VIDEO ERROR CONCEALMENT BASED ON DATA HIDING IN THE 3D WAVELET DOMAIN

F. Battisti, M. Carli, A. Neri

Universita' degli Studi Roma TRE
Applied Electronics Dept.
Roma, Italy

ABSTRACT

In this contribution a novel method for video error control and concealment is proposed. The system is based on the use of data hiding techniques for transmitting the extra information needed to recover the data lost during transmission. The embedded data is a sub sampled binary version of each key frame of the shot. The embedding is performed by using the Quantization Index Modulation (QIM) scheme in the 3D wavelet transform domain. The experimental results show the effectiveness of the proposed approach.

***Index Terms*—** Data hiding, video error concealment, multimedia communication.

1. INTRODUCTION

Digital video is a commonly used medium for sharing information among people. Significant developments in compression and transmission techniques have been performed in the last decades, which have made possible to deliver high quality video through different kind of networks. In particular, the advent of digital video systems is making possible the creation of new telecommunication services (e.g., direct broadcast satellite, digital television, high definition TV, video teleconferencing, and Internet TV). Digital video is a very flexible media: it can be enhanced, transmitted, translated across different standards, and displayed by using several different devices.
However, the delivering of video over noisy communication channels may cause the degradation of the transmitted data due to packet losses, delay jitter, protocol errors, buffer overflows, etc.. These degradations may produce artifacts which can impair the perceived quality of the received video. Therefore, there is the need for error concealment techniques that can detect and conceal the effects of communications errors resulting in a more satisfying perceived quality.
To this aim, many solutions have been proposed in literature for developing robust image and video coding techniques [1] [2]. According the role of encoder-decoder in the recovering errors, they can be classified into three main categories:

- *Forward Error Correction* if only the encoder is employed to introduce enough redundancy in the stream for correcting errors,
- *Error Concealment* when only post-processing techniques are employed at the decoder for masking errors,
- and *Interactive Error Control* when the encoder and the decoder work together in error handling.

In particular, error concealment methods are usually realized in a two steps paradigm. In the first phase, based on some spatial and temporal information, the presence of an error is detected. Then, the received video stream is modified to recover as much as possible of the quality of the original data.
Several error concealment techniques using either statistical methods to detect and correct errors or that depend on critical information from the transmitter, like resynchronization markers, have been proposed in literature. Some of them employ blockwise deterministic models in spatial or transform domain for accounting the spatial dependency within images [3] [4]. In [5], the authors estimate the missing Discrete Cosine Transform (DCT) coefficients based on spatial smoothing constraint on the reconstructed blocks. Other approaches are based for example on edge models for directional interpolation in the spatial domain [6] or fuzzy logic reasoning [7].

The authors in previous works [8] [9] proposed a different approach for error concealment based on the use of data hiding. These techniques are usually adopted for security purposes [10]. Generally, a copyright information, or mark, is embedded into the original data before its distribution over the network, so that no other user is able to produce a copy identical to the original without the embedded message [1].
To be effective every embedding system should satisfy three main constraints:

- Invisibility: the mark should not affect the perceptual quality of the video and should not produce noticeable distortions in the received data.
- Robustness (to data alterations): the mark should not be modified by malicious (an attempt to alter the mark) or

unintentional (compression, transmission, or filtering) operations.

- Security: the mark should not be removed from the video, even if the embedding scheme is known.

These constraints are not independent. The increase of the robustness, for example, generally increases the visibility of the mark.
We propose to use the data hiding framework for transmitting extra information inside the data itself, without significantly increase of the amount of data to be transmitted. More in details, in our application the mark is a low dynamic (4-bits), low resolution version of the most representative frame of a shot, named key frame. The recovery of the mark, after the transmission process, can help in increasing the overall perceived quality, if any of the frames in the shot, due to channel error or losses, are impaired. In fact, at the receiver side, the mark is extracted from the decoded video. The eventually corrupted frames are then restored with the aid of the hidden key frame. Specific areas that are lost through transmission are selected from the extracted mark and replaced in the original frame, thus enhancing its perceptual quality.
Examples of the use of data hiding as an error control tool for video transmission over error-prone channels can be found in state of the art works such as in [8] [11] [12].
The proposed method differs from our previous works because the hidden information is directly a low resolution version of key frames of the video sequence. This information is hidden into the video, before the encoder. The receiver can then use this information to conceal impulsive errors. The detected information is also useful to conceal multiple frame losses, by inserting the extracted and resized mark in place of the missed frames.
The proposed method uses the Quantization Index Modulation (QIM) embedding algorithm applied in the three dimensional Discrete Wavelet Transform domain (3D-DWT) [13]. Thanks to the computational simplicity of the QIM detection algorithm, the embedded information can be easily extracted from the video. In this way, the bit stream has an extra functionality for error concealment, without increasing the bandwidth occupation during the file transfer. Since the presence of this additional content is not perceived by a human user, decoders can even ignore this functionality. Furthermore, the DWT presents lower computation complexity than the previously adopted DCT. A comparison of the proposed technique to previous variant is provided along with an extensive analysis of its performance.

The paper is organized as follows. In Section 2, the proposed watermarking method for error concealment is detailed. Simulations have been performed to analyze the visibility of the embedding method and the robustness against the H.264/AVC coder [14]. Section 3 presents the experimental results and the comparison with previous methods. Conclusions are drawn in Sections 4.

2. KEY FRAME HIDING IN THE DWT DOMAIN

In the proposed concealment system, the original video sequence is partitioned into sets of consecutive frames, called shots. Let $N_r \times N_c$ pixels be the frame size; a shot consisting of N_f frames can be represented by a 3D matrix V of size $N_r \times N_c \times N_f$ pixels. The optimal number of frames to be considered in a shot depends on several factors as the dynamics of the scene.
Video segmentation into shots is a topic that has been deeply investigated during the recent past in the framework of video indexing, synthetic summary generation, and video retrieval. Shot boundary detection and evaluation of the key frame that best represents the shot content is beyond the scope of the present work. Here, we assume that a state of the art video-segmentation algorithm is incorporated into the data hiding scheme, while we focus our discussion on the hiding of the low resolution version of the key frame provided by the video-segmentation into the videostream itself.
Auxiliary information as the length of the n^{th} shot and the key frame are hidden into the $(n-1)^{th}$ shot of the video sequence itself. Therefore, the information extracted from the $(n-1)^{th}$ shot at the decoder side refers to the next shot to be received. The receiver can use this information to conceal errors in the n^{th} shot.

2.1. Key frame embedding

The overall embedding procedure can be summarized as follows:

1. A low resolution version of the key frame $s(n)$ that summarizes the n^{th} shot (provided by the shot segmentation algorithm) is created by low pass filtering and down-sampling by a factor of two the full resolution frame.

   Then, the low resolution key frame is quantized so that, for each pixel, only the four most significant bits are retained. The corresponding four bit plane arrays of size $\frac{N_r}{2} \times \frac{N_c}{2}$ constitute our mark $s^{low}(n)$.

2. The 3D DWT $\mathcal{V}(n-1)$ of the $(n-1)^{th}$ shot $V(n-1)$ is computed.

3. The mark $s^{low}(n)$, extracted from the n^{th} shot, is embedded into $\mathcal{V}$(n-1) related to the (n-1)th shot (the host signal), using the QIM method [15]. More specifically, the mark is embedded twice: first the QIM is applied to the 4 most significant bits of $\mathcal{V}$(n-1), and, then, to the four less significant bits of $\mathcal{V}$(n-1).

   In practice, the embedding algorithm modifies the 3D-DWT coefficients by quantizing them. According to the value of the bit to be hidden, the quantizer chooses between one of two scalar uniform quantizers, shifted by Δ/2, where Δ is the quantization step of the embedding. By applying the QIM on $\mathcal{V}$(n-1), we obtain a matrix of watermarked coefficients $\mathcal{V}^M$(n-1).

4. The Inverse Discrete Wavelet Transform (3D-IDWT) is applied to $\mathcal{V}^M$(n-1), thus obtaining the watermarked shot V^M(n-1).

2.2. Key frame extraction

The procedure for extracting the hidden information can be summarized as follow:

1. The decoded videostream is first segmented into shots V^D(n). This operation can be performed by applying the same segmentation algorithm used in the embedding phase. As an alternative, as described before, the shot duration can be regarded as a side information that can be transferred by means of any auxiliary channel, including the covert channel. Thus, the duration of the n^{th} shot can be hidden into the 3D-DWT of the (n-1)th shot, together with the low resolution key frame $s^{low}(n)$. Let us observe that V^D(n) can differ from V^M(n) because of coding losses and transmission errors.

2. The length of the n^{th} shot and the low resolution key frame $s^{low}(n)$ are extracted from V^D(n-1) by means of the QIM detection algorithm [15]. The knowledge of the quantization step Δ is required.

3. The low dynamic, low resolution key frame is then upsampled in order to obtain the frame used for error concealment.

3. SIMULATION RESULTS

Many tests have been performed to verify the effectiveness of the proposed watermarking-concealment framework. To test the visibility of the hidden key frame and its robustness to H.264/AVC coding, a MATLAB simulator has been implemented.
For sake of simplicity, and for controlling the computational complexity, the videos have been segmented into shots of 8 frames, and the middle frame has been retained as key frame. Obviously, use of more sophisticated shot boundary detection and key frame construction could outperform this simple one. The reported results can thus be regarded as minimal performance achievable by the proposed error concealment scheme.
The results reported here refer to a QCIF video sequence (176×144 pixels) of 80 frames (10 shots) at an original bit/rate of 4.455 Kbit/s. To test the robustness of the data hiding scheme, the marked sequence has been coded with the H264 standard coder (JM software, ver. 13.2) at different bit/rates, namely: 256 Kb/s, 384 Kb/s, and 512 Kb/s.
The frames have been coded according the scheme *IPPPIPP-PIPPPI*, where *I* denotes an Intra frame and *P* corresponds to a Predicted frame. An example of the distortions introduced by the encoder in the reconstructed key-frame for both the 3D-DCT-based and 3D-DWT-based data hiding schemes is reported in Figures 1, 2, and 3. It can be easily verified that the proposed DWT method outperforms the previously presented DCT-based scheme.

(a) (b)

Fig. 1. Original Frame (a) and subsampled version (b).

(a) (b)

Fig. 2. Watermarked frame (a) DCT method, Delta = 40, extracted mark (b) coded at 256 Kb/s.

(a) (b)

Fig. 3. Watermarked frame (a) DWT method, Delta = 40, extracted mark (b) coded at 256 Kb/s.

To quantitatively assess the perceived quality of the marked video, three different objective quality metrics have been used: the Peak Signal to Noise Ratio (PSNR), the Weighted PSNR (WPSNR), and the Video Quality Metric (VQM) developed by the ITS/NTIA [16].

Δ	3D-DCT based	3D-DWT based
20	0.01	0.02
30	0.02	0.04
40	0.04	0.06
50	0.07	0.08
60	0.12	0.09
70	0.16	0.1
80	0.20	0.13
90	0.24	0.14
100	0.27	0.16
110	0.31	0.17
120	0.34	0.19
130	0.37	0.20

Table 1. VQM versus quantization step.

Let us recall that, given the original frame F and the frame F' whose quality has to be assessed, the WPSNR is defined as follows:

$$WPSNR(dB) = 10 \log_{10} \frac{\max(F)^2}{\|NVF(F' - F)\|^2}, \quad (1)$$

where the noise visibility function NVF is 1 in flat regions and 0 in textured regions and edges [17]. As expected, the plots of the average WPSNR show that increasing Δ degrades the video quality.

Figures 4, 5 and 6 show the PSNR and WPSNR versus the quantization step, computed between the embedded and the extracted watermark at 256 Kbit/s, 384 kbits/s, and 512 Kbit/s for both DWT and DCT based hiding schemes. Specifically:

- DCT Mark corresponds to the distortion between the embedded and the extracted mark by using the 3D-DCT based method.
- DWT Mark corresponds to the distortion between the embedded and the extracted mark by using the 3D-DWT based method.

In Table 1, the NTIA VQM for videos with the key frames embedded in the DWT and DCT domains are reported. Let us recall that VQM scores are in the range $[0, 1]$, where 0 indicates excellent quality. The table shows that with the same quantization step a higher quality of the sequence marked in the DWT domain is obtained. The simulations indicate that the H264 encoder quantization parameter (QP) is critical regarding the robustness of the embedding method. In particular, for the same coder output bit-rate, the quality of the reconstructed key frame strongly decreases if the QP is not constant. The performed simulations show that the minimum value of Δ for which the embedding is robust enough to the coding strictly depends on the compression rate, e.g. on the QP. Moreover, the QP should be constant during the H.264/AVC encoding.

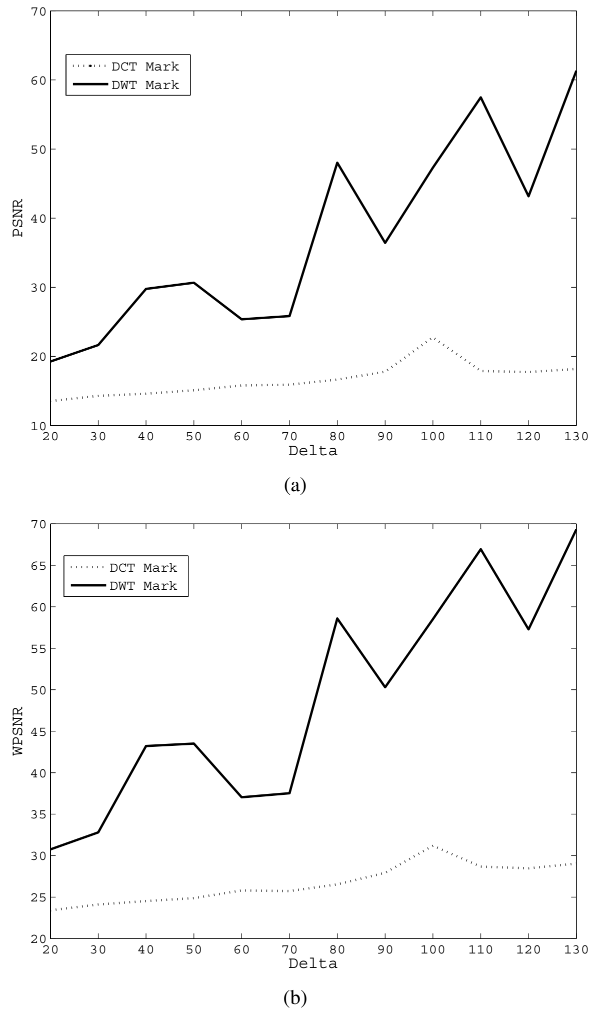


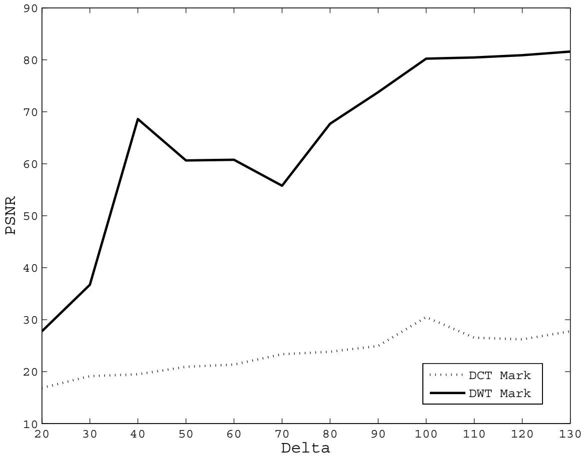
Fig. 4. PSNR (a) and WPSNR (b) versus the quantization step computed between the embedded and the extracted watermark, at 256 Kbit/s.

4. CONCLUSIONS

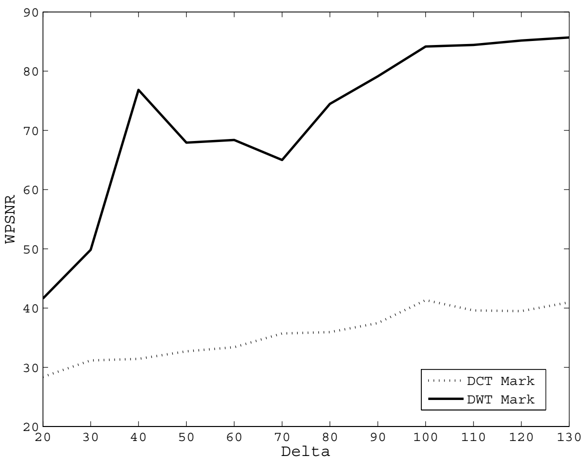
In this paper we have presented a video watermarking scheme for error concealment purpose over error-prone video transmission. The performed simulations show the effectiveness of the proposed method with noticeable improvements in the performances with respect previous works.

When using the QIM algorithm, if the quantization step Δ increases, the quantization error increases. Therefore the quality of the video decreases. On the other hand, increasing Δ increases the distance between the two scalar quantizers. Thus the robustness of the embedded mark increases. A trade-off between visibility and robustness of the embedding method has to be defined, setting the best range of values for the embedding parameters.

The DWT coefficients where the key frame is embedded can

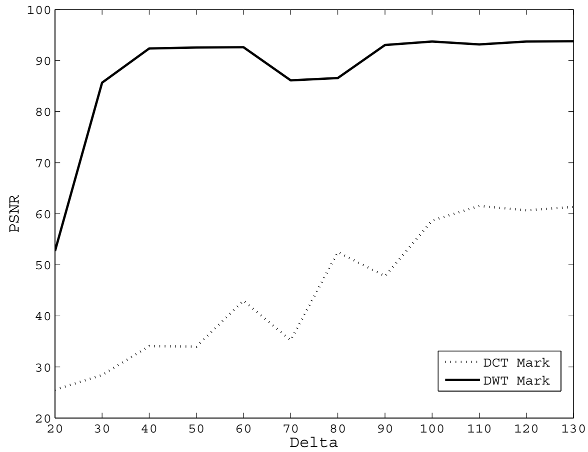


(a)

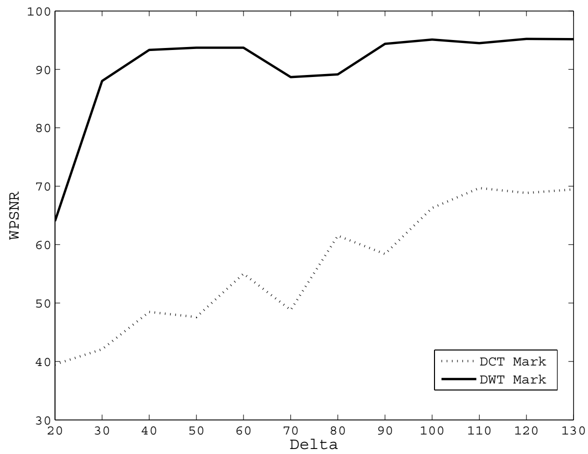


(b)

Fig. 5. PSNR (a) and WPSNR (b) versus the quantization step computed between the embedded and the extracted watermark, at 384 Kbit/s.



(a)



(b)

Fig. 6. PSNR (a) and WPSNR (b) versus the quantization step computed between the embedded and the extracted watermark, at 512 Kbit/s.

be considered as a parameter too. In the performed simulations, it was hidden into the 3D-DWT coefficients corresponding to the high frequencies. The left-high zone, corresponding to the low frequencies 3D-DWT coefficients is left untouched since any modification of them strongly affects the quality of the reconstructed frames.

5. REFERENCES

[1] M. Wu and B. Liu, "Data hiding in image and video: Part i-fundamental issues and solutions," *IEEE Transactions on Image Processing*, vol. 12, no. 6, pp. 685–695, June 2003.

[2] Y. Wang and Z. Qin-Fan, "Error control and concealment for video communication: a review," *Proceeding of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[3] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projection onto convex set," in *IEEE Transaction Image Processing*, 1995, vol. 4, pp. 470–477.

[4] P. Salama, N.B. Shroff, E.J. Coyle, and E.J. Delp, "Error concealment techniques for encoded video streams," in *Proc. International Conference Image Processing*, 1995, vol. I, pp. 9–12.

[5] J.W. Park, J.W. Kim, and S.U. Lee, "Dct coefficients recovery-based error concealment technique and its application to the mpeg-2 bit stream error," *IEEE Transaction Circuits Systems Video Technology*, vol. 7, pp. 845–854, Dec. 1997.

[6] W. Zeng and B. Liu, "Geometric-structure-based directional filtering for error concealment in image/video transmission," in *Proc. SPIE Conf. Wireless Data Transmission, Photonics East*, 1995, vol. 2601, p. 145156.

[7] X. Lee, Y. Zhang, and A. Leon-Garcia, "Information loss recovery for block-based image coding techniquesa fuzzy logic approach," *IEEE Trans. Image Processing*, vol. 4, pp. 259–273, Mar 1995.

[8] C.B. Adsumilli, M.C.Q. Farias, S.K. Mitra, and M. Carli, "A robust error concealment technique using data hiding for image and video transmission over lossy channels," *IEEE Transaction on Circuit and Systems for Video Technology*, vol. 15, no. 11, pp. 1394–1406, November 2005.

[9] C.B. Adsumilli, V. Razzauti, M. Carli, and A.Neri, "Videowatermarking in 3D DCT domain," in *Proc. IEEE International Workshop on Spectral Methods and Multirate Signal Processing, SMMSP 2005*, June 2005.

[10] T. Kalker, "Considerations on watermarking security," October 2001, pp. 201 –206.

[11] A. Yilmaz and A.A. Alatan, "Error concealment of video sequences by data hiding," in *International Conference Image Processing - ICIP*, 2003, vol. 3, pp. 679–682.

[12] M. Verza, N. Conci, G. Boato, and F.G.B. De Natale, "Multiple description coding using data hiding and regions of interest for broadcasting applications," *International Journal of Digital Multimedia Broadcasting*, vol. 2009, 2009.

[13] G. W. Wornell and B. Chen, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transaction on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[14] ITU-T, "Recommendation H.264, ISO/IEC 14496 10 version 1, Advanced video coding for generic audiovisual services," Tech. Rep., 2003.

[15] B. Chen and G.W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423 –1443, May 2001.

[16] VQM, "Video Quality Metric (VQM) software," Tech. Rep., Download http://www.its.bldrdoc.gov/n3/video/, v.2.0 edition.

[17] A. Watson, G. Jang, J. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transaction on Image Processing*, vol. 6, no. 8, August 1997.