# A HYBRID CONSTRAINED UNEQUAL ERROR PROTECTION AND DATA HIDING SCHEME FOR PACKET VIDEO TRANSMISSION

*Chowdary B. Adsumilli*[*], *Mylène C. Q. de Farias*[*], *Marco Carli*[†], *Sanjit K. Mitra*[*]

[*]Dept. of Electrical and Computer Engineering, University of California - Santa Barbara, USA
[†]Dept. of Electrical Engineering, University of ROME TRE - Rome, ITALY
Email: {chowdary,mylene,mitra}@ece.ucsb.edu, carli@ele.uniroma3.it

**ABSTRACT**

In this paper, a novel hybrid scheme with constrained Unequal Error Protection (UEP) and data hiding is proposed which maximizes the perceptual quality of the video at the end user to compensate for the effects of channel losses. The technique involves (1) implementing a *forcing function* which weighs the objective perceptual quality of the video frame based on a hidden mark signal to give optimum protection level in the packet, and (2) utilizes a data hiding mechanism to embed a second level wavelet approximation coefficients of the frame in itself. An optimum UEP in the transmitted packets is sought using a constrained optimization approach. Simulation results show that the proposed technique outperforms existing non-adaptive error concealment approaches in improving the perceptual quality, especially for higher loss probabilities.

## 1. INTRODUCTION

The internet has witnessed a rapid growth in deployment of Web-based multimedia applications during recent years. Applications that require a minimum quality level, as in the case of video, need the transmitter/receiver system to be enabled to perform end to end congestion control and quality adaptation to match the delivered stream. Error resilient transmission coding and error concealment decoding techniques have been introduced to overcome network losses and meet the quality requirements.

In this work, constrained optimization is implemented by dynamically varying the Unequal Error Protection (UEP) level in a packet based on network channel conditions. A "forcing function" [1] is estimated based on the channel loss characteristics which provide the bandwidth and the latency constraints. The UEP level in the packet is modified in accordance with this forcing function and the video data such that the visual quality is maximized at the end user. An error concealment technique using data hiding [2] is implemented in a novel way to transmit a lower resolution version of the video in itself without increasing its bit rate. The UEP technique is adapted to data hiding such that high protection is given to the data embedded frequencies.

Similar approaches using adaptive UEP have been evaluated in [3] and [4]. While Mohr *et al.* use unconstrained optimization, a constraint on system probability of failure rather than on channel conditions is applied in [3]. In doing so, Grangetto *et al.* have enforced a tight bound on minimum achievable Peak Signal to Noise Ratio (PSNR) but have not prevented any channel inflicted quality loss. It is argued here that channel constraints have to be implemented adaptively and the packet structure changed dynamically for the model suggested in [3] to work effectively. These approaches, however, do not adapt to any error concealment techniques.

The remainder of the paper is organized as follows. Section 2 describes in detail the technique of unequal error protection that is proposed here as a constrained optimization problem. Section 3 explains the error concealment reconstruction algorithm. The experiments done using the proposed technique and the results obtained are discussed in Section 4. In Section 5, conclusions are drawn based on the obtained results.

## 2. CONSTRAINED UNEQUAL ERROR PROTECTION

The proposed technique incorporates constrained optimization of the UEP level in the packet structure to achieve maximum expected subjective perceptual quality. Objective measures that accurately define subjective visual quality include SNR measurement, frequency domain masking/pooling, and detection thresholds [5]. In this paper, however, *PSNR* is adopted. The proposed optimization can be defined as finding an optimum UEP level that maximizes the end user PSNR first as an unconstrained optimization with a *forcing function* and then constraining it using variable bandwidth and latency constraints.

The optimization of UEP is done based on a *forcing function*, which can be defined as the function that determines the error protection level to be employed in a packet based on the embedded mark signal for the given channel conditions, i.e., for a given combination of effective bandwidth and loss characteristics, the forcing function evaluates the required error protection bits in the packet to be transmitted by giving higher protection to the mark embedded data. It therefore indirectly "forces" a UEP level for the packet transmission.

Let this forcing function be represented by $f(\mathbf{A}_i)$ where $\mathbf{A}_i$ is a column vector representing the $i^{th}$ packet sent. Mathematically, the unconstrained problem can be defined as follows: Let $\mu$ be the UEP ratio (protection level) in the packet. Then, for a given $f(\mathbf{A}_i)$, we need to find the value of $\mu$ that satisfies

$$\max_{i\in[0,N]}\{PSNR/f(\mathbf{A}_i) \quad \forall \mathbf{A}_i\}, \tag{1}$$

where *N* is the total number of packets. For simplicity, we assume $f(\mathbf{A}_i)$ to be the performance curve that includes the maximum area on the probability of successful arrival of the mark embedded packet.

This work was supported by University of California MICRO grant with matching support from Conexant Corp., Lucent Technologies, Philips Laboratories, and Microsoft Corp.

The system performance of a standard network protocol (UDP here) is observed for varying bandwidths over time for the consideration of constrained optimization problem. Once the transmission rate ($R$) matches the bandwidth ($B$) within the Acceptable Range ($AR$), the rate is either fixed or lowered. Similarly, bounds on latency are also applied. For this, the upper bound of latency is considered to be the time difference ($T$) between two *I* frames in the video transmission. Let $t_1$ be the time required to transmit the entire frame with UEP ratio of $\mu$. It is given by

$$t_1 = \frac{N.S.(1-\mu)}{B+AR}, \tag{2}$$

where *S* = Packet size; *N* = total number of packets; *B* = bandwidth, and $\mu \in [0, 1]$. The latency constraint is then given by

$$|t_1| < t_0 + T, \tag{3}$$

where $t_0$ is the time required to transmit the entire frame without any delay bounds.

The bandwidth constraint indirectly renders limiting of the total number of packets transmitted for a given $t_1$ in Eq. (2). Hence, even though a total number of *N* packets are required to transmit the entire video frame, a total of say $\tau$ packets can only be transmitted by a reduced bandwidth in the given time constraint of Eq. (3). This is particularly true for mark signal embedded packet transmissions where fewer source bits are sent in each packet due to higher protection. Hence, the bandwidth constraint imposes indirectly a reduction threshold on the total number of packets given as

$$\tau = N.\frac{\mu_1}{\mu_\tau}, \tag{4}$$

where $\mu_1$ is the UEP ratio that satisfies $t_1$ in Eq. (2) and $\mu_\tau$ is the UEP ratio for the current packet transmission. The variation in bandwidth can have considerable effect on the quality of the video transmitted. The end user subjective visual quality acceptance variation gives us the acceptable range ($AR$) of bandwidths to achieve a *constant* perceptual quality. This implies that the rate of transmission should adapt to bandwidth variations with an allowable range of $AR$, which gives us the bandwidth constraint as

$$|R - B| \leq AR. \tag{5}$$

Hence, the constrained optimization problem can now be formulated as: Find a $\mu$ that satisfies

$$\begin{aligned} \max_{i\in[0,k]} \quad & \{PSNR/f(\mathbf{A}_i) \quad \forall \mathbf{A}_i\}, \\ s.t: \quad & \\ & |t_1| < t_0 + T, \\ & |R - B| \leq AR, \end{aligned} \tag{6}$$

where $k = min\{N, \tau\}$. As the amount of data hidden in a packet increases, so does the packet protection, thus decreasing $\tau$.

Equation (6) also requires that the PSNR be maximized based on the weighted protection provided by the forcing function. As described, the forcing function $f(\mathbf{A}_i)$ considered here is the function that maximizes the area under the curve of perceptual quality with variation in probability of successful transmission of the packet $\mathbf{A}_i$. For a Rayleigh fading channel, the maximum area under the quality-bit rate curve would be its expectation. $f(\mathbf{A}_i)$ is therefore assumed to be

$$\begin{aligned} f(\mathbf{A}_i) &= E\{PSNR\}\forall i \in [0, k] \\ &= \sum_{i=1}^{k} PSNR * (1 - L_i), \end{aligned} \tag{7}$$

where $L_i$ is the loss probability for the packet $\mathbf{A}_i$. $i$ varies till $k$ packets and as the data hiding in the frame increases, $\tau$ decreases making $k = \tau$, thus increasing the effect of the forcing function by making the latency constraint much harder.

To account for the channel effects, consider $\mathbf{A}_i$ to be a column vector of size $S \times 1$. Let $\mathbf{D}$ be the frame vector formed by lexicographic ordering of all the $\mathbf{A}_i$s. Therefore, $\mathbf{D}$ is given by

$$\mathbf{D} = [\mathbf{A}_1^T \mathbf{A}_2^T ... \mathbf{A}_k^T], \tag{8}$$

where $k$ is as defined as in Eq. (6). The size of $\mathbf{D}$ is $1 \times Sk$ bits or $1 \times k$ packets. $\mathbf{C}$, the channel output, can then be given as

$$\mathbf{C} = diag(\mathbf{P}^T \mathbf{D}), \tag{9}$$

where $\mathbf{P}$ is the binary probability loss vector of the channel with a predefined loss percentage. $\mathbf{P}$ is a row vector of size $1 \times k$ and is randomly generated. The simulation details of $\mathbf{P}$ are given in Section 4. Since each element of $\mathbf{P}$ is multiplied with each $\mathbf{A}_i$ in $\mathbf{D}$, $\mathbf{C}$ is a vector of $k$ packets and now contains the received set of packets which are decoded and the image is reconstructed.

## 3. ERROR CONCEALMENT WITH DATA HIDING

Using data hiding techniques, redundancy is added to the transmitted video sequence frame data without increasing its bit rate. The basic approach consists of independently embedding the original information from the video frames into the data stream as hidden data. At the receiver, this hidden data is extracted which provides additional information about the received frame and so can be used for detecting and concealing errors.

### 3.1. The Embedding Part

The data hiding technique used here is a modified version of the Cox's watermarking algorithm [6]. The block diagram of the embedding algorithm is shown in Fig. 1. In this technique, a 2-D image (marker) is embedded into the discrete cosine transform (DCT) coefficients of each frame of the video. A spread spectrum technique is then used to hide the marker, by multiplying it with a pseudo-random noise before embedding it into the video. This marker, which is an approximation of the current frame, has the purpose of adding the redundancy to the video in a way to make it possible to recover any lost data.

Due to the limited embedding capacity of the algorithm, it is practically not feasible to embed the whole frame into itself. In this work, therefore, the discrete wavelet transform (DWT) and dithering techniques have been used to reduce the amount of data to be embedded such that the algorithm embeds maximum information while still catering to the feasibility issues. The dithering techniques used here is Floyd-Steinberg error diffusion method [7]. The approximation coefficients are half-toned before being embedded.

The 2-D DWT of the frame is first computed. A two-level DWT is performed again on the approximation coefficients such that the image obtained, the marker, is one-fourth the size of the

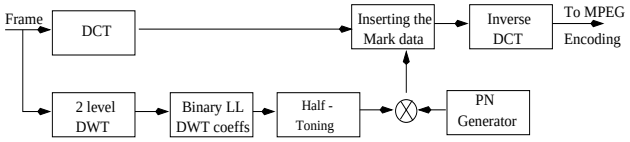


**Fig. 1**. Block diagram of the embedding alogorithm

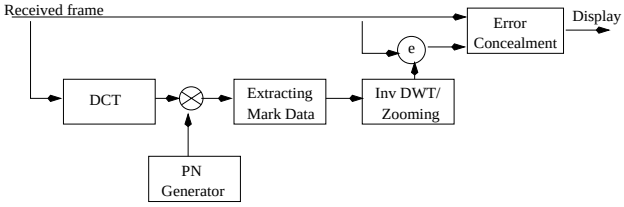


**Fig. 2**. Block diagram of the retrieval algorithm

original frame. A dithered image is then generated from the reduced size image. One marker is used for each frame. Each pixel of the marker is repeated 3 times with the second level high-high (HH) diagonal coefficients in a $2 \times 2$ matrix format. This repetition allows the decoder to recover the mark from the data in a more robust fashion.

Mathematically, the marker generated for the $i$-th frame, $\mathbf{f}_i$, can be represented as $\mathbf{m}_i$. Here, $\mathbf{f}_i$ is of size $m \times n$ and $\mathbf{m}_i$ is $\frac{m}{4} \times \frac{n}{4}$. Let $T$ represent the transformation of error diffusion and $\mathbf{w}_i$ be the resulting watermark. A zero mean, unit variance pseudo-noise image is then randomly generated with a known seed. A unique pseudo-noise image is generated for each frame of the video. For a generic $i$-th frame, $\mathbf{f}_i$ of a video sequence, the final watermark $\tilde{\mathbf{w}}_i$ is obtained by multiplying the mark image $\mathbf{w}_i$ with the pseudo-noise image, $\mathbf{p}_i$:

$$\tilde{\mathbf{w}}_i = \mathbf{w}_i . \mathbf{p}_i = T(\mathbf{m}_i) . \mathbf{p}_i . \tag{10}$$

An important note here is that the pseudo noise matrix is of size $m \times n$. Four mark images, three approximation and one diagonal coefficient matrices, each of size $\frac{m}{4} \times \frac{n}{4}$ are multiplied with $\mathbf{p}_i$ on a pixel-by-pixel basis. The DCT coefficients of the luminance channel of the frame $\mathbf{f}_i$ are computed. The watermark $\tilde{\mathbf{w}}_i$ is then scaled by a factor $\alpha$, and added to a set of these coefficients. The resulting data, $\mathbf{Y}_{i,l}$ is given by

$$\mathbf{Y}_{i,l} = DCT(\mathbf{f}_i)_l + \alpha . \tilde{\mathbf{w}}_{i,l} \tag{11}$$

where $l$ varies from 1 to 4. Here, $l$ is the number of times $\tilde{\mathbf{w}}_i$ is embedded into various frequencies of $\mathbf{f}_i$.

The mark is added only to the mid-frequencies DCT coefficients. The range of frequencies where the watermark is inserted is strongly dependent on the application. For the purpose of delivering a high quality video through a lossy channel, the mid-frequencies are a good choice. Inserting the mark in the low-frequencies would cause visible artifacts in the video, while inserting it in the high frequencies would make it more prone to channel errors.

### 3.2. The Retrieval Part

The block diagram of the retrieval technique is shown in Fig. 2. The DCT coefficients of the luminance channel are computed. The coefficients where the mark was inserted are extracted and are multiplied by the corresponding pseudo-noise image $\mathbf{p}_i$ as shown in Eq. (12). The pseudo-noise image generated is same as that at the transmitter side.

$$\mathbf{G}^{'}{}_{i,l} = DCT(\mathbf{f}^{'}{}_i)_l \tag{12}$$

Here, it is inherently assumed that the receiver knows the seed for generating the pseudo-noise image. An issue of concern with this assumption is that it might lead to possible synchronization problems when severe channel errors cause loss of frames. This can in turn be handled by embedding the frame order number $i$ into the frame itself. The receiver side pseudo noise generator algorithm can be driven by the recovered value.

It is also assumed that the receiver knows the initial position in the DCT domain coefficients where the mark was inserted. The result is averaged for the 4 pixels ($2 \times 2$ matrix form) and the binary mark is extracted by taking the sign of this average as shown in Eq. (13).

$$\tilde{\mathbf{w}}_{ri} = sgn\left\{ \frac{1}{16} \sum_{l=1}^{4} \mathbf{G}^{'}{}_{i,l} . \mathbf{p}_i \right\} . \tag{13}$$

The repetition of the mark allows for an increase in the robustness of the designed system to transmission error. The marker will be degraded due to the DCT transformation, the pseudo-noise image, and to all the losses the video usually suffers, such as compression, channel noise, etc. The error in the watermark due to these errors can be estimated by:

$$\mathbf{e}_i = \tilde{\mathbf{w}}_i - \tilde{\mathbf{w}}_{ri} \tag{14}$$

It has been shown by Campisi *et al.* [8] that this approach enables a fairly large amount of hidden data to be embedded without significantly affecting the perceptual quality of the encoded image. Once the binary marker is extracted, a 2-D inverse DWT is performed on it along with the high-high diagonal coefficients. This is then zoomed by up-sampling and passing through a low pass filter to obtain an $m \times n$ image. The resulting image is compared with the current frame to detect and conceal the corrupted blocks by substituting the appropriate data.

## 4. SIMULATION RESULTS

The algorithm proposed in Sections 2 and 3 has been implemented using conventional UDP transmission with a simulated packet loss. The following assumptions are made for simplicity with regard to implementation of the algorithm: (1) the binary loss probability of the channel is assumed to be constant for a given network bandwidth, (2) the source transmission rate is assumed to be less than the maximum channel bandwidth, (3) no re-transmissions occur, and (4) bit errors over successfully received packets are negligible.

Haar wavelet is used for calculating the two-level approximate and diagonal wavelet coefficients. The value of $\alpha$, the scaling factor for embedding, was fixed at $0.6$. The embedded frame is then source coded and channel protected with the defined UEP scheme. The compressed output stream is vectorized and multiplied with a vector $P$, that is randomly generated using Monte Carlo simulation. About $1000$ varying packet loss simulations were generated

**Table 1**. Performance of the proposed algorithm. PSNR (or P) in dB for a fixed mean loss 15% and variance 2.5%; $\alpha$=0.6.

| | $\mu$ = 0.2 | | $\mu$ = 0.3 | | $\mu$ = 0.4 | |
|---|---|---|---|---|---|---|
| *Frame* | $P_r$ | $P_{ec}$ | $P_r$ | $P_{ec}$ | $P_r$ | $P_{ec}$ |
| Wind | 25.61 | 36.74 | 27.98 | 38.11 | 26.92 | 37.71 |
| Camera | 23.60 | 33.78 | 24.82 | 34.49 | 24.48 | 33.87 |
| Psycho | 21.68 | 36.55 | 23.70 | 37.35 | 22.58 | 36.86 |
| News | 26.99 | 35.03 | 27.51 | 36.01 | 27.34 | 35.46 |
| Surf | 24.33 | 35.07 | 26.19 | 36.49 | 25.21 | 35.75 |
| Dog | 22.07 | 35.19 | 23.78 | 36.32 | 22.87 | 35.63 |

independently for each transmission and their statistical average is taken to obtain the probability loss vector.

Table. 1 summarizes the results of the experiment. For each value of $\mu$, the PSNR of the received image ($PSNR_r$) and the PSNR of the error concealed image ($PSNR_{ec}$) were noted. Six images/video sequence frames were considered. The experiment was repeated for $\mu = 0.2, 0.3$ and $0.4$ for all six frames. In each case, the UEP was varied and chosen such that the total number of packets remain well within the constraints defined in Section 2.

A sample result is shown in Fig. 3 for the Cameraman image with the parameter values: $\mu$ = 0.3, $\alpha$ = 0.6, mean loss = 15%, loss variance = 2.5%. The received frame had a $PSNR_r$ = 24.82 and the error concealed image had a $PSNR_{ec}$ = 34.49. These frames are shown in Figs. 3(a), (b) and (c) respectively. Fig. 3(d) was obtained by localized scaling error concealment. Here, the scaling of the watermark image was done very locally relative to the area of the packet loss by using a localization kernel. The kernel used here was of the size of the lost data area and the $PSNR_{lc}$ obtained was 35.48. Even though the PSNR variation is not substantial when compared to the error concealed frame, there is much improvement in the perceptual quality. However, better results in terms of PSNR are expected if the kernel size is varied.

**Fig. 3**. (a) Original frame, (b) Received frame with mean loss = 15%, variance = 2.5%; $\mu$ = 0.3; PSNR = 24.8176, (c) Error concealed frame; $\alpha$ = 0.6; PSNR = 34.4933, and (d) Localized scaling error concealed frame; PSNR = 35.4776.

## 5. CONCLUSION

A constrained UEP technique combined with data hiding is proposed for lossy video communication that aims at achieving the maximum perceptual quality at the end user. The data hiding technique employed is efficient in protecting the low-low wavelet coefficients of the frame by embedding multiple copies of these coefficients in the frame itself. A UEP technique that maximizes the PSNR with a constraint on a channel conditions is formulated that ensures more protection be given to the embedded data hidden packets.

Simulation results of the proposed algorithm on various video sequence frames are presented and analyzed for varying UEP ratios, and scaling factors. A localized scaling error concealment technique is implemented for improving the perceptual quality of the frames. It can be seen from these results that the proposed algorithm outperforms other existing techniques for video transmission over lossy channels.

## 6. REFERENCES


[1] C.B. Adsumilli, Y.H. Hu, "A dynamically adaptive constrained unequal error protection scheme for video transmission over wireless channels," *Proc. Intl. Workshop on Multimedia Signal Processing Conf.*, in press, USA, Dec 2002.

[2] M. Carli, D. Bailey, M. Farias, S.K. Mitra, "Error control and concealment for video transmission using data hiding," *Proc. Wireless Personal Multimedia Comm.*, in press, USA, Oct 2002.

[3] M. Grangetto, E. Magli, M. Marzo, G. Olmo, "Guaranteeing Quality of Service for Image Transmission by means of Hybrid Loss Protection," *Proc. Intl. Conf. on Multimedia and expo*, Vol. 2, pp. 469-472, Switzerland, Aug 2002.

[4] A.E. Mohr, E.A. Riskin, R.E. Ladner, "Unequal Loss Protection: Graceful Degradation over Packet Erasure Channels Through Forward Error Correction," *IEEE Journal on Selected Areas in Comm.*, Vol. 18, No. 7, pp. 819-828, June 2000.

[5] M. Farias, M. Carli, J.M. Foley, S.K. Mitra, "Detectability and annoyance of artifacts in watermarked digital videos," *Proc. XI European Signal Processing Conf.*, France, Sept 2002.

[6] I. Cox, J. Kilian, F. Leighton, T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. on Image Processing*, Vol. 6, No. 12, pp. 1673-1687, Dec 1997.

[7] J. Buchanan, L. Streit, "Threshold - diffuse hybrid half - toning methods," *Proc. Western Computer Graphics Symposium (SKIGRAPH)*, pp. 79-90, Canada, 1997.

[8] P. Campisi, M. Carli, G. Giunta, A. Neri, "Tracing Watermarking for Multimedia Communication Quality Assessment," *Proc. IEEE Intl. Conf. on Comm.*, Vol. 2, pp. 1154-1158, USA, May 2002.