The Choice of Watermark Domain in the Presence of Compression

Chuhong Fei, Deepa Kundur and Raymond Kwong
Edward S. Rogers Sr. Department of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario, Canada M5S 3G4
Email: fei@control.toronto.edu, deepa@comm.toronto.edu, kwong@control.toronto.edu

Abstract

In this paper, we determine the watermark domain that maximizes data hiding capacity. We focus on the situation in which the watermarked signal undergoes lossy compression involving quantization in a specified compression domain. A novel linear model for the process of quantization is proposed which leads to analytical results estimating the data hiding capacity for various watermarking domains. Using this framework we predict appropriate transforms for robust spread spectrum data hiding in the face of JPEG compression. Simulation results verify our theoretical observations. We find that a repetition code used in conjunction with spread spectrum watermarking in a different domain than employed for compression improves data hiding capacity.

1. Introduction

In this paper, we focus on identifying general rules of thumb for reliable high capacity watermarking in the presence of compression. Our intent is to take a communication analogy for watermarking and answer the fundamental questions:

- For lossy compression involving quantization in a specific transform domain, what domain is best suited for reliable watermark embedding and extraction.
- How much information can I reliably hide?

These questions lead us to a more analytic and information theoretic approach to addressing the problem of data hiding in the presence of compression. We hope that the insights gained through this work may be applied directly to previously proposed and future robust watermarking algorithms to enhance performance.

Watermarking is emerging as a technology useful to not only copy protection and tamper assessment applications, but for broadcast monitoring and signal tagging. For the two latter applications, sophisticated attacks are not necessarily the leading threat. Instead, practical compression is the most common form of incidental distortion that limits the robustness or capacity of a data hiding or watermarking scheme. Thus, in this work we address the problem of high capacity watermarking in the presence of perceptual coding. Currently, the most common lossy image compression standard is JPEG. Therefore, we focus on deriving capacity results for watermarking in various transform domains assuming that the watermarked image undergoes JPEG compression prior to watermark detection or extraction.

Several papers have dealt with integrating watermark embedding with compression. Wolfgang *et al*. [6] suggest that one must use the same transforms for watermarking as for compression to maximize robustness. Their simulations, however, are somewhat inconclusive as the results do not strongly support their hypothesis. Kundur *et al* [3] finds that use of the same transform for embedding and compression is not optimal, and suggests that complementary transforms will result in superior performance. Ramkumar*et al* [4, 5] indicate that use of transforms with poor energy compaction properties, which are not suitable for compression, provides greater watermark capacity.

In this work, we conclude that the use of differing transforms results in the best capacity when a repetition code is used for watermark embedding (i.e., the same watermark sequence is repeatedly embedded in the image). The use of the same transform yields good results when one long watermark sequence is embedded in the image. Overall, the data hiding capacity is greater when a repetition code is used with spread spectrum watermarking in a domain different than used for compression. The results are derived analytically and verified through simulations.

This paper is organized into six sections. The next section introduces the specific problem that we consider. Section 3 introduces the figures of merit and models used to derive our results. Novel models and analysis results are presented in Section 4, followed by simulation results and final remarks in Sections 5 and 6.

2. Problem Formulation

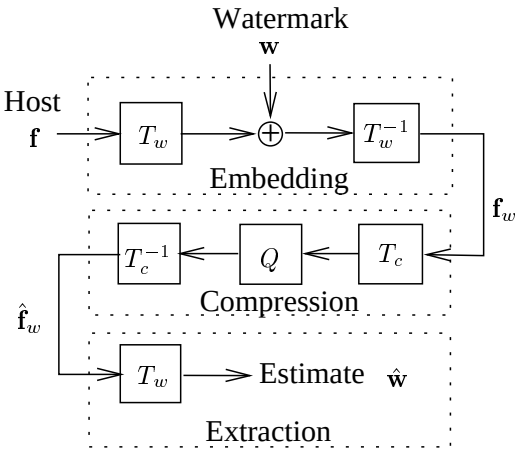

Figure 1. The proposed joint watermarking and compression scenario.

The block diagram of a typical watermark embedding scheme in the presence of lossy compression is shown in Figure 1. There are three basic stages: watermark embedding, lossy compression and watermark detection. Generally, the embedding process occurs in a *watermark domain*. An orthogonal transformation T_w is applied to the host image $\mathbf{f}$. The transformation decomposes the host image $\mathbf{f}$ into coefficients to which the watermark is embedded. Taking the inverse transform T_w^{-1} produces watermarked image $\mathbf{f}_w$ in pixel domain which is designed to be perceptually identical to the original image $\mathbf{f}$. Most commonly used transforms include discrete cosine transform (DCT), wavelet transforms, the Hadamard transform, the discrete sine transform, the discrete Fourier transform (DFT) and the Karhunen Loeve transform (KLT).

We consider the situation in which such compression is applied *after* watermark embedding. Lossy compression is a quantization process in a *compression domain* T_c such as DCT domain for JPEG. The resulting compressed watermarked image is $\hat{\mathbf{f}}_w$.

At the receiver, the hidden message $\hat{\mathbf{w}}$ is extracted from the "corrupted" watermarked image $\hat{\mathbf{f}}_w$ in the watermark domain. The existence of the original watermark $\mathbf{w}$ within $\hat{\mathbf{f}}_w$ is detected by calculating the correlation between the original watermark $\mathbf{w}$ and the extracted watermark $\hat{\mathbf{w}}$. If the correlation coefficient is above a given threshold the watermark is considered to be detected, otherwise, the watermark is considered not to be present in the image.

It is of great interest to determine the best transformation domain in which to devise robust watermark embedding methods given that compression occurs in a specific domain, for example, DCT for JPEG. In this paper, we attempt to investigate analytically the choice of watermark domain for high capacity data embedding in the presence of compression.

3. Models and Measures

In the section, we introduce the basic models and figures of merit used in our analysis work.

3.1. Spread Spectrum Watermarking

Many proposed watermarking schemes borrow ideas from spread spectrum communications [1]. They embed a watermark by adding a pseudo-noise (PN) sequence with low amplitude to the host image. This specific PN sequence is detected using a correlation receiver.

Let $x = [x_1, x_2, \cdots, x_N]$ be the host image coefficients in watermark domain. The watermark consists of a sequence of numbers, $\mathbf{w} = [w_1, w_2, \cdots, w_N]$ with a given statistical distribution, such as a normal distribution $\mathcal{N}(0, 1)$ with zero mean and a unit variance. The watermark is embedded into the coefficients $\mathbf{x}$ according to the relationship

$$y_i = x_i + a_i w_i, \tag{1}$$

where a_i is a scaling parameter which determines the extent to which one can alter x_i to keep the perceptual fidelity of the image; y_i is watermarked coefficient.

To verify the presence of the watermark $\mathbf{w}$, the similarity between the watermark domain coefficients $\hat{\mathbf{w}}$ of the possibly tampered image $\hat{f}_w$, and the original watermark $\mathbf{w}$ is measured. The similarity measure is given by the normalized correlation coefficient.

$$\rho(\mathbf{w}, \hat{\mathbf{w}}) = \frac{\sum_{i=1}^{N} w_i \hat{w}_i}{\sqrt{\sum_{i=1}^{N} w_i^2} \sqrt{\sum_{i=1}^{N} \hat{w}_i^2}}, \tag{2}$$

where $\hat{\mathbf{w}} = [\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_N]$.

In this work, we assume that watermarking embedding and detection takes place in this manner.

3.2. Watermark Capacity

One popular analogy for watermark embedding and detection in the presence of compression is data communications over a noisy channel. Communicating the watermark is analogous to transmission of the watermark through an associated *watermark channel*.

For this watermark channel, there exists a *watermark capacity* which is the maximum number of bits that can be hidden and recovered successfully (i.e., with an arbitrarily

low probability of bit error) from the possibly tampered watermarked image. The more robust and effective a watermarking scheme, the greater its watermark capacity which makes this measure ideal for evaluation of the success of any robust data hiding scheme.

Our task in this paper is to identify appropriate watermark domains for robustness against JPEG compression. By using watermark capacity as our figure of merit, our goal is essentially to determine the watermark transform domain which maximizes this capacity.

3.3. Noise

In the case of *blind* watermarking the host image is not available for extraction and the associated watermark channel has two sources of noise: 1) the noise due to the original image; 2) the attack noise due to compression/decompression process [4, 5].

A common model of an image involves representing it as a wide sense stationary Gaussian stochastic process with a specified covariance [2]. Let $\{f_{ij}\}$ denote two-dimensional image sequence in the spatial domain defined on a rectangular grid; f_{ij} is assumed to be a two-dimensional stationary stochastic process. One typical covariance function model used in image processing is [2]

$$\text{Cov}(f_{ij}, f_{mn}) = \sigma^2 \rho_1^{|i-m|} \rho_2^{|j-n|} \tag{3}$$

where ρ_1, ρ_2 are one step column and row correlation parameters.

A linear orthogonal transformation of the image may be represented as follows

$$X = TFT^{-1} \tag{4}$$

where $F \triangleq \{f_{ij}; 1 \leq i \leq N, 1 \leq j \leq N\}$ is the image matrix; T is an orthogonal transform matrix; X is the coefficient matrix in a transform domain.

Let K_1, K_2 be two matrices with elements $K_1(i,j) = \rho_1^{|i-j|}$, and $K_2(i,j) = \rho_2^{|i-j|}$ respectively. From (3) and (4), the covariances of the transform coefficients X are derived to be

$$\text{Cov}(x_{ij}, x_{mn}) = \sigma^2 K_1^*(i,m) K_2^*(j,n) \tag{5}$$

where $K_1^* = TK_1T^{-1}$ and $K_2^* = TK_2T^{-1}$.

Given the basic models and assumptions established in this section, we next introduce a novel model for compression in order to derive our capacity results.

4. A Novel Compression Attack Model

4.1. Quantizer Model

The lossy compression we consider in this paper involves quantization of signal coefficients in a *compression domain* such as the DCT domain (for JPEG). We denote the associated transformation from the pixel domain to the compression domain with T_c. Both the host image signal and the watermark signal pass through a quantizer, as shown in Figure 2, where w is the watermark information and x is the host image in the T_c domain.

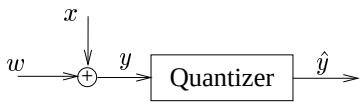

Figure 2. Quantizer. The watermark w and host image x jointly undergo quantization to produce the watermarked quantized signal $\hat{y}$.

The existence of a nonlinear element such as a quantizer in the watermark channel makes it difficult to analyze the relationship between the original watermark and extracted watermark. However, since watermark detection involves computing the correlation coefficient between the original and extracted watermark (i.e., the input and the output of the watermark channel), we treat this channel as a black box and propose a more tractable linear model which still captures the essential characteristics of the effect of compression on w in terms of correlation.

Suppose x and w are two independent random variables with zero mean. The transform T_c coefficient value of the watermarked image prior to quantization is given by $y = x + w$. Let Δ be the quantization step. Then the quantized coefficient is given as follows

$$\hat{y} = [y]_\Delta = \text{round}(\frac{y}{\Delta})\Delta, \tag{6}$$

where round$(\cdot)$ denotes rounding to the nearest integer, and $[\cdot]_\Delta$ denotes the quantization operation with step Δ.

Our novel additive model for quantization is to replace $\hat{y}$ with

$$\tilde{y} = \alpha w + \beta x, \tag{7}$$

where α and β are two parameters set such that the output of the novel additive model has the same power $E\{\hat{y}^2\}$, and the same correlation to w, $E\{w\hat{y}\}$. That is, α and β are set such that

$$E\{\tilde{y}^2\} = E\{\hat{y}^2\}, \tag{8}$$

$$E\{w\tilde{y}\} = E\{w\hat{y}\}. \tag{9}$$

It can be shown that the following assignments obey equations (8) and (9):

$$\alpha = \frac{E\{w\hat{y}\}}{E\{w^2\}}, \tag{10}$$

$$\beta = \sqrt{\frac{E\{\hat{y}^2\} - \alpha^2 E\{w^2\}}{E\{x^2\}}}. \tag{11}$$

The model of (7 essentially tries to account for the varying degree of effect of the quantization process on the host and watermark components of y. Clearly, because w is much lower in amplitude than x for transparency of the watermark, the influence of w after quantization on $\hat{y}$ will be much smaller if not negligible than the influence on x. Thus, we expect α to be smaller than β.

4.2. The Scheme

In this section we incorporate the novel linear quantization model discussed in Section 4.1 into our analysis framework. Figure 3 shows an overall representation of the watermarking and quantization-based compression process. Spread spectrum watermarking occurs in the T_w domain and compression in the T_c domain where T_w and T_c are both orthogonal transformation matrices.

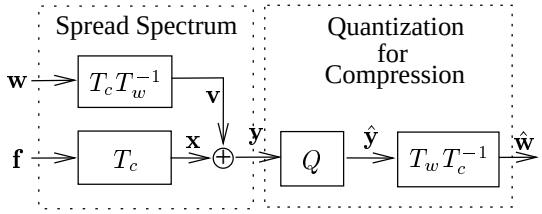

Figure 3. Representation of the overall watermarking and compression process.

The popular JPEG compression standard is based on the 8×8 block DCT. In order to evaluate the effects of different watermark domains given that compression occurs in the DCT domain, we also consider 8×8 image segment transforms for watermarking.

Although an image is often represented by a two-dimensional signal, for our purposes we regard the image signal as a one-dimensional sequence acquired by a columnwise reordering operation. After column by column scanning, each 8×8 block field is regarded as a 64×1 vector. Let $n = 64$ be the length of the vector. Suppose $\mathbf{f} = [f_1, f_2, \cdots, f_n]^T$ is an image block in the pixel domain, $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T$ is the image coefficient in the compression domain, and $\mathbf{w} = [w_1, w_2, \cdots, w_n]^T$ is the watermark signal in the watermark domain. From Figure 3, we see that

$$x = T_c f, \tag{12}$$

$$v = T_c T_w^{-1} w. \tag{13}$$

In practical lossy compression, the quantization is applied to all coefficients in varying degrees depending on a quantization table. By modeling the quantization of coefficient i as described in Section 4.1 using parameters α_i and β_i, we can establish a relationship between the watermark signal embedded and its extracted version after quantization. Figure 4 shows the overall coupled parallel channel model.

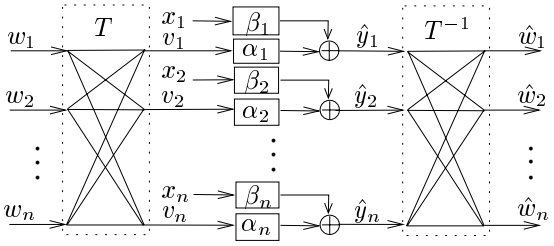

Figure 4. Equivalent parallel additive watermark channel model.

Let $A = \text{diag}(\alpha_1, \alpha_2, \cdots, \alpha_N)$, $B = \text{diag}(\beta_1, \beta_2, \cdots, \beta_n)$. Then $\hat{y} = Av + Bx$. Let

$$T = T_c T_w^{-1}. \tag{14}$$

From (13), we have

$$\hat{w} = T^{-1}\hat{y} = T^{-1}ATw + T^{-1}Bx. \tag{15}$$

Since T_w, T_c are orthogonal matrices, T is orthogonal matrix as well. $T^{-1} = T^T$. Let $S = T^T A T$, $z = T^T B x$, then

$$\hat{w} = Sw + z. \tag{16}$$

Generally, S is not a diagonal matrix, so there exists interference between watermark inputs in different parallel channels. However, it represents a symmetric interference channel for which the channel capacity depends on the dependence between watermark inputs in different channels. In this paper, we only consider two extreme cases of dependence: (1) fully dependent watermark inputs, or (2) independent watermark inputs. To calculate the capacity, we make the following assumptions:

- The image signal $\mathbf{f}$ is Gaussian distribution as discussed in Section 3.3, so $\mathbf{x}$, $\mathbf{z}$ are all normally distributed random variables whose variances can be derived from $\mathbf{f}$.

- In channel i, the watermark signal component w_i is also Gaussian distributed with variance $\sigma^2_{w_i}$.

Case 1: Fully dependent watermark inputs

Suppose $w_1, w_2, \cdots, w_n$ are fully dependent, namely for all i, $w_i = \sigma_{w_i} u$ where u is a standard normal distribution $\mathcal{N}(0,1)$ and σ_{w_i} is the amplitude of watermark signal in channel i. This corresponds to the situation in which the watermark is repeated throughout the signal. We can exploit the full correlation and treat the watermark components from other channels w_j, for all $j \neq i$, as equivalent to the input signal w_i. The only noise contribution is from z_i. Therefore,

$$\hat{w}_i = z_i + \sum_{j=1}^{n} s_{ij} w_j \tag{17}$$

where $z_i = \sum_{k=1}^{n} T_{ki}\beta_k x_k$ and $s_{ij} = \sum_{k=1}^{n} T_{ki}\alpha_k T_{kj}$. Then, the channel capacity of channel i is

$$C_i = \frac{1}{2}\log_2(1 + \frac{(\sum_{j=1}^{n} s_{ij}\sigma_{w_j})^2}{\sigma^2_{z_i}}) \tag{18}$$

where $\sigma^2_{z_i} = \sum_{k=1}^{n} T^2_{ki}\beta^2_k\sigma^2_{x_k}$. The overall channel capacity is $C = \sum_{i=1}^{n} C_i$.

Case 2: Independent watermark inputs

Suppose w_i and w_j for $i \neq j$ are independent. This corresponds to the situation in which one long spread spectrum watermark sequence is embedded. For channel i, other watermark signals $w_j, j \neq i$ are regarded as noise. Therefore,

$$\hat{w}_i = s_{ii} w_i + z_i + \sum_{j=1, j\neq i}^{n} s_{ij} w_j, \tag{19}$$

where $z_i = \sum_{k=1}^{n} T_{ki}\beta_k x_k$ and $s_{ij} = \sum_{k=1}^{n} T_{ki}\alpha_k T_{kj}$. Then the channel capacity is

$$C_i = \frac{1}{2}\log_2(1 + \frac{s^2_{ii}\sigma^2_{w_i}}{\sigma^2_{z_i} + \sum_{j\neq i} s^2_{ij}\sigma^2_{w_j}}), \tag{20}$$

where $\sigma^2_{z_i} = \sum_{k=1}^{n} T^2_{ki}\beta^2_k\sigma^2_{x_k}$. Once again, the overall channel capacity is $C = \sum_{i=1}^{n} C_i$.

5. Investigative Results

We consider several watermark transform T_w domains: (1) pixel, i.e. the identity transform; (2) Karhunen Loeve transform (KLT); (3) Discrete Cosine transform (DCT); (4) Hadamard transform; (5) Wavelet transform, in particular, Daubechies wavelet; (6) Slant transform. Given a JPEG compression quality level, the associated quantization table is fixed and known, and the channel capacity of each potential transform T_w can be calculated for cases 1 and 2.

For comparison, we assume that the amplitude of watermark in all channels is equal to a constant, i.e. for all i, $\sigma^2_{w_i} = a$. And we choose $a = 4$ to guarantee imperceptibility of the watermark embedding. Two parameters ρ_1, ρ_2 in image covariance model given by Equation (3) are estimated from the classic image of Lena (also used in our verification simulation results) by least-mean-square-error method.

Figure 5 shows the predicted channel capacity of six different watermark transforms for Case 1 when the watermark inputs in different channels are fully dependent. From the results, we can see that Hadamard and wavelet transforms are better for robust data hiding than other transforms. In particular, Hadamard Transform is the best transform when JPEG compression occurs in the very common range of JPEG quality $40 \sim 85$, and the pixel domain is the worst for all cases of watermarking. The results for Case 2 when the watermark inputs are all independent in different channels are shown in Figure 6. For any JPEG quality, KLT and DCT are the two best among all these transforms and SLANT is also good transform, better than Hadamard and Wavelet. The pixel domain is still the worst domain for watermarking.

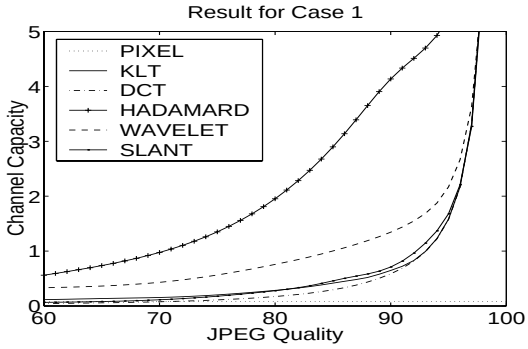

Figure 5. Channel capacity of different watermark transforms for Case 1.

In order to verify our analytical results, we implemented spread spectrum watermarking simulations on a real test image of Lena. A watermark sequence with normal distribution $\mathcal{N}(0,4)$ is embedded into each decomposition coefficient channel. Since capacity is a theoretical measure which cannot be directly computed from our watermarking simulation, we estimate the capacity of the individual channel by formula $C = \frac{1}{2}\log_2(\frac{1}{1-\rho^2})$ where ρ is the correlation coefficient between the original watermark and the extracted watermark and it is assumed that the extracted watermark undergoes additive white Gaussian noise. Figure 7 and 8 show the overall channel capacity based on simulation on

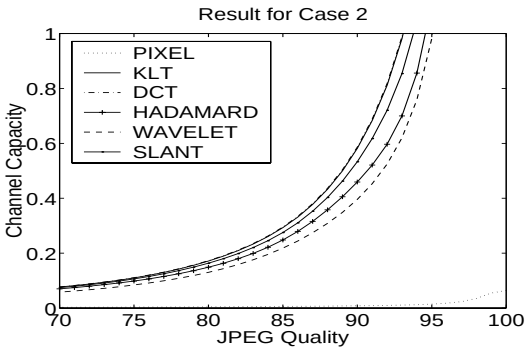

Figure 6. Channel capacity of different watermark transforms for Case 2.

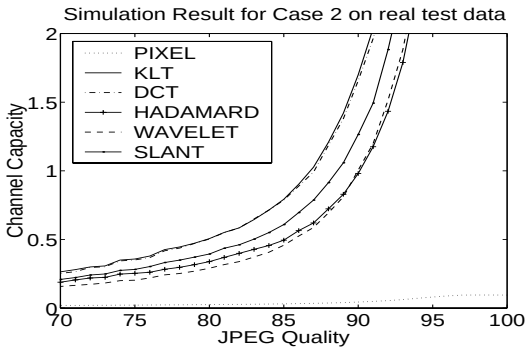

Figure 8. Simulation results on test image Lena for Case 2.

test image Lena for Case 1 and Case 2 respectively based on our analytical model. We can see that the experimental results nearly follow our theoretical calculations.

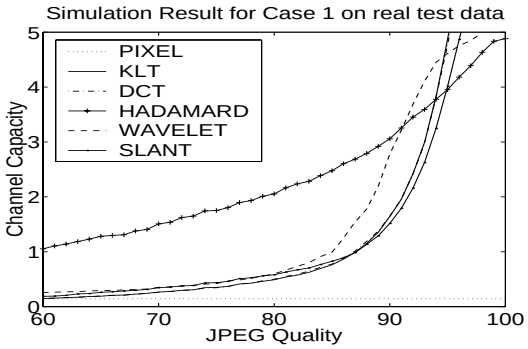

Figure 7. Simulation results on test image Lena for Case 1.

6. Concluding Remarks

We have described one novel approach to analytically evaluate the behavior of watermarking in a watermark domain given that lossy compression occurs. By approximating a non-linear quantizer as a linear model, channel capacity is calculated to measure the efficiency of the associated watermark channel. Based on this novel model, we find that the Hadamard transform is better than other commonly used transforms in the case that the watermarks embedded in different decomposition channels are fully dependent. On the other hand, in the case that the watermarks embedded in different decomposition channels are independent, it is better to choose watermark domain to be the same as compression domain, i.e. DCT for JPEG. We also have shown that the experimental results nearly follow the theoretical calculation, which shows that our underlying model is at least, in part, sound.

We focus on spread spectrum-based watermarking in the paper, future research will be devoted to investigate quantization based watermarking algorithm and try to determine the best transform domain for quantization based watermarking algorithm given that JPEG compression occurs. In addition, we are currently working on using our compression attack model to develop a highly quantization-robust blind data hiding algorithm.

References

[1] I. J. Cox, J. Killian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. Technical Report 95-10, NEC Research Insitute, 1995.

[2] A. K. Jain. Advances in mathematical models for image processing. *Proceedings of the IEEE*, 69(5):502–528, May 1981.

[3] D. Kundur and D. Hatzinakos. Mismatching perceptual models for effective watermarking in the presence of compression. In A. G. Tescher, editor, *Proc. SPIE, Multimedia Systems and Application II*, volume 3845, pages 29–42, September 1999.

[4] M. Ramkumar and A. N. Akansu. Theoretical capacity measures for data hiding in compressed images. In *Proc. SPIE, Voice, Video and Data Communications*, volume 3528, pages 482–492, November 1998.

[5] M. Ramkumar, A. N. Akansu, and A. Alatan. On the choice of transforms for data hiding in compressed video. In *IEEE ICASSP*, volume vi, pages 3049–3052, Phoenix, Arizona, March 1999.

[6] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp. The effect of matching watermark and compression transforms in compressed color images. In *Proc. IEEE Int. Conference in Image Processing*, volume 1, October 1998.