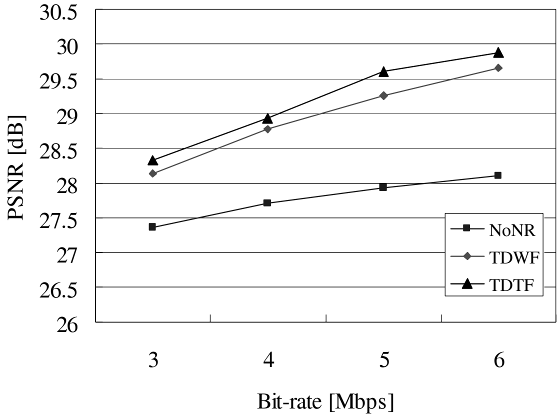

Fig. 5. PSNR performance for various bit-rates.

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] N. Vasconcelos and R. Dufaux, "Pre and post-filtering for low bit-rate video coding," *Proc. Int. Conf. Image Processing*, vol. 1, pp. 291-294, Santa Barbara, California, Oct. 1997.

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

[4] M. Biloslavo, G. Ramponi, S. Olivieri, and L. Albani, "Joint Kalman-based noise filtering and motion-compensated video coding for low bit-rate videoconferencing," *IEEE International Conference on Image Processing*, pp. 992-995, 2000.

[5] 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.

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

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

[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*, vol. 2, pp. 211-214, 2003.

[9] J. Lee, K. W. Lim, B. C. Song, and J. B. Ra, "A fast multi-resolution block matching algorithm and its LSI architecture for low bit-rate video coding," *IEEE Trans. Circ. Syst. for Video Technol.,*vol. 11, no. 12, pp. 1289-1301, Dec. 2001.