

Joint Optimization of Data Hiding and Video Compression

Jithendra K. Paruchuri and Sen-ching Samson Cheung
Center for Visualization and Virtual Environments and Department of ECE
University of Kentucky, Lexington KY 40507
{jkparu0@engr.uky.edu, cheung@engr.uky.edu}

Abstract—From copyright protection to error concealment, video data hiding has found usage in a great number of applications. Recently proposed applications such as privacy data preservation require huge amount of information to be hidden inside a compressed video bitstream. Since data hiding disturbs the underlying statistical patterns of the source data, it adversely affects the performance of compression which are designed based on the statistical properties of the data. As such, it is imperative to design a data hiding scheme that is compatible with the compression algorithm and at the same time, introduces as little perceptual distortion as possible. In this paper, we propose a novel compression-domain video data-hiding algorithm that determines the optimal embedding strategy to minimize both the output perceptual distortion and the output bit rate. The hidden data is embedded into selective Discrete Cosine Transform (DCT) coefficients which are found in most video compression standards. The coefficients are selected based on minimizing a cost function that combines both distortion and bit rate via a user-controlled weighting. Two methods are proposed – exhaustive search and fast Lagrangian approximation. While the former produces optimal results, the latter approach is significantly faster and amenable to real-time implementation.

I. INTRODUCTION

Data hiding has been used in various applications like copyright protection, authentication, fingerprinting, error concealment, broadcast monitoring, covert communication, etc. Each application imposes different types of constraints in terms of capacity, security and robustness. In [1], we have proposed data hiding for privacy information preservation. Privacy is protected by obfuscating images of individuals from the video and the original data is preserved by hiding it in the compressed bit stream of the modified video. This is particularly useful when a condition arises to prove the authenticity of the modified video. Data hiding, however, is not the only approach – the authors in [2] and [3] scramble the pixels of the specific image objects for privacy protection. With the appropriate private key, the scrambling can be undone to retrieve the original. The drawback of these techniques is that it cannot be used with any other video modification techniques besides scrambling. Using data hiding for privacy data preservation is more flexible as it completely isolates preservation from modification. It can handle advanced modification techniques such as object removal using video inpainting as in [4].

Generally, there are two popular data embedding and extracting approaches – spread spectrum and quantization index

modulation (QIM)[5]. We use a QIM based approach, where quantization point is chosen based on parity of a hidden bit. This is chosen because it can handle more information compared to spread spectrum techniques. Even with compression, the size of the privacy information in our application usually exceeds 3000 bits per a 352×288 frame. Obviously, the perceptual quality after embedding such a large payload is of great concern.

In [6], the authors proposed to hide large volume of information into the nonzero DCT terms after quantization. This method cannot provide sufficient embedding capacity for our application because surveillance videos have high temporal redundancy, so more than 80% of the DCT coefficients will be zero in the inter-coded frames. In [7], the authors proposed to implement the embedding in both zero and non-zero DCT coefficients but only in macro blocks with low inter frame velocity. This framework deals only with minimizing perceptual distortion without considering the increase in bit rate. Our algorithm considers both rate and distortion and produces an optimal distribution of hidden bits among various DCT blocks. Our earlier work in [1] discussed a heuristic method to reach a compromise in distortion and output rate but no formal optimization is performed. In [8], rate-distortion optimization is used for data hiding in MPEG audio. Their scheme maximizes the channel capacity used for data hiding with a given distortion constraint.

In contrast, our problem is, given a fixed channel capacity, how to minimize simultaneously the distortion and rate resulted from data hiding. Our main contribution is an optimization framework to combine both the distortion and rate together as a single cost function and to use it in identifying the optimal locations to hide data. This allows a significant amount of information to be embedded into compressed bitstreams without disproportional increase in either output bit rate or perceptual distortion.

The rest of the paper is organized as follows: In Section II, we explain the constraints imposed on data hiding for the privacy preserving application. The perceptual distortion measure and embedding procedure are also explained in details. In Section III, we explain the cost function to be optimized followed by the algorithms using exhaustive search and Lagrange multipliers. In Section IV, we analyze the experimental results of rate-distortion based data hiding algorithms and conclude the paper in Section V.

II. DATA HIDING PROCEDURE

Privacy data preservation demands high embedding capacity with minimal perceptual distortion. As privacy protection is typically applied to surveillance video which needs to be stored for a prolong period of time, the video must be compressed to minimize the amount of storage space needed. Our earlier results in [1] showed a disproportional increase in the compressed bit rate after the data hiding process. The main reason for this increase is the introduction of rare run length patterns by data hiding that are incompatible with the entropy coding. Furthermore, as the output sequence needs to be decoded with a regular decoder, the embedding process must be inside the motion compensation loop. The noise introduced by the hidden data causes a decline in efficiency of the motion compensation process. As a result, the strategy proposed in [1] of choosing fixed locations for data hiding in all frames significantly reduces the compression performance. The solution is to follow an embedding strategy that can simultaneously minimize the distortion and the output bit rate.

In our proposed system, the information to be embedded is the compressed variable-bit-rate stream of the portion of the video that is affected by a video modification process designed to shield privacy data. The embedding process is done at frame level so that the decoder can reconstruct the privacy information as soon as the compressed bitstream of the same frame has arrived. Data is hidden only in the luminance DCT blocks which typically occupy the largest portion of the bit stream. The following subsections explain the embedding technique and the perceptual distortion measure used in our proposed system.

A. Embedding Process

To embed the data in the compressed bitstream, we follow the parity embedding approach which is a special case of QIM. We quantize the Discrete Cosine Transform (DCT) coefficients to odd or even quantization points depending on the bit to be embedded in that coefficient. Let $c(i, j, k)$ and $q(i, j, k)$ be the (i, j) -th coefficient of the k -th DCT block before and after quantization respectively. To embed a bit x into the (i, j, k) -th coefficient, we change $q(i, j, k)$ to $\tilde{q}(i, j, k)$ using the following embedding procedure:

- 1) If x is 0 and $q(i, j, k)$ is even, add or subtract one from $q(i, j, k)$ to make it odd. The decision of increment or decrement is chosen to minimize the difference between the reconstructed value and $c(i, j, k)$.
- 2) If x is 1 and $q(i, j, k)$ is odd, add or subtract one from $q(i, j, k)$ to make it even. The decision of increment or decrement is chosen to minimize the difference between the reconstructed value and $c(i, j, k)$.
- 3) $q(i, j, k)$ remains unchanged otherwise.

Following the above procedure, each DCT coefficient can embed at most one bit. Decoding is shown in Equation (1):

$$x = (\tilde{q}(i, j, k) + 1) \bmod 2 \quad (1)$$

This embedding, however, is not invertible. If the quantization stepsize used is QP, the embedding doubles the maximum

quantization distortion from QP to 2QP. This implies that the reconstructed video will be different from the originally compressed version. At a fine quantization level, the visual difference is minimal. On the other hand, such an embedding method is not suitable for applications that demands the original video to be unaltered by the data hiding process.

B. Perceptual Distortion

Common distortion measures like mean square does not work for our goal of finding the optimal DCT coefficients to embed data bits: given the number of bits to be embedded, the mean square distortion will always be the same regardless of which DCT coefficients are used as DCT is an orthogonal transform. Instead, we adopt the DCT perceptual model described in [9]. Consider the luminance and contrast masking of human visual system as described in [9], the final perceptual mask $s(i, j, k)$ that indicates the maximum permissible alternation to the (i, j) th coefficient of the k th 8×8 DCT block of an image can be calculated as:

$$s(i, j, k) = \max [t_L(i, j, k), |c(i, j, k)|^w t_L(i, j, k)^{1-w}] \quad (2)$$

with

$$t_L(i, j, k) = t(i, j) \left(\frac{c(0, 0, k)}{c_0} \right)^{\alpha_T} \quad (3)$$

for $i, j \in \{0, 1, \dots, 7\}$. $t(i, j)$ is the frequency sensitivity threshold, $w = 0.7$ is a constant, $c(0, 0, k)$ is the DC term of block k , $\alpha_T = 0.649$ is a constant, and c_0 is the average luminance of the image. With this perceptual model, we can compute a perceptual distortion value for each DCT coefficient in the current frame as:

$$D(i, j, k) = \frac{QP}{s(i, j, k)} \quad (4)$$

where QP is the quantization parameter used for that coefficient and $s(i, j, k)$ is the perceptual mask value calculated in Equation (2). Perceptually, just a few highly distorted coefficients account for more distortion than many mildly distorted ones. So an L_4 norm pooling is employed for calculating total distortion within a block or for the entire frame:

$$D = \left(\sum_{i,j,k} |D(i, j, k)|^4 \right)^{\frac{1}{4}} \quad (5)$$

III. PROPOSED RATE-DISTORTION OPTIMIZED DATA HIDING SCHEME

In our joint data hiding and compression framework, we aim at minimizing the output bit rate R and the perceptual distortion D caused by embedding M bits into the DCT coefficients. By using a user-specified control parameter δ , we combine the rate and distortion into a single cost function as follows:

$$C = (1 - \delta) \cdot N_F \cdot D + \delta \cdot R \quad (6)$$

N_F is used to normalized the dynamic range of D and R . δ is selected based on the particular application which may favor the least amount of distortion by setting δ close to zero, or

the least amount of bit rate increase by setting δ close to one. In the following two sections, we describe two approaches to minimize C given a target number of bits to be embedded. In order to avoid any overhead in communicating the embedding positions to the decoder, both of these approaches compute the optimal positions based on the previously decoded DCT frame so that the process can be repeated at the decoder.

A. Exhaustive Search

Minimizing bit rate and distortion with a constraint on the target number of bits to be hidden is a combinatorial optimization problem. Distortion is easy to handle because it is a fixed value throughout the embedding process for a particular coefficient. Output rate depends on the run-length patterns and thus changes after each embedding. To obtain the true optimal selection of DCT coefficients, one needs to build a trellis of available DCT coefficients versus bits to be embedded and exhaust all possible paths. Since the embedding in one coefficient affects multiple run-length patterns, the number of states maintained during the traversal of the trellis grows exponentially with the number of bits to be embedded. This is clearly impossible to carry out in practice.

Instead, we propose to minimize the combined cost of distortion and rate as described in Equation (6) by a greedy approach. Each data bit is embedded at the minimum cost position. The cost function is then updated according to the change of bit rate after each embedding. The same process is repeated until all the data bits are embedded. To minimize the complexity in searching for the position with the lowest cost, we use a B-tree to keep the cost for each coefficient sorted at all time. The insertion and deletion from the B-tree are based on the highly-optimized embedded database package BerkeleyDB. Still, its running time depends on the number of bits to be embedded and is usually far from real time. A 10 sec of CIF video(352x288) at 30 fps took hours to complete the embedding. Furthermore this greedy algorithm needs random access to the entire frame contents during the optimization process and thus may not be suitable for embedded processors that can only process a portion of a video frame at any one time. Due to these shortcomings, we propose a much more efficient algorithm in the following section based on approximating the combinatorial optimization problem as a continuous optimization problem using the Lagrangian Multipliers.

B. Lagrangian Approximation

Lagrangian method turns a constrained optimization problem into an unconstrained one, and is commonly used in rate-distortion optimized video compression. In our data hiding framework, the constrained optimization can be formulated as follows:

$$\min_{\Gamma} C(\Gamma) \text{ subjected to } M = N \quad (7)$$

where M is the variable that denotes the number of bits to be embedded, N is the target number of bits to be embedded, C is the cost function as described in Equation (6) and Γ is a possible selection of N DCT coefficients for embedding the data.

Using Lagrangian Multiplier, this constrained optimization is equivalent to the following unconstrained optimization:

$$\min_{\Gamma} \Theta(\Gamma, \lambda) \text{ with } \Theta(\Gamma, \lambda) = C(\Gamma) + \lambda \cdot (M - N) \quad (8)$$

We can further simplify Equation (8) by decomposing it into the sum of similar quantities from each DCT block k :

$$\Theta(\Gamma, \lambda) = \sum_k C_k(\Gamma_k) + \lambda \cdot \left(\sum_k M_k - N \right) \quad (9)$$

Here we make two assumptions: first, we approximate the number of bits M_k to be embedded in block k as a real value. Second, instead of having C_k being a function of the selection of DCT coefficients Γ_k , we assume C_k as a differentiable function of M_k . In other words, we assume the existence of a one-to-one relationship between M_k and Γ_k . Based on these assumptions, the optimal solution to (8) must satisfy

$$\frac{dC_k}{dM_k} = -\lambda \quad (10)$$

for all k .

To prepare for the above optimization, we need to first generate the curves between cost and number of bits for all the DCT blocks. M_k ranges from 0 to 64 for a 8×8 DCT block. The cost function, as described in Equation (6) consists of both the distortion and the rate. The distortion is calculated based on perceptual masking and pooling as described in section II-B. Rate increase is directly calculated by embedding a '0' bit onto the existing DCT block using the run-length and entropy coding in the compression process. Instead of computing the best possible selection of coefficients for each M_k which is a very time-consuming process, we heuristically fix the selection order starting from the highest frequency coefficients and traversing back in a zigzag fashion.

We further approximate the discrete data points to a continuous one by fitting a second order curve as in Figure 1.

$$C_k \approx a_k \cdot M_k^2 + b_k \cdot M_k + c_k \quad (11)$$

The final step is to search for an optimal slope that satisfies our embedding constraint. The slope can be found as follows

$$\frac{dC_k}{dM_k} = 2 \cdot a_k \cdot M_k + b_k = -\lambda \quad (12)$$

To meet the minimum embedding constraint, the total number of bits embedded from each DCT block must be equal to N :

$$N = \sum_k M_k = -\lambda \cdot \sum_k \left[\frac{1}{2 \cdot a_k} \right] - \sum_k \left[\frac{b_k}{2 \cdot a_k} \right]$$

Thus, λ can be determined as follows:

$$\lambda = - \frac{N + \sum_k \left[\frac{b_k}{2 \cdot a_k} \right]}{\sum_k \left[\frac{1}{2 \cdot a_k} \right]} \quad (13)$$

Since the actual problem is a discrete one, we use λ from Equation (13) as an initial slope and search for the exact slope in its neighborhood to match our target requirements. At this optimal slope on each curve, we can identify the number of data bits M_k to be embedded at each DCT block.

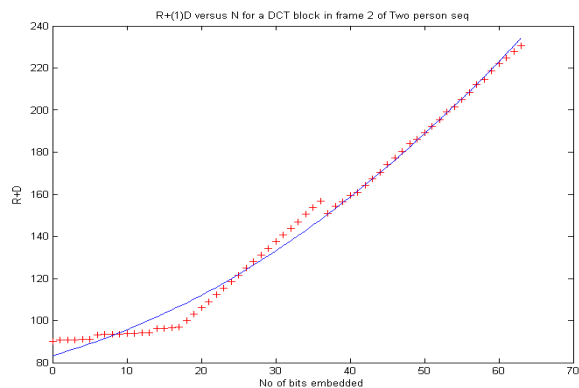


Fig. 1. Sample Distribution Curve between Cost and M

IV. EXPERIMENTS

We tested our rate distortion based data hiding algorithms on two standard test sequences: “hall monitor” and “foreman”. Both sequences has 299 frames and are in CIF format (352×288). The “hall monitor” sequence is modified by inpainting one of the persons with an adaptive estimate of the background. The image of the person being removed is compressed using a regular H.263 encoder with constant quantization. This privacy information averages 2700 bits per frame but can fluctuate from 1000 bits to 20000 bits for individual frame. Results of the Exhaustive Search method and the Lagrangian Approximation on the “hall monitor” sequence for QP=10 are shown in the first table. The PSNR value stays almost the same at 32.7 db for all the cases. $\delta = 0$ denotes pure distortion optimization and $\delta = 1$ indicates rate optimization. The results obtained by the Exhaustive Search are better than those of the Lagrangian Approximation but the former takes almost 7 hours compared to 15 minutes by the latter one to embed the hidden information. Our results also significantly outperforms our prior results of 922 kbps reported in [1]. Figure 2 shows one sample frame using the Lagrangian Approximation with different δ . The presence of hidden data is not visible for both $\delta = 0$ and $\delta = 0.5$ and becomes marginally visible at $\delta = 1$. The second table shows similar results for embedding the same privacy information in the Foreman sequence at two different quantization parameters.

Hall Monitor Sequence (QP=10)

δ	Lagrangian Approximation		Exhaustive Search	
	Rate	Distortion	Rate	Distortion
Original	119.15 kbps	0	119.15 kbps	0
0	636.89 kbps	18.9	517.56 kbps	13.48
0.5	562.03 kbps	25.3	502.07 kbps	18.84
1	370.26 kbps	104.7	310.52 kbps	140.46

V. CONCLUSIONS

In this paper, we have presented a new rate distortion optimized video data-hiding algorithm which distributes hidden data among various DCT blocks. Our results show that

Foreman Sequence

δ	QP = 10		QP = 5	
	Rate	Distortion	Rate	Distortion
Original	406.30 kbps	0	1100.98 kbps	0
0	742.60 kbps	8.54	1383.78 kbps	4.5
0.5	678.17 kbps	8.83	1278.31 kbps	5.14
1	609.80 kbps	10.81	1261.09 kbps	5.16



Fig. 2. 234th frame of Hall Monitor Sequence after data hiding for QP = 10. Top Left: No Watermark ; Top Right: $\delta = 0$; Bottom Left: $\delta = 0.5$; Bottom Right: $\delta = 1$

it performs better than equal distribution or only distortion-oriented distribution of hidden bits.

REFERENCES

- [1] W. Zhang, S.-C. Cheung, and M. Chen, “Hiding privacy information in video surveillance system,” in *Proceedings of the 12th IEEE International Conference on Image Processing*, Genova, Italy, Sept. 2005, pp. 868–871.
- [2] T. E. Boulton, “Pico: Privacy through invertible cryptographic obscuration,” in *Computer Vision for Interactive and Intelligent Environments - the Dr. Bradley D. Carter Workshop Series*, 2005.
- [3] F. Dufaux and T. Ebrahimi, “Scrambling for video surveillance with privacy,” *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW’06)*, p. 160, 2006.
- [4] S.-C. S. Cheung, J. Zhao, and M. V. Venkatesh, “Efficient object-based video inpainting,” in *Proceedings of IEEE International Conference on Image Processing (ICIP 06)*, 2006, pp. 705–708. [Online]. Available: <http://dblp.uni-trier.de/db/conf/icip/icip2006.html#CheungZV06>
- [5] Chen and Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” in *ISIT: Proceedings IEEE International Symposium on Information Theory*, 2000. [Online]. Available: <http://citeseer.ist.psu.edu/chen99quantization.html>
- [6] K. Solanki, N. Jacobsen, S. Chandrasekaran, U. Madhow, and B. Manjunath, “High-volume data hiding in images: introducing perceptual criteria into quantization based embedding,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal processing (ICASSP’02)*, vol. 4, 2002, pp. 3485–3488.
- [7] A. Sur and J. Mukherjee, “Adaptive data hiding in compressed video domain,” in *ICVGIP*, 2006, pp. 738–748.
- [8] P. Prandoni and M. Vetterli, “R/D optimal data hiding,” in *SPIE Proceedings on Security and Watermarking of Multimedia Contents*, 1999, pp. 375–385. [Online]. Available: citeseer.ist.psu.edu/prandoni99rd.html
- [9] I. Cox, M. Miller, and J. Bloom, *Digital Watermarking*. Morgan Kaufmann Publishers, 2002.