Adaptive MPEG-2 Video Data Hiding Scheme

Anindya Sarkar, Upamanyu Madhow, Shivkumar Chandrasekaran and
Bangalore. S. Manjunath

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

ABSTRACT

We have investigated adaptive mechanisms for high-volume transform-domain data hiding in MPEG-2 video which can be tuned to sustain varying levels of compression attacks. The data is hidden in the uncompressed domain by scalar quantization index modulation (QIM) on a selected set of low-frequency discrete cosine transform (DCT) coefficients. We propose an adaptive hiding scheme where the embedding rate is varied according to the type of frame and the reference quantization parameter (decided according to MPEG-2 rate control scheme) for that frame. For a 1.5 Mbps video and a frame-rate of 25 frames/sec, we are able to embed almost 7500 bits/sec. Also, the adaptive scheme hides 20% more data and incurs significantly less frame errors (frames for which the embedded data is not fully recovered) than the non-adaptive scheme. Our embedding scheme incurs insertions and deletions at the decoder which may cause de-synchronization and decoding failure. This problem is solved by the use of powerful turbo-like codes and erasures at the encoder. The channel capacity estimate gives an idea of the minimum code redundancy factor required for reliable decoding of hidden data transmitted through the channel. To that end, we have modeled the MPEG-2 video channel using the transition probability matrices given by the data hiding procedure, using which we compute the (hiding scheme dependent) channel capacity.

Keywords: MPEG-2 compression, Quantization Index Modulation (QIM), quantization parameter per macroblock, channel capacity estimate

1. INTRODUCTION

Video data hiding presents a more challenging task compared to image data hiding. Digital video is a very promising host candidate that can carry a large amount of data (payload) and its potential for secret communications is largely unexplored. Since a video is formed from a sequence of frames, it presents the data hider with the possibility to embed and send a large amount of data. Digital video is compressed - both spatially using transforms such as the discrete cosine transform (DCT) and temporally using motion compensation, before being transmitted. Now, after the host video, with embedded data, has been subjected to compression, it may often be that the distortion for the transform domain coefficients is large enough so as to make the retrieval of the hidden data impossible. The perturbation of the coefficients depends on the severity of the compression. In image coding, we modulate the coefficients such that their perturbation is within a pre-specified bound and the amount of perturbation can be predicted from the compression in the spatial domain. In video, due to the exploitation of the spatial and temporal redundancies, there is enhanced compression and this makes the coefficient perturbation difficult to predict.

The bandwidth of the transmission medium determines the allowable bitrate for the video. It is a challenge to design a robust data hiding system where the hiding parameters scale properly with reduced bitrates and one has to trade-off (a significant amount of) data embedding capacity to obtain robustness at higher compression rates. E.g. to make an image data hiding system robust to JPEG[1] compression, we need to hide at a lower quality factor, than the JPEG compression attack. For a video compression format, say MPEG-2,[2] the number of variable parameters are many more (e.g. variation in bitrate, Group of Pictures size, and so on).

Further author information: (Send correspondence to Anindya Sarkar)
Anindya Sarkar: E-mail: anindya@ece.ucsb.edu, Telephone: 1-805-893-5682

For MPEG-2 compression, we wish to make the video data-hiding scheme robust to variation in these parameters.

In our prior work,[3] a “Selective Embedding” based robust scheme was proposed for high-volume data hiding in still images. We have studied the feasibility of extending this scheme for video data hiding (considering MPEG-2 compressed video). We have proposed an adaptive hiding scheme, which considers the video as a sequence of frames, where the data embedding rate is varied depending upon the frame quality and the type of frame (I/P/B in MPEG-2 scheme). Due to the increased perturbation/distortion (because of compression) for videos, greater redundancy needs to be inserted in the error correcting code used in the hiding scheme for videos.

2. PAST WORK - SELECTIVE EMBEDDING IN COEFFICIENTS SCHEME

In our scheme, the input video is decompressed at first into a sequence of frames, as shown in Fig. 1 which provides a detailed block diagram representation of the video data hiding method. We embed data in the luminance (Y) component only. Here, we have employed the “Selective Embedding in Coefficients” (SEC) scheme for hiding.[3] In the SEC scheme, 8×8 DCT of non-overlapping blocks are taken and the coefficients are divided by the JPEG quantization matrix at design quality factor. A uniform quantizer of step size Δ is used on the DCT domain coefficients of the host image. Data is embedded through the choice of the scalar quantizer. The quantization index modulation (QIM)[4] scheme uses even and odd multiples of Δ to store 0 and 1, respectively. Only those quantized DCT coefficients which lie in a certain low frequency band and whose magnitude exceeds a certain threshold are used for hiding - hence “selective embedding” occurs. For a coefficient below the threshold, an erasure occurs. The embedded bitstream is then encoded and transmitted.

The decoder does not know explicitly the exact locations where the data is hidden but it uses the same local criteria (low frequency band for embedding and use of a magnitude threshold) as the encoder to guess the locations of hidden data. Channel noise may cause an insertion (decoder guessing incorrectly that there is hidden data) or a deletion (decoder guessing incorrectly that there is no data) which may lead to de-synchronization and decoding failure. We use turbo-like codes with strong error-correction capability and channel erasures at the encoder[3] to account for this problem. Using the entire set of coefficients that lie in a designated low-frequency band, long codewords can be constructed to achieve very good correction ability. We use repeat-accumulate (RA) codes[5] in our experiments because of their simplicity and near-capacity performance for the erasure channels. A rate $\frac{1}{q}$ RA encoder involves q-fold repetition, pseudorandom interleaving and accumulation of the resultant bit-stream.

Thus, once the data is embedded using the SEC scheme, error resilience is added using the RA codes. The modified frames are then converted to a MPEG-2 video. The video is subjected to MPEG-2 compression attacks (change in bitrate, variation in the Group of Pictures (GOP) size). Then, the received video is decoded to a sequence of frames, from which decoding (of the embedded data per frame) is performed iteratively using the sum-product algorithm.[6]

3. ADAPTIVENESS OF THE HIDING SCHEME FOR MPEG-2 COMPRESSION STANDARD

While hiding data in a frame, we can vary these 4 parameters:

- the first (n) number of AC DCT coefficients, chosen by zigzag scan, to embed data per 8×8 block of the Y component of a raw frame
- the RA-code redundancy factor (q)
- the quantization interval used in the QIM scheme (Δ as in Fig. 2)
- the design quality factor (QF) used for quantization of the DCT coefficients.

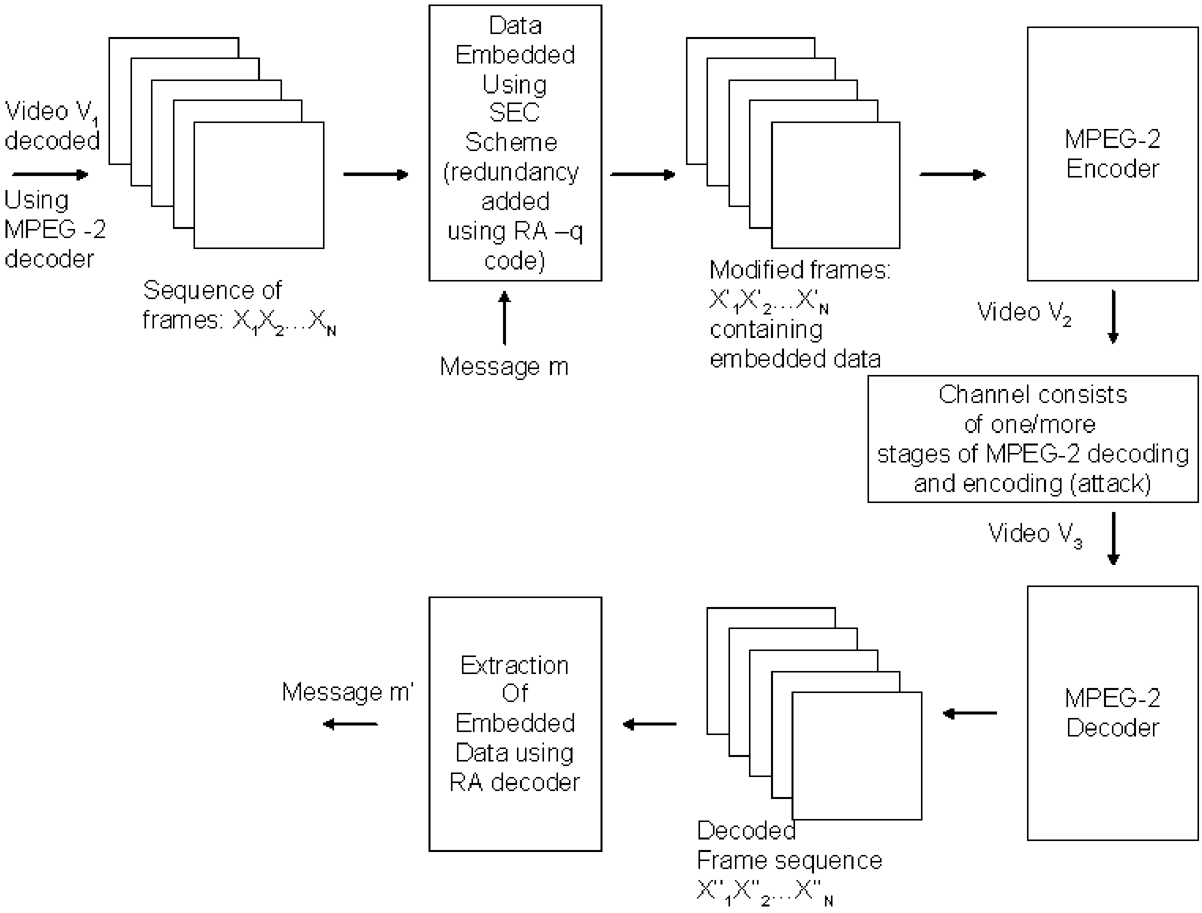

Figure 1. The video data hiding framework: data is hidden in the raw frames in the transform domain

The effective number of data bits hidden per 8×8 block = $\frac{n}{q}$.

Let us now explain the different levels of adaptiveness used in our scheme:

1. The first level of adaptability in the scheme is hiding different amounts in I, P and B frames. Due to bidirectional prediction, B frames are represented using the minimum bits while due to the intra-coding and absence of motion vector representation, I frames require the maximum bits. Naturally, the distortion incurred is minimum for I and maximum for B frames. Therefore, we can afford to hide more in I frames and recover it after compression attacks, compared to B frames.
2. While deciding on the parameter which will control the amount of data we can hide in a frame, we have borrowed from the adaptive rate control scheme of MPEG-2. From[2,7] , we can say that the reference quantization parameter derived per macroblock is related to the video bitrate, the frame-rate and the spatial activity in the macroblock. A smaller quantization parameter indicates greater spatial activity and finer quantization; computing the average (averaged over the macroblocks in a frame) quantization parameter (we shall call it *mquant* according to the notation in[7]) per frame gives us an estimate of the number of bits used by MPEG-2 encoder to represent the frame .
3. The relationship of the data hiding capability with the spatial activity in a macroblock can be explained as follows: for the hiding scheme (SEC) to be robust to noise, we hide only in those quantized DCT coefficients

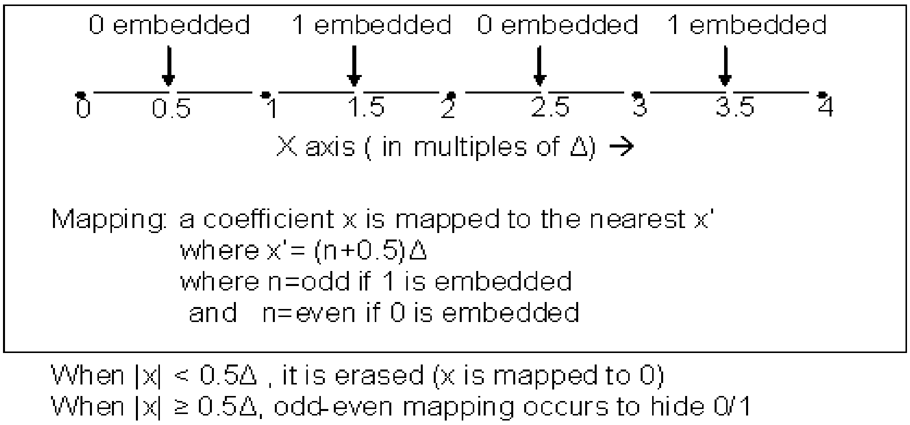

Figure 2. The scalar version of QIM scheme used in data hiding

that exceed a certain threshold. For a very smooth region in the image, most of the AC DCT terms will be close to 0, leaving fewer terms eligible for hiding. In blocks of higher variance, the magnitude of DCT coefficients will be more, in general, resulting in more available locations for data embedding (these frames will have a lower value of *mquant*). Thus, we hide more in frames having lower *mquant* and vice versa .

3.1. Explanation of MPEG-2 based Quantization Parameter

We need to explain some MPEG-2 based terminologies which decide the quantization parameters for each macroblock. MPEG-2 uses an adaptive rate control scheme where depending on the type of frame and target bits allocated for that frame, the reference quantization parameter Q_j (for the j^{th} macroblock) is varied. From MPEG-2 standard terminology and formula sheet in,[7]

$$Q_j = \left(\frac{d_j \times 31}{r}\right) \tag{1}$$

where r is the "reaction parameter" given by:

$$r = \frac{2 \times bit\_rate}{picture\_rate} \tag{2}$$

where *bit_rate* is the bit rate specified for the video stream (say, 1.5 Mbps, 4 Mbps, and so on), *picture_rate* is the frame rate (we use a frame rate of 25 frames/sec) and d_j is a measure of the fullness of virtual buffers at macroblock j.[2,7]

In MPEG-2, these d_j terms vary depending on whether they belong to I, P or B frames. Then, depending on the spatial activity in each macroblock, the Q_j term is modulated to derive the quantization parameter $mquant_j$, that is used to quantize the macroblock. A spatial activity measure for the j^{th} macroblock is computed from the four luminance frame-organized sub-blocks (n=1..4) and the four luminance field-organized sub-blocks (n=5..8) using the original pixel values. We present the following expressions from the standard MPEG-2 formula sheet:

$$act_j = 1 + \min(vblk_1, vblk_2, ..., vblk_8) \tag{3}$$

$$vblk_n = \frac{1}{64} \times \sum_{k=1}^{64} (P_k^n - P\_mean_n)^2; 1 \leq n \leq 8 \tag{4}$$

$$P\_mean_n = \frac{1}{64} \times \sum_{k=1}^{64} P_k^n; 1 \leq n \leq 8 \tag{5}$$

$$\tag{6}$$

where P_k are the sample values in the n^{th} original 8×8 block. Here, act_j is called the activity of the j^{th} macroblock and it depends on the minimum variance of the 8×8 blocks. Let avg_act denote the average activity value. The normalized activity value N_act_j for the j^{th} macroblock is given by:

$$N\_act_j = \frac{(2 \times act_j) + avg\_act}{act_j + (2 \times avg\_act)} \tag{7}$$

The quantization parameter $mquant_j$ is given by:

$$mquant_j = Q_j \times N\_act_j \tag{8}$$

4. COMPUTING THE DATA HIDING CAPACITY

The question remains as to the optimal selection of the hiding parameters. For transmission rates $R = \frac{n}{q}$ (where the first (n) number of AC DCT coefficients, chosen by zigzag scan, are used to embed data per 8×8 block of the Y component of a raw frame, and q is the RA-code redundancy factor) which are less than the channel capacity C,[8] we can achieve perfect data recovery. Letting $q_{eff} = \frac{q}{n}$, we have : $R < C \Rightarrow \frac{1}{q_{eff}} < C \Rightarrow q_{eff} > \frac{1}{C}$. Thus, obtaining a proper q factor motivated the study of the channel capacity estimate.

Let us consider Fig. 3. Say, X is the input binary message to be embedded in a frame. We pass it through a channel encoder using a RA-q code to get a binary sequence Y which modifies the quantized DCT coefficients to embed 0 or 1. In the frame, if a quantized DCT coefficient $\in [-\frac{\Delta}{2}, \frac{\Delta}{2}]$, we regard it as an erasure. Thus, depending on the value of the quantized DCT coefficient, the data embedded in it is 0,1 or e (e denotes an erasure). Let us call this sequence of embedded data Z. Once we have hidden the data in the raw frames, we encode the frames to form a MPEG-2 video and transmit it.

After decoding the MPEG-2 video, we extract the embedded bits from the quantized DCT coefficients for each frame. The logic used is: 0 or 1 is extracted depending on whether the quantized coefficient is an even or odd multiple of Δ. Also, if the quantized coefficient $\in [-\frac{\Delta}{2}, \frac{\Delta}{2}]$, we assume it to be an erasure. Let the sequence of 0,1 and e at the decoder side be Z'. Now, depending on the sequence Z' (the initial log likelihood ratios (LLR) used in the iterative turbo decoding process are fixed according to the Z' values), we apply RA-q decoding and obtain a reconstructed version X' of the original input message X.

From the data hiding experiments, we obtain the transition probability matrices from Y to Z and from Z to Z'. The matrices depend on the parameters used in the data hiding scheme (n, Δ and QF) and in the MPEG-2 compression.

For our problem, the capacity can be rephrased as the data hiding capability in a frame given a certain hiding method and a certain compression scheme. For capacity computation, we need to maximize the mutual information between the input (Y) and output (Z') terms in the probabilistic part of the channel.

$$I(Y, Z') = \max_{p(y)} \sum_{y \in \{0,1\},\ z' \in \{0,1,e\}} p(y, z') \log \left\{ \frac{p(y|z')}{p(y)} \right\} \tag{9}$$

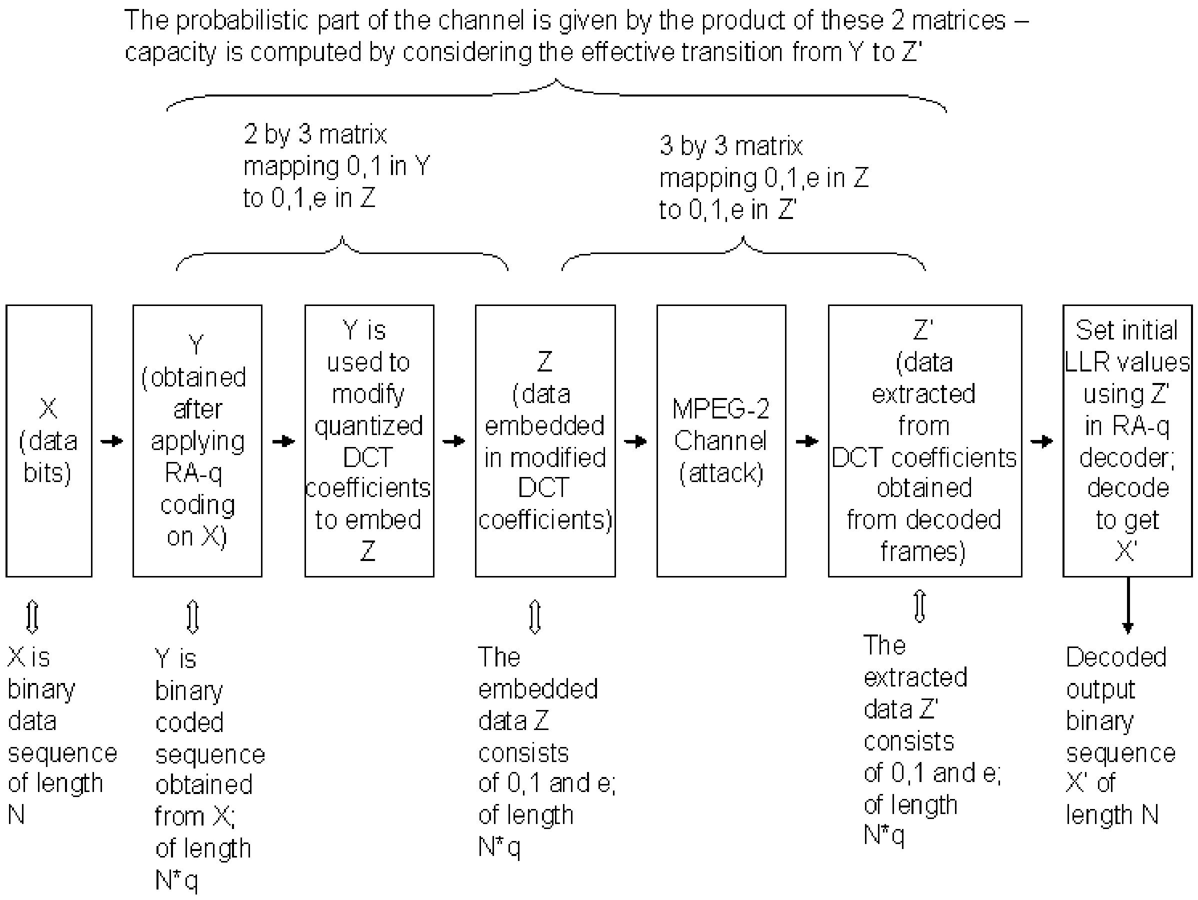

Figure 3. Computation of the data hiding capacity depends on the 2 by 3 transition probability matrix mapping Y to Z' (LLR = Log Likelihood ratio)

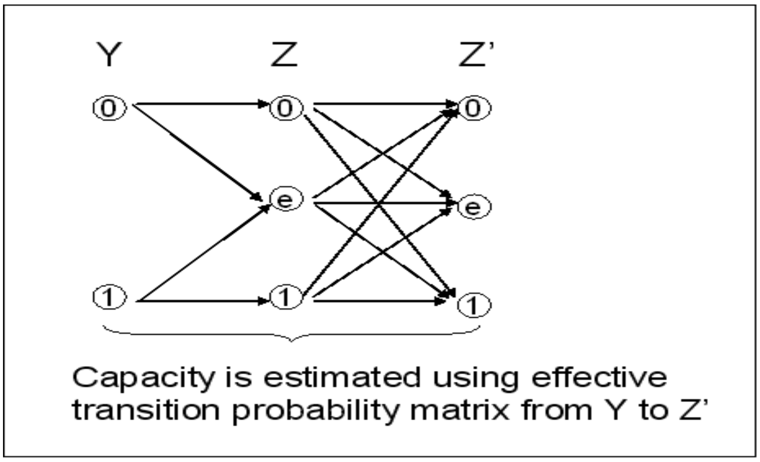

Figure 4. Binary sequence Y gets mapped to {0,1,e} in Z', Y and Z' having been shown in Fig. 3

For certain assumed priors $p(y)$ and the transition probability matrices (from Y to Z and from Z to Z', as shown in Fig. 4) being known, the probability terms involved in Eq. 9 can be computed. Now, the input message X (as in Fig. 3) is user-dependent. Since we use a pseudorandom interleaver in the RA-q code, we expect the sequence Y to have almost equal number of 0's and 1's in general. Thus, in almost all cases, the input probability distribution of $p(y)$ that achieves capacity is $p(y = 0) = p(y = 1) = 0.5$.

Once we compute the capacity value (C) for a frame, the optimum effective redundancy factor (q_{eff}) should be $\lceil \frac{1}{C} \rceil$. After fixing n, Δ and QF (hiding parameters) and the parameters used in MPEG-2 compression, we can compute C which in turn helps in estimating q. In our experiments, we found that for I and P frames, the q factor thus calculated (even lesser values suffice), helped in full data recovery but for B frames, we often needed a much higher q_{eff} than that given by $\lceil \frac{1}{C} \rceil$. Thus, the computed capacity is underestimated for I and P frames and overestimated for B frames.

5. EXPERIMENTS AND RESULTS

We now describe the actual experiments. We vary the data hiding parameters n, q and QF, depending on *mquant* and the frame type. Earlier in the discussion, "channel" was used for the transmission medium (mapping Z to Z' in Fig. 3). Here, we use the term "channel" for a hiding setup where a certain set of hiding parameters are used (mapping Y to Z in Fig. 3). We divide the frames into 3 channels (depending on the frame type) and each (I/P/B) channel is sub-divided further into 3-5 channels (depending on *mquant* for the frame). Thus, if we can use any one of 9(15) parameter sets in an adaptive framework, we say that 9(15) "channels" are being used. For frames with higher *mquant* (B frames), we use lower $\frac{n}{q}$ (and thus hide effectively less data) and lower QF (use of a lower quality factor matrix corresponds to coarser quantization and increased noise robustness) compared to those with lower *mquant* (I frames). For perceptual transparency, we use Δ=1.

Here, we describe the parameter allocation for the 15 channel case, which was empirically decided upon through experimentation. In the adaptive rate control scheme of MPEG-2,[7] the *mquant* parameter is clipped to the range [1..31]. Let $(3i-2)^{th}$, $(3i-1)^{th}$ and $(3i)^{th}$ channels, for $1 \leq i \leq 5$, correspond to I, P and B frames, respectively.

$$\left.\begin{array}{l} \text{(I frames) } n_1 = 6, \;\; q_1 = 15, \;\; QF_1 = 50, \\ \text{(P frames) } n_2 = 4, \;\; q_2 = 15, \;\; QF_2 = 40, \\ \text{(B frames) } n_3 = 2, \;\; q_3 = 15, \;\; QF_3 = 30, \end{array}\right\} \;\; 11 \leq mquant \leq 15$$

$$\begin{array}{l} n_i = n_{i-3} - 1, \;\; q_i = q_{i-3} + 4, \;\; QF_i = QF_{i-3} - 5, \;\; 4 \leq i \leq 6, \;\; 16 \leq mquant \leq 22 \\ n_i = n_{i-6} + 1, \;\; q_i = q_{i-6} - 2, \;\; QF_i = QF_{i-6} + 5, \;\; 7 \leq i \leq 9, \;\; 6 \leq mquant \leq 10 \\ n_i = n_{i-9} - 1, \;\; q_i = q_{i-9} + 7, \;\; QF_i = QF_{i-9} - 5, \;\; 10 \leq i \leq 12, \;\; mquant > 22 \\ n_i = n_{i-12} + 2, \;\; q_i = q_{i-12} - 2, \;\; QF_i = QF_{i-12} + 5, \;\; 13 \leq i \leq 15, \;\; mquant \leq 5 \end{array}$$

For a multi-channel approach, the receiver (decoder) should be aware of the set of possible parameters that can be used. If the hiding parameters match at the encoder and decoder sides and the noise due to compression attacks is small enough, the RA decoding generally converges in a few iterations. Thus, we vary the parameters (at the decoder side) over a set of allowable parameters, and then choose that set for which the RA decoding does converge.

We performed the experiments on 1.5 Mbps videos having a GOP length of 12 frames (distance between consecutive I and P frames=3). The noise introduced by the transmission medium was due to 2 sets of MPEG-2 attacks at varying bitrates (4 & 1.5 Mbps in one case and 8 & 1.5 Mbps in the other) and GOP length was varied from 6 to 24. For the 3 channel case, we had n=6, 4 and 2 for I, P and B frames, respectively, while n=3 for single channel case; the code redundancy factor q was set to 15. To compare the perceptual degradation introduced by the adaptive scheme, the following objective metrics[9] have been used:

1. spatial noise, which is measured per frame and it depends on the magnitude of the 2D Discrete Fourier Transform (DFT) of the luminance component of the difference (between original and recovered frame) signal, considered over a certain frequency band,
2. temporal flicker, caused by the difference in noise pattern across different frames. It is measured by taking 2D DFT of the temporal variation of the difference signal.

The absolute values of these distortion measures vary with DFT size (we use a 256×256 point DFT) and other scaling methods; their relative values help to understand how the perceptual quality varies with changing GOP size and the number of channels. The results are expressed in Table 1. The average number of data bits embedded per frame depends on the number of channels used and not on the parameters in the compression attack.

The notations used in Table 1 are as follows:

- FER = Frame Error Rate (fraction of the total frames for which the turbo decoder does not converge)
- PSNR = Peak-Signal to RMS-Noise Ratio, measured in dB
- 3 Channels - I/P/B channels are not sub divided further using *mquant*
- 9(15) Channels - each I/P/B channel is sub-divided into 3(5) channels using *mquant*

Table 1. Performance of the hiding scheme using different number of channels and varying compression parameters: "Bits per frame" and "FER" denote the per-frame data embedding rate and "frame error rate", respectively. The perceptual measures "Noise" (spatial noise) and "Flicker"[9] and PSNR (in dB) are averaged over all the frames.

GOP size	No. of Channels	Bits per Frame	4 Mbps & 1.5 Mbps				8 Mbps & 1.5 Mbps			
			FER	Noise	Flicker	PSNR	FER	Noise	Flicker	PSNR
6	1	240.0	.140	1114.4	21.54	36.40	.134	1087.5	21.08	36.62
	3	227.2	.091	1117.7	22.89	36.38	.085	1091.2	22.49	36.60
	9	277.2	.103	1123.7	22.56	36.34	.074	1096.5	22.16	36.56
	15	294.2	.082	1130.6	22.41	36.26	.074	1101.9	21.99	36.49
12	1	240.0	.088	1062.9	20.74	36.99	.077	1035.4	20.23	37.23
	3	227.2	.048	1068.9	22.25	36.95	.045	1039.0	21.69	37.21
	9	277.2	.028	1073.0	21.86	36.91	.020	1044.6	21.35	37.16
	15	294.2	.011	1079.6	21.70	36.84	.014	1049.9	21.17	37.10
15	1	240.0	.068	1059.2	20.73	37.05	.054	1030.7	20.31	37.30
	3	227.2	.037	1063.0	22.21	37.03	.025	1033.5	21.78	37.28
	9	277.2	.014	1067.0	21.81	36.99	.017	1037.8	21.35	37.25
	15	294.2	.011	1073.6	21.68	36.92	.011	1043.1	21.20	37.18
18	1	240.0	.057	1055.4	20.54	37.09	.042	1025.1	20.14	37.35
	3	227.2	.022	1059.4	22.13	37.07	.017	1027.8	21.69	37.34
	9	277.2	.020	1065.4	21.82	37.02	.011	1033.6	21.29	37.29
	15	294.2	.020	1070.7	21.66	36.95	.011	1039.0	21.11	37.23
24	1	240.0	.065	1050.7	20.51	37.16	.051	1022.9	20.07	37.40
	3	227.2	.017	1055.5	22.02	37.13	.022	1026.1	21.59	37.38
	9	277.2	.017	1059.9	21.70	37.09	.008	1031.1	21.20	37.34
	15	294.2	.008	1066.2	21.48	37.02	.005	1035.8	21.00	37.28

As seen from Table 1, using a higher number of channels, we are able to embed about 20% more data while the frame error rate is also reduced. If we use a single channel, we shall end up either incurring errors in smooth regions (with less hiding capability) or hiding less than the maximum possible in frames having greater spatial activity. From Table 1, the PSNR difference between single and 15-channel schemes is found to be less than 0.2 dB. From the relative variation of the noise and flicker terms, we see that the perceptual quality improves with increase in the GOP size. Also, the distortion increases with increase in adaptivity of the scheme. The flicker values show a curious trend - it actually decreases while going from 3 to 15 channels. The noise depends on the difference signal while the flicker depends on its time derivative. The apparently anomalous behavior of flicker can be thus explained - the difference signal is more dominant for 15-channel than for 3-channel case but its temporal variation is more marked for 3-channel case. Reducing the perceptual distortion provides scope for future work.

In MPEG-2, the encoding of a B or P frame depends on other frames - in frame-by-frame hiding, this temporal relationship is not maintained. E.g. for B frames, we reconstruct a B frame using two P frames or an I and a P frame; now, due to data hiding, there have already been modifications of the reference I and P frames - therefore, the distortion for the B frame is strongly dependent on the amount of noise introduced in these reference frames. Therefore, to improve the perceptual quality of the video in the still-image based hiding approach, the temporal information (and thus knowledge about the distortion of the reference frames involved in the prediction process) should also be included as an additional parameter.

6. CONCLUSION

In this paper, we have presented an adaptive hiding scheme where the embedding rate is varied for MPEG-2 video depending on the type of frame (I/P/B) and on the MPEG-2 reference quantization parameter, which depends on the spatial activity in the macroblocks of a frame. As there are a variety of parameters at which the encoder may hide the data, the different parameter sets to be used need to be pre-decided and made known to the decoder. However, at runtime, the parameter set per frame need not be explicitly sent to the decoder since the RA decoding generally converges only if the decoder uses the right combination from its known set of parameters. Though this adaptive scheme is able to embed more data than the fixed-parameter based method and the frame error rate is also minimized, the temporal flicker and spatial noise terms are increased for the adaptive case. In our future work, we wish to incorporate methodologies to minimize the temporal flicker by including the per-frame distortion information in our frame-by-frame based hiding framework.

ACKNOWLEDGMENTS

This research is supported in part by a grant from ONR # N00014-05-1-0816.

REFERENCES

1. G. K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM* **34**(4), pp. 30–44, 1991.
2. B. G. Haskell, A. Puri, and A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, Chapman and Hill, 1997.
3. K. Solanki, N. Jacobsen, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Robust image-adaptive data hiding based on erasure and error correction," *IEEE Trans. on Image Processing* **13**, pp. 1627 –1639, Dec 2004.
4. B. Chen and G. W. Wornell, "Quantization Index Modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Info. Theory* **47**, pp. 1423–1443, May 2001.
5. D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for turbo-like codes," in *36th Allerton Conf. on Communications, Control, and Computing*, pp. 201–210, Sept. 1998.
6. F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Info. Theory* **47**, pp. 498–519, Feb. 2001.
7. "http://www.mpeg.org/MPEG/MSSG/tm5/Ch10/Ch10.html." Chapter on Rate Control and Quantization Control - from www.mpeg.org.
8. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.
9. S. Winkler, E. D. Gelasca, and T. Ebrahimi, "Toward perceptual metrics for video watermark evaluation," in *Proc. of SPIE, Applications of Digital Image Processing*, pp. 5203–371–378, Aug. 2003.