

# AN ENCODER-EMBEDDED VIDEO DENOISING FILTER BASED ON THE TEMPORAL LMMSE ESTIMATOR

Liwei Guo<sup>†</sup>, Oscar C. Au<sup>†</sup>, Mengyao Ma<sup>‡</sup> and Zhiqin Liang<sup>†</sup>

<sup>†</sup>Dept. of Electrical and Electronic Engineering, <sup>‡</sup>Dept. of Computer Science  
Hong Kong University of Science and Technology  
email: {eeglw, eeau, myma, zhiqin}@ust.hk

## ABSTRACT

Noise not only degrades the visual quality of video contents, but also significantly affects the coding efficiency. Based on the temporal linear minimum mean square error (LMMSE) estimator, an innovative denoising filter is proposed in this paper. The proposed filter only requires simple operations manipulating on the individual residue coefficients and can be seamlessly integrated into video encoders. Compared to traditional filter-encoder cascaded scheme, embedding the proposed filter into the video encoder can save a large amount of computation. The experimental results show that with the proposed filter embedded, both the noise suppression capability and the coding efficiency of the video encoder can be dramatically improved. Furthermore, as a purely temporal filter, it can well preserve the fine details of video contents and satisfactory visual quality can be achieved.

## 1. INTRODUCTION

The importance of video technology is constantly growing with the ever increasing use of television and video systems in consumer, commercial, medical and scientific applications. However, digital video sequences are always contaminated by noise, during acquisition, recording, processing and transmission. The presence of noise not only degrades the subjective quality, but also affects the efficiency of subsequent processing, especially for video coding.

For conventional hybrid video coding architecture, temporal redundancy is exploited by motion estimation (ME) and motion compensation (MC) to achieve high coding efficiency. However, noise is inherently independent among frames. Thus the presence of noise reduces temporal redundancy among consecutive frames and this can decrease the accuracy of the motion vectors (MVs) obtained in ME. Moreover, even if the motion vectors are accurate, the noise will make the residue frame noisy. As the noise typically exhibits little spatial redundancy, the energy of the noisy residue frame can not be compacted by discrete cosine transform (DCT). As a result coding performance can drop very significantly, with a large part of bit rate wasted on coding the noise. It is thus highly desirable to remove the noise while preserving the original video contents.

In order to increase visual quality and coding efficiency, much effort has been made on developing video denoising methods. Conventionally, denoising filtering is considered as a process independent of video encoding, and most denoising-coding systems apply denoising filters [1–4] before the video encoding. Obviously, the complexity of such a cascaded scheme is the superimposition of the filter and the encoder, which is always very high. In this paper, an innovative linear temporal denoising filter which can be seamlessly embedded into the video encoder is proposed. The proposed filter

not only shares ME module with the video encoder, but also reduces the temporal filtering to very simple linear operations on residue coefficients. Compared to the cascaded scheme, the proposed embedded scheme is much simpler and a lot of computation can be saved. As an approximation to the temporal LMMSE estimator, the proposed filter can achieve much better denoising performance than other encoder-embedded denoising filters, such as the transform domain temporal filter (TDTF) [5]. Furthermore, as it is a purely temporal filter, blurring will not be caused in the filtering and fine spatial details can be well preserved.

This paper is organized as follows. Section 2 presents the proposed denoising filter based on the temporal LMMSE estimator. Section 3 describes the implementation of the proposed filter. Experimental results are shown in section 4, and section 5 gives a conclusion to this paper.

## 2. THE PROPOSED ENCODER-EMBEDDED DENOISING FILTER

### 2.1. Mathematical model

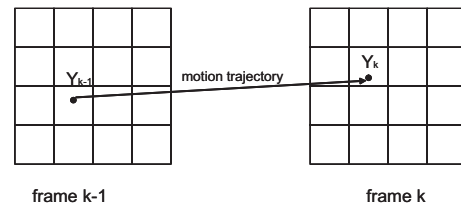


Fig. 1. System model

Fig. 1 shows two neighboring frames with time indices  $k-1$  and  $k$ . Here we consider tracking a point object denoted by the black dot in the figure. Suppose this sequence is contaminated by additive Gaussian noise, the observed values  $Y_{k-1}$  and  $Y_k$  are not the actual intensity values  $X_{k-1}$  and  $X_k$  but the summation of original intensity and noise, which can be expressed as follows,

$$Y_k = X_k + N_k \quad (1)$$

where  $N_k$  represents the undesirable noise in frame  $k$  that should be suppressed.  $N_k$  is assumed to be zero-mean with variance  $\sigma_k^2$ . This additive noise is independent over space and time, and independent of  $X_k$ . As this point object is corrupted by noise,  $X_k$  is unknown and modelled as a random variable with mean  $\bar{X}_k$  and variance  $\sigma_{X(k)}^2$ .

Due to inaccuracy of motion estimation, change in light condition and change in the object itself, the pixel intensity is changing

from frame to frame. The pixel intensity values of two consecutive frames can be expressed as follows

$$X_k = X_{k-1} + Z_{k-1} \quad (2)$$

where  $Z_{k-1}$  is the residue and assumed to be independent of  $N_k$  and  $X_{k-1}$ . Traditionally,  $Z_{k-1}$  is modelled as a zero-mean random variable with variance  $\sigma_{Z(k-1)}^2$  [3]. In this paper non-zero mean  $\bar{Z}_{k-1}$  is assumed for  $Z_{k-1}$  to make our model more accurate.

Equations (1) and (2) can be set up for each pixel in a frame. Our model assumes that each of these pixels can be tracked in the video sequence all the time. Sometimes this assumption fails due to scene change or object overlapping. In these cases, these two equations still hold by introducing large residue variance.

## 2.2. The temporal LMMSE estimator

With regards to filters, "optimal" is always referred to as minimum mean square error (MMSE). LMMSE estimator is such an optimal filter which only has linear operations. Our target is to find the LMMSE estimator of  $X_k$  based on the two latest observations,  $Y_k$  and  $Y_{k-1}$ :

$$Y_{k-1} = X_{k-1} + N_{k-1}, \quad (3)$$

$$Y_k = X_k + N_k. \quad (4)$$

As random variables  $Y_k$  and  $Y_{k-1}$  have different means, the target filter is a non-homogenous one and a constant  $d$  is included to reduce the estimation error. The LMMSE estimator has the following form

$$\tilde{X}_k = w_1 Y_{k-1} + w_2 Y_k + d \quad (5)$$

where  $\tilde{X}_k$  is the estimated value of  $X_k$ ,  $w_1$  and  $w_2$  are the weights of input data.

The estimation error is denoted as

$$e_k = X_k - \tilde{X}_k \quad (6)$$

To minimize  $E[e_k^2]$ , orthogonality must be satisfied, meaning the estimation error must be orthogonal to the input data. Combining equations (3), (4), (5), and (6) with the orthogonality, the coefficients of the temporal LMMSE estimator are given as follows:

$$w_1 = \frac{\sigma_k^2 \sigma_{X(k-1)}^2}{\Delta}, \quad (7)$$

$$w_2 = \frac{\sigma_{X(k)}^2 \sigma_{k-1}^2 + \sigma_{X(k-1)}^2 (\sigma_{X(k)}^2 - \sigma_{X(k-1)}^2)}{\Delta}, \quad (8)$$

$$d = (1 - w_2) \bar{X}_k - w_1 \bar{X}_{k-1}. \quad (9)$$

where

$$\Delta = (\sigma_{X(k-1)}^2 + \sigma_{k-1}^2)(\sigma_{X(k)}^2 + \sigma_k^2) - \sigma_{X(k-1)}^2 \sigma_{X(k)}^2.$$

Since  $d$  has very complex form, here it is expressed in terms of  $w_1$ ,  $w_2$  and  $\bar{X}_k$ ,  $\bar{X}_{k-1}$ .

## 2.3. The proposed embedded temporal filter based on the temporal LMMSE estimator

The above filter is supposed to be the optimal one among all the linear temporal filters. However, it can not be directly embedded into the video encoder. Note that conventionally, only the processed frames are stored in the system, which means  $Y_{k-1} = \tilde{X}_{k-1}$  and the actual estimator should be expressed as follow,

$$\tilde{X}_k = w_1 \tilde{X}_{k-1} + w_2 Y_k + d \quad (10)$$

For video denoising system,  $\tilde{X}_{k-1}$  is the denoised version of  $X_{k-1}$  and it is reasonable to assume there is no noise in  $\tilde{X}_{k-1}$ , meaning  $N_{k-1}$  is 0. As  $Z_{k-1}$  is assumed to be independent of  $X_{k-1}$ ,

$$\bar{Z}_{k-1} = \bar{X}_k - \bar{X}_{k-1}, \quad (11)$$

$$\sigma_{Z(k-1)}^2 = \sigma_{X(k)}^2 - \sigma_{X(k-1)}^2. \quad (12)$$

Substituting equations (11), (12) and  $\sigma_{k-1}^2 = 0$  into equations (7), (8) and (9), the filter can be dramatically simplified:

$$w_1 = \frac{\sigma_k^2}{\sigma_{Z(k-1)}^2 + \sigma_k^2}, \quad (13)$$

$$w_2 = \frac{\sigma_{Z(k-1)}^2}{\sigma_{Z(k-1)}^2 + \sigma_k^2}, \quad (14)$$

$$d = w_1 \bar{Z}_{k-1}. \quad (15)$$

The mean square estimation error is

$$E[e_k^2] = \frac{1}{\sigma_{Z(k-1)}^{-2} + \sigma_k^{-2}} \quad (16)$$

Now the impact of each parameter can be analyzed. On one hand,  $w_2$  decreases as  $\sigma_k^2$  increases. This is consistent with the fact that if the current observation is heavily corrupted with noise, low confidence should be given. On the other hand, large  $\sigma_{Z(k-1)}^2$  forces  $w_1$  to be small. Note that  $\sigma_{Z(k-1)}^2$  is the variance of the difference between  $X_k$  and  $\tilde{X}_{k-1}$ . If this variance is large, probably scene change or object overlapping happens. In these cases, consecutive frames may be very different and involving pixels in the previous frame can not suppress noise but incur "ghosting". So the contribution of  $\tilde{X}_{k-1}$  to the estimation must be limited. The constant  $d$  is used to compensate the mean shift between consecutive frames. Note that when there is no noise ( $\sigma_k^2 = 0$ ), no filtering will be applied ( $w_1 = d = 0$ ,  $w_2 = 1$ ), so the proposed filter will not affect clean video.

Note that the summation of  $w_1$  and  $w_2$  is 1. Combing with equations (2), (4) and (10), this filter can be rewritten as follows:

$$\begin{aligned} \tilde{X}_k &= w_1 \tilde{X}_{k-1} + w_2 (\tilde{X}_{k-1} + Z_{k-1} + N_k) + w_1 \bar{Z}_{k-1} \\ &= \tilde{X}_{k-1} + w_2 Z_{k-1} + w_1 \bar{Z}_{k-1} \end{aligned} \quad (17)$$

where  $Z_{k-1} = Z_{k-1} + N_k$ . Note that the mean of  $Z_{k-1}$  is also  $\bar{Z}_{k-1}$ , and the variance is  $(\sigma_{Z(k-1)}^2 + \sigma_k^2)$  as  $Z_{k-1}$  and  $N_k$  are independent.

Suppose there is a video encoder and  $Y_k$  is the pixel to be encoded,  $\tilde{X}_k$  is the output corresponding to  $Y_k$ . Here,  $\tilde{X}_{k-1}$  is the prediction of  $Y_k$  in the previous frame and  $Z_{k-1}$  is the noisy residue. The proposed filter can be implemented by adding scaled residue ( $w_2 Z_{k-1}$ ) and a constant ( $w_1 \bar{Z}_{k-1}$ ) to the prediction ( $\tilde{X}_{k-1}$ ). So this filter can be gracefully incorporated into the ME/MC module of video encoders.

## 3. IMPLEMENTATION

A video encoder with the proposed filter embedded is presented in Fig.2. The pixel to be encoded is  $Y_k$ . After ME,  $\tilde{X}_{k-1}$  is found as the prediction of  $Y_k$ , and the residue  $Z_{k-1}$  is extracted. Based on the parameter estimation, a specific weight  $w_2$  is calculated to scale down  $Z_{k-1}$ . Meanwhile, another term  $w_1 \bar{Z}_{k-1}$  is added back to the scaled down residue to compensate mean fluctuation. With MC,

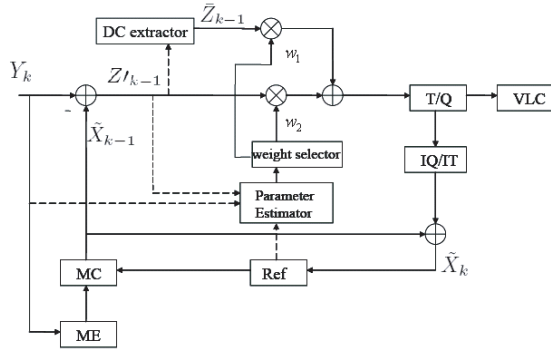


Fig. 2. Video encoder with the proposed filter embedded

$\tilde{X}_k$ , which will be the output at the decoder, is obtained and stored in the reference buffer for following filtering and compression. Note that  $w_1$  and  $w_2$  are only utilized in the encoder-side, no need to be transmitted. And there is no modification for the decoder.

Accurate parameters are of significant importance to denoising performance. The following parameters of the filter have to be estimated :

1.  $\sigma_k^2$ . This is the variance of undesirable noise. As the noise is assumed to be uniform in the whole frame, smooth regions must have smaller variances than other regions and these variances are mainly contributed by noise. To estimate  $\sigma_k^2$ , the frame is divided into blocks of same size. As too large block might overlap the texture region while too small block does not have enough samples to reveal the statistics, the size of the block must be carefully designed. According to our experiments, 16x16 is a suitable choice. Then for each block the variance is calculated, and the average of the minimum 10 variances is used as the variance of noise.

2.  $\bar{Z}_{k-1}$  and  $\sigma_{Z(k-1)}^2$ . They are the mean and the variance of  $Z_{k-1}$ , respectively. After ME, the noisy residue  $Z_{k-1}$  is obtained.  $Z_{k-1}$  is a random variable and what we can observe is one of its realizations. As neighboring residues probably have similar characteristics, they can be seen as various realizations of  $Z_{k-1}$ , meaning that ensemble characteristics can be estimated by spatial characteristics. Specifically, the residue frame is divided into 16x16 blocks, and the mean and the variance are calculated for every block. As  $N_k$  is zero-mean, the mean of the block is used as  $\bar{Z}_{k-1}$ . Subtracting  $\sigma_k^2$  from the block variance,  $\sigma_{Z(k-1)}^2$  can be obtained. Obviously the parameters are block-wise and adaptive over space.

#### 4. EXPERIMENTAL RESULTS

Experiments are conducted to evaluate the proposed denoising filter. Three test video sequences *Akiyo*, *Foreman*, and *News* in CIF format are used. Gaussian noises with variances 49 and 169 are added to the luminance component of these test sequences. The first 100 frames of each sequence are encoded by original JM 8.2 (JM) and the JM 8.2 embedded with the proposed filter (Proposed), respectively. For comparison, another encoder-embedded denoising filter TDTF [5] is also embedded into JM 8.2 (JM\_TDTF) to encode these noisy sequences. The first frame is I frame and the rest are P frames with rate control enabled.

The comparison of coding performance are shown in Fig.3 and 4. It can be seen that the video encoder with the proposed filter

embedded has much better rate-distortion (R-D) performance than JM\_TDTF and JM. It is interesting to note for JM, sometimes the coding performance is better in low bitrates, especially when noise is serious. The same observation was reported by Kim in [6]. This is because when bit rate is low, some DCT coefficients contributed by noise are quantized to zero. So coarse quantization provides a filtering effect and partially removes noise.

To evaluate the subjective quality of the encoded video, some reconstructed frames are shown in Fig.5 and 6. Although quantization can remove some noise, the frames encoded by JM are still very noisy. And a lot of "mosaics" are observed due to high frequency DCT coefficients from noise. JM\_TDTF has better noise suppression capability than JM. However, some fine details are blurred in the filtering, such as the teeth of Foreman. With the proposed filter embedded, the noise suppression capability of the video encoder can be dramatically improved. Furthermore, as the proposed filter is a purely temporal filter and the filtering strength is determined based on the LMMSE estimator, the spatial details can be well preserved and satisfactory visual quality can be achieved.

#### 5. CONCLUSIONS

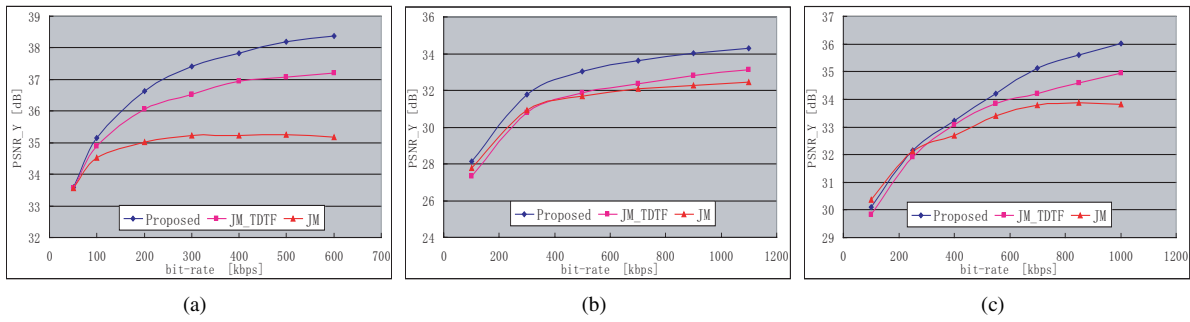
In this paper, an innovative denoising filter based on the temporal LMMSE estimator is presented. The proposed filter only requires simple operations on the residue coefficients. Such feature makes it suitable to be embedded into the traditional hybrid video encoders and a lot of computation can be saved. Experimental results show that incorporating the proposed filter into video encoders can dramatically improve both the subjective quality and the coding efficiency of the noisy sequences.

#### 6. ACKNOWLEDGMENT

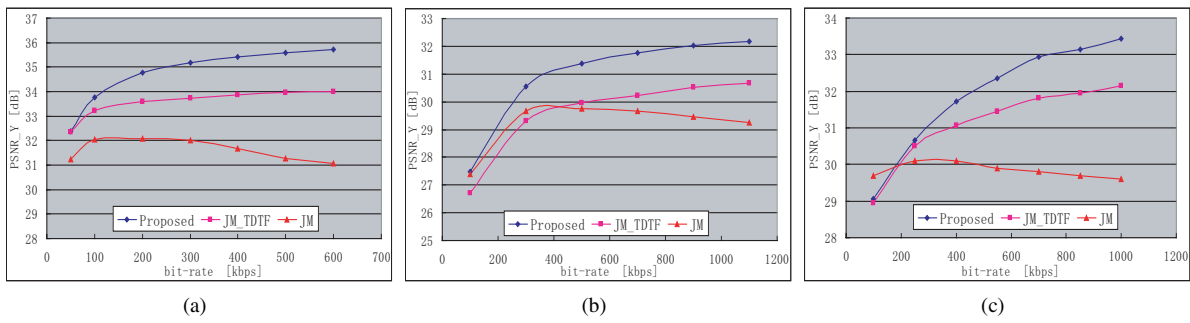
This work has been supported by the Innovation and Technology Commission of the Hong Kong Special Administrative Region, China (project no. ITS/122/03 and project no. GHP/033/05).

#### 7. REFERENCES

- [1] K.J.Boo and N.K.Bose, "A motion-compensated spatio-temporal filter for image sequences with signal-dependent noise," in *IEEE Trans. Circuits Syst. Video Technol.*, June 1998, vol. 8, pp. 287–298.
- [2] Mehmet K.Ozkan, M.Ibrahim Sezan, and A.Murat Tekalp, "Adaptive motion-compensated filtering of noisy image sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 277–290, Aug. 1993.
- [3] R.Dugad and N.Ahuja, "Video denoising by combining kalman and wiener estimates," in *Proc.ICIP'99*, Oct. 1999, pp. 152–156.
- [4] Tai-Wai Chan, Oscar C. Au, Tak-Song Chong, and Wing-San Chau, "A novel content-adaptive video denoising filter," in *Proc. ISCAS'05*, May 2005, pp. 6060–6283.
- [5] B.C.Song and K.W.Chun, "Motion-compensated temporal filtering for denoising in video encoder," *Electron. Lett.*, vol. 40, Jun 2004.
- [6] S.D.Kimand S.K. Jang, M.J.Kim, and J.B.Ra, "Efficient block-based coding of noise images by combining pre-filtering and dct," in *Proc. ISCAS'99*, May 1999, pp. 37–40.



**Fig. 3.** R-D performance, noise variance=49 a) *Akiyo*, b) *Foreman*, c) *News*.



**Fig. 4.** R-D performance, noise variance=169 a) *Akiyo*, b) *Foreman*, c) *News*.



**Fig. 5.** Frame 58 of *Foreman* (corrupted by noise with variance 49), encoded at 900kbps.



**Fig. 6.** Frame 14 of *News* (corrupted by noise with variance 169), encoded at 700kbps.