ERROR CONCEALMENT IN VIDEO COMMUNICATIONS USING DPCM BIT STREAM EMBEDDING

Chowdary B. Adsumilli and Sanjit K. Mitra

Department of Electrical and Computer Engineering
University of California, Santa Barbara
Santa Barbara, CA - 93106-9560, USA
E-mail: (chowdary, mitra)@ece.ucsb.edu

ABSTRACT

An error concealment using data hiding (ECDH) technique is proposed to improve the end user perceptual quality of videos that are affected by transmission defects. A 2×2 set of coefficients from each block of block-based 2D DCT of the video frame is encoded using DPCM and embedded in the frame itself using a low gain spread spectrum watermarking technique. The DPCM bit stream is ordered into a binary block-image, which is approximately 4 times smaller than the video frame, and embedded in the frame's mid-frequencies. At the receiver, the extracted DPCM bits are lexicographically ordered and a reference is reconstructed for error concealing the channel loss affected video frame. The technique is closely compared with a similar approach which embeds halftoned bits of a low resolution version of the video frame in itself. Also, a comparison is done when the DPCM encoded bits are transmitted as side information through the channel rather than as embedded data. Experimental results show that the technique that uses DPCM bit stream embedding is more effective than the other two ECDH techniques as well as most other error concealment algorithms.

1. INTRODUCTION

Multiple error detection and concealment algorithms, which either stress on recovering the lost data in a more efficient way or removing/hiding the errors in an effective manner, have been proposed in the literature for video communications. Lately, new approaches to improve the recovery of lost data have been introduced which rely on data hiding. These techniques use data hiding to embed certain key features of the image/video frame in itself, which are extracted at the receiver and used as a reference to better recover the image/video frame from transmission losses and are called error concealment using data hiding (ECDH) techniques.

In this work, we consider a novel ECDH algorithm, where the embedded reference is the encoded energy content of the frame itself. A set of 4 2D DCT coefficients of each block are obtained in a 2×2 matrix format and encoded using DPCM. A spread spectrum watermarking algorithm is used to embed the DPCM bit stream in the mid-frequency range of the video frame. To reduce the encoder complexity and cater to the feasibility issues, embedding operation is performed only in the intra-coded frames of the video. The algorithm compares closely with the ECDH algorithm that embed the halftoned bits of the low resolution version of the video frame [1]. A detailed performance comparison of both techniques is provided here.

In this paper, the image or video frame that acts as a host to the embedded data is called as the parent image and the 2D marker data that is embedded is called as the child image. Furthermore, a two-part variant of the ECDH techniques is to transmit the child image data as side information through the channel rather than embedding it. We call this variant error concealment with side information (ECSI). The performance of ECDH and ECSI are compared. However, a detailed comparison of these techniques along the performance comparison of ECDH with other error concealment techniques can be found in [1].

The paper is organized as follows. Section 2 gives a detailed review of the work done in the field of ECDH. Section 3 explains the proposed technique and its variants. It also points out the advantages and disadvantages of one over the other. Section 4 details the obtained results and analyzes them, while Section 5 draws conclusions based on the analysis.

2. PREVIOUS WORK

The use of data hiding as an error control tool was first introduced by Liu and Li [2]. They extracted the important information in an image and embedded it into the host image with multi-dimensional lattice encoding. In the work that followed, certain key features were extracted from the image and these features were encoded and data hidden in the original image as a concealment tool [3],[4]. A region of interest based coded bit stream embedding in the region-of background wavelet coefficients was employed by Wang and Ji [5]. This technique gives better results when perception based encoding is employed.

Munadi *et al.* extended the concept of key feature extraction and embedding to inter-frame coding for video [6]. In their scheme, the most important feature is embedded into the prediction error of the current frame. Yilmaz and Alatan proposed embedding a combination of edge oriented information, block bit-length, and parity bits in intra-frames [7].

A common problem with existing ECDH techniques is that only one or a few selected set of key features are used for embedding. These features may not necessarily follow the loss characteristics of the channel employed. Moreover, recovery of high level information such as local texture may not be required for video

This work was supported in part by National Science Foundation grant No. CCR-0105404, in part by a University of California MICRO grant with matching supports from Philips Research Laboratories and Microsoft Corporation.

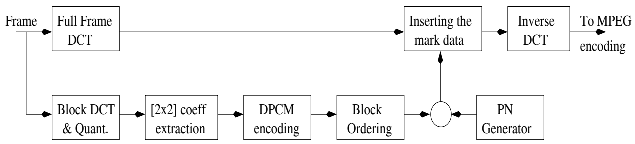


Fig. 1. Block diagram of the embedding algorithm.

(especially in the case of low bit rate video) transmission. Another recurring problem (except with [2]) is that these key features are not encoded before/while embedding. This may lead to reduction of quality of the reconstructed reference. The proposed algorithm is designed to avoid these two problems.

A set of concealment techniques that do not use data hiding while giving high levels of performance have been proposed in the literature [8],[9]. However, it has been shown that ECDH gives higher performance than these techniques and are therefore not considered in detail here [1].

3. ERROR CONCEALMENT USING DATA HIDING

Using data hiding techniques, redundancy is added to the transmitted video sequence frame data. It is important to underline that the proposed technique does not overload the communication channel by requiring feedback or any retransmission of damaged blocks.

3.1. DPCM encoded child image

The proposed scheme is divided into an embedding part and a retrieval part and each part is explained separately.

3.1.1. The Embedding Part

The data hiding technique used here is a modified version of the Cox's watermarking algorithm [10]. The block diagram of the embedding algorithm is shown in Fig. 1. In this technique, a block 2D DCT of a video frame is obtained and each block is quantized using the JPEG quantizer. A set of 4 coefficients for each block (one DC and three adjacent AC coefficients) is selected in a 2×2 matrix format. These coefficients are DPCM encoded and the encoded bit stream is ordered into a binary block, which forms the 2D marker, $\mathbf{m}_i$, for the i-th key frame. One marker is used for each intra-coded frame. An identifier is added at the end of DPCM code of each block for synchronization and effective decoder operation.

The set of coefficients selected for embedding is based on the encoder embedding capacity, compression quality factor, DPCM code length, and size of the marker required. In our case, by using a 2×2 set of coefficients, the marker is of size $\frac{m}{2} \times \frac{n}{2}$ for a video frame of size $m \times n$, i.e., the marker is $\frac{1}{4}$-th the size of the video frame.

A unique zero mean unit variance random pseudo-noise image, $\mathbf{p}_i$, of size $\frac{m}{2} \times \frac{n}{2}$ is generated with a known seed for each intra-coded frame of the video. For a generic i-th key frame $\mathbf{f}_i$, of a video sequence, the embedding watermark $\mathbf{w}_i$ is obtained by multiplying $\mathbf{m}_i$ with the pseudo-noise image $\mathbf{p}_i$.

The computed DCT coefficients of the luminance channel of the frame $\mathbf{f}_i$ are denoted as $\mathbf{F}_i$. The watermark $\mathbf{w}_i$ is then scaled by a factor α, and added to a set of these coefficients starting at the initial frequencies of (Δ_1, Δ_2). The resulting image $\mathbf{Y}_i$ is given

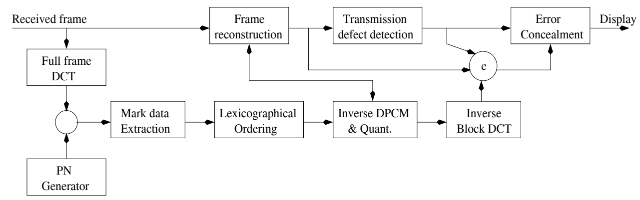


Fig. 2. Block diagram of the retrieval algorithm.

by

$$Y_i(K + \Delta_1, L + \Delta_2) = F_i(K + \Delta_1, L + \Delta_2) + \alpha.w_i(k, l) \quad (1)$$

where k and l correspond to the pixel location in the spatial domain, and K and L correspond to the coefficient location in the DCT domain. Here, $Y_i(.,.)$, $F_i(.,.)$, and $w_i(.,.)$ represent the individual component values of matrices $\mathbf{Y}_i$, $\mathbf{F}_i$, and $\mathbf{w}_i$, respectively. Note that $\Delta_1 \in [0, \frac{m}{2}]$ and $\Delta_2 \in [0, \frac{n}{2}]$. $\mathbf{Y}_i$ is then inverse transformed, encoded and transmitted.

In this work, the final watermark is added only to the mid-frequency DCT coefficients. For high quality video transmission, the mid-frequencies are a good choice. Inserting the watermark in the low-frequencies would cause visible artifacts in the image, while inserting it in the high frequencies would make it more prone to channel induced defects.

3.1.2. The Retrieval Part

The block diagram of the retrieval technique is shown in Fig. 2. The DCT coefficients of the luminance channel of the received frame $\mathbf{y}_{ri}$, denoted by $\mathbf{Y}_{ri}$, are computed as

$$\mathbf{Y}_{ri} = DCT_2(\mathbf{y}_{ri}) \quad (2)$$

where DCT_2 represents the 2-D DCT operation.

These coefficients are then multiplied by the corresponding pseudo-noise image $\mathbf{p}_i$. It is assumed that the receiver knows the seed for generating the pseudo-noise image and the initial frequencies, (Δ_1, Δ_2), where the mark was inserted. The binary marker is extracted by taking the sign of the result of the multiplication. This is given as

$$m_{ri}(k, l) = sgn\left\{Y_{ri}(K + \Delta_1, L + \Delta_2).p_i(k, l)\right\}. \quad (3)$$

where $Y_{ri}(.,.)$, $p_i(.,.)$, and $w_{ri}(.,.)$ are the individual component values of matrices $\mathbf{Y}_{ri}$, $\mathbf{p}_i$, and the extracted marker $\mathbf{m}_{ri}$, respectively. Note here that the values of $\mathbf{m}_{ri}$ greater than 0 are assigned a value of 1 and those that are equal to or less than 0 are assigned a value 0 to make the resulting image binary. Also note that the size of $\mathbf{m}_{ri}$ is $\frac{m}{2} \times \frac{n}{2}$.

The extracted data is lexicographically ordered into a single bit stream. Here, it is assumed that the receiver is aware of the identifier code at the end of DPCM code of each block. This assumption can be validated by using the eob (end of block) identifier in JPEG-like encoding. It can also be encoded, embedded, and transmitted along with the child image. Since the identifier bits are embedded (in full frame DCT of the parent image) before compression, this would not de-synchronize the existing eob of compression code. The reference image is reconstructed by obtaining the inverse DPCM values and using them as a 2×2 set of coefficients of each block. Other AC coefficients are assumed 0 for each block. These block coefficients are inverse quantized and inverse block DCT transformed to obtain the reference $\hat{\mathbf{f}}_i$, which is used for error concealing the lossy parent image.

3.2. Low resolution halftoned child image

The modified Cox's watermarking algorithm is used in this technique too. However, the 2D marker is chosen to be an approximation of the current frame. A second level 2D-DWT is performed on the video frame to obtain an image that is $\frac{1}{16}$-th the size of the original frame. A half-toned image, the marker $\mathbf{m}_i$, is then generated from the reduced size image using Floyd-Steinberg error diffusion algorithm. After the marker is generated, a repetition operation is performed in which each pixel of the marker is repeated 4 times in a 2×2 matrix format to increase its robustness. Note that the resulting image $\tilde{\mathbf{m}}_i$ is of size $\frac{m}{2} \times \frac{n}{2}$ even though $\mathbf{m}_i$ is $\frac{m}{4} \times \frac{n}{4}$.

The binary marker is embedded and extracted using Eqs. (1)-(3) with a small change to Eq. (3). Due to the repetition, the result of the multiplication needs to be averaged over 4 pixels before its sign is considered. Therefore, the size of $\mathbf{m}_{ri}$ is $\frac{m}{4} \times \frac{n}{4}$. After inverse halftoning, an estimate of the marker, $\hat{\mathbf{m}}_i$, undergoes inverse 2D DWT and zooming to form the reference $\hat{\mathbf{f}}_i$ which is compared with current received frame $\mathbf{y}_{ri}$ to conceal transmission errors.

A more detailed explanation of this algorithm can be found in [1].

3.3. Comparison of ECSI, ECDH with DPCM, and halftoning

For a fixed packet loss probability, the bits required to compress the original video frame are fewer than the bits required to compress the embedded frame. However, this difference is small and therefore will not compensate for the bits required to compress or encode the side information. This means that the transmission of halftone or DPCM bits as side information would require more bits than the difference in compression of mark-embedded and unmarked frames. Also, any additional computational complexity advantage that ECSI has over ECDH (since ECSI does not have to perform the embedding operation) is nullified by the complexity of encoding side information. Furthermore, since the side information bits are not as protected as the embedded bits, transmission loss incurs more errors in the received reference image in case of ECSI than ECDH. If higher protection is given to the side information bits such that the performance of ECSI is comparable to that of ECDH at higher packet loss probabilities, then the encoding complexity of ECSI increases. Therefore, ECDH provides more optimal point in the performance-complexity curve than ECSI. An extended analysis of these trade-offs along with PSNR vs. loss rate curves is given in [1].

The advantages of using ECDH with DPCM approach over ECDH with halftoning are that firstly, fewer number of bits are required to be embedded. As a comparison, the number of bits required to embed the child image after halftoning are of the order of 36000 for an image of size 240×320. However, this is reduced to around 4000 in the case of DPCM bit stream embedding. In fact, this reduction allows us to embed the DPCM bit stream of the parent image with its full resolution instead of using a low resolution version as a reference. Secondly, the complexity of the transceivers are greatly reduced. Not only are the DWT and halftone operations not required at the encoder, but also the zooming and/or the inverse DWT operations in the receiver are not used. Instead, JPEG-like operations such as quantization and DPCM encoding of DPCM coefficients are used thus reducing the complexity of the codec. Thirdly, the performance of the error concealment and the localized error concealment algorithms are much improved over ECDH with halftoning. This is because of two reasons: (1) the reference can be extracted to higher quality than the halftoned version, and (2) since fewer bits are used, α could be reduced and therefore the parent image would suffer from lesser watermarking artifacts. And lastly, since the DPCM bit stream of the reference image is embedded, it is more secure to outside attacks and transmission errors than just embedding halftone bits.

4. EXPERIMENTAL RESULTS AND ANALYSIS

The experiments were performed with a sample set of videos of a fixed size, 240×320. *ns-2* simulator was used to generate the wireless transmission loss scenarios. A Gilbert-Elliot loss model is used for generating the two-state Gaussian packet loss distribution with a predefined mean and variance. For simplicity of experimentation, it is assumed that: (1) single and 2-bit errors of successfully received packets were negligible, and if the number of bit errors exceed 2, the packet is lost, (2) no re-transmissions occur, (3) the loss probability is constant for a given channel bandwidth, and (4) packet size is fixed at 1500 kB for a given video transmission.

The embedding operation is performed with α value of 3. At the receiver, the packet loss areas are located and error concealed. In the explanation that follows, the frames that are error concealed in this manner will be referred to as EC frames. Since only a set of coefficients are embedded, the reconstructed reference would not have high frequency content. Therefore, the lossy areas in the video frame which have been error concealed would look smoother than the neighboring areas. For this reason, the high frequency content around a 16×16 pixel-area surrounding the loss are locally scaled to improve EC image's perceptual quality. The frames that are processed in this fashion will be referred to as LEC (local-scaled error concealed) frames.

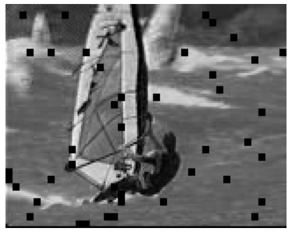
Fig. 3 shows the performance of the proposed algorithm and that of ECDH with halftoning. Figs. 3(a) and (b) show the original video frame and packet loss effected received frame respectively. Figs. 3(c) and (d) show the error concealed and localized error concealed frames obtained using ECDH with halftone bit stream embedding, and Figs. 3(e) and (f) show the error concealed and localized error concealed frames obtained using ECDH with DPCM bit stream embedding respectively.

Table 1. Algorithm performance in PSNR (dB) for ECDH with halftoning, ECDH with DPCM and ECSI.

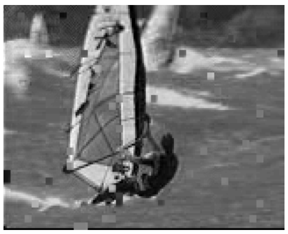
Video	HT_{EC}	HT_{LEC}	DP_{EC}	DP_{LEC}	ECSI
Foreman	25.83	29.52	27.31	30.77	29.13
News	26.42	29.89	27.95	31.06	28.68
Football	27.04	29.58	28.11	30.92	29.01
Flower	25.87	28.98	28.03	30.16	29.22
Hockey	25.35	29.44	27.39	31.27	29.08

Fig. 4 shows the PSNR vs. loss rate plots of the proposed algorithm along with its comparison to ECDH with halftone bits embedding. As observed from the figure, both the EC and the LEC versions of the proposed algorithm perform better than ECDH with halftone bits embedding. A performance comparison of ECDH

(a) Original image

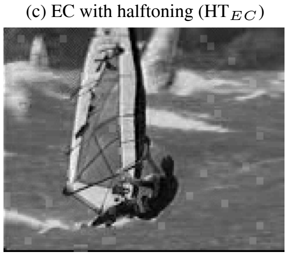


(b) Received image



(c) EC with halftoning (HT$_{EC}$)

(d) LEC with halftoning (HT$_{LEC}$)



(e) EC with DPCM (DP$_{EC}$)

(f) LEC with DPCM (DP$_{LEC}$)

Fig. 3. The received image is obtained for a mean packet loss probability of 0.15 and variance 2.5%. The PSNR values of the images are (b) 20.2943, (c) 27.3112, (d) 30.8620, (e) 28.8613, and (f) 33.1337. The value of α used in both cases was 3.

with DPCM bit stream embedding, ECDH with halftone bits embedding, and ECSI is provided in Table. 1 for various videos. From the figures and the table, we can conclude that ECDH with DPCM bit stream embedding (in the LEC case) outperforms other techniques.

5. CONCLUSIONS

A novel ECDH algorithm for video communications is developed where a 2×2 set of compact energy coefficients of a video frame are DPCM encoded and embedded into itself using spread spectrum watermarking. The performance of this algorithm is compared to ECDH with halftoning and ECSI transmission of DPCM bit stream.

A possible future direction would be to implement a temporal embedding algorithm where the motion vector information of the P- and B-frames can be embedded into the I-frame. This would represent a three-dimensional (spatial as well as temporal domain) embedding. The number of frames required for such 3D embedding and the number of bits that we can embed are yet to researched. Furthermore, spread spectrum embedding may not be ideal for certain low bit rate application due to its higher spreading gains, higher entropy levels, and sensitivity to attacks. A better alternative is quantization-based embedding. Ideally, an embedding algorithm suitable for this application needs to developed and could be another possible future direction.

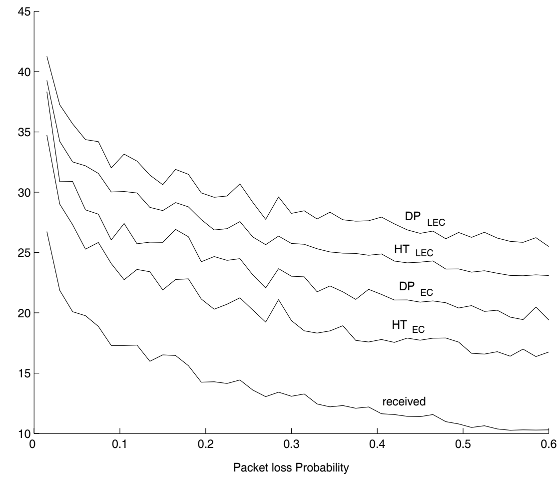


Fig. 4. PSNR vs. loss rate curves of ECDH with halftoning and ECDH with DPCM encoding.

6. REFERENCES

[1] C.B. Adsumilli, M.C. Farias, M. Carli, and S.K. Mitra, "A robust error concealment technique using data hiding for image and video transmission over lossy channels," *IEEE Trans. on Circuits and Systems for Video Technology*, accepted.

[2] Y. Liu and Y. Li, "Error concealment for digital images using data hiding," *Proc. DSP Workshop*, Texas, October 2000.

[3] Y. Shao, L. Zhang, G. Wu, and X. Lin, "Reconstruction of missing blocks in image transmission by using self-embedding," *Proc. Intl. Symp. on Intelligent Multimedia, Video and Speech Processing*, Hong Kong, May 2001, pp. 535-538.

[4] P. Yin, B. Liu, and H.H. Yu, "Error concealment using data hiding," *Proc. IEEE ICASSP*, vol. 3, May 2001, pp. 1453-1456.

[5] J. Wang and L. Ji, "A region and data hiding based error concealment scheme for images," *IEEE Trans. on Consumer Electronics*, vol. 47, no. 2, May 2001, pp. 257-262.

[6] K. Munadi, M. Kurosaki, and H. Kiya, "Error concealment using digital watermarking technique for interframe video coding," *Proc. Intl. Technical Conf. on Circuits/Systems, Computers, and Communications*, July 2002, pp. 599-602.

[7] A. Yilmaz and A.A. Alatan, "Error concealment of video sequences by data hiding," *Proc. ICIP*, vol. 3, September 2003, pp. 679-682.

[8] P. Salama, N.B. Shroff, E.J. Coyle, and E.J. Delp, "Error concealment techniques for encoded video streams," *Proc. ICIP*, vol. 1, October 1995, pp. 9-12.

[9] X. Li and M.T. Orchard, "Novel sequential error concealment techniques using orientation adaptive interpolation," *IEEE Trans. on Circuits and Systems on Video Technology*, vol. 12, no. 10, October 2002, pp. 857-864.

[10] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, vol. 6, no. 12, 1997, pp. 1673-1687.