ESTIMATING STEGANOGRAPHIC CAPACITY FOR ODD-EVEN BASED EMBEDDING AND ITS USE IN INDIVIDUAL COMPENSATION

A. Sarkar and B. S. Manjunath

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

ABSTRACT

We present a method to compute the steganographic capacity for images, with odd-even based hiding in the quantized discrete cosine transform domain. The method has been generalized for varying orders of co-occurrence statistics for statistical restoration based steganography. We further utilize this capacity estimate to hide the maximum possible data per individual frequency stream, while ensuring that the first order histograms of individual frequency coefficients remain matched. We also show that certain frequency components are more useful for steganalysis after first order statistical restoration is performed for a certain band of select frequencies.

Index Terms— steganography, steganographic capacity, steganalysis, statistical restoration, individual compensation

1. INTRODUCTION

Steganography is the art of secure communication where the very existence of the communication cannot be detected while steganalysis is the art of detecting the presence of the secret communication. The steganographer has two conflicting requirements - he has to imperceptibly embed a certain amount of data in an innocuous looking host signal (the *cover*), and also ensure that there is minimal statistical difference between the cover and the *stego* (signal containing hidden data). The concept of ϵ-secure steganography was introduced by Cachin [1]. He proposed an information-theoretic model for steganography where security is assured if the relative entropy (Kullback-Leibler divergence) between the cover and stego is less than a predefined constant ϵ. Cachin's work thus provides a theoretical framework to define the steganographic security.

Fridrich et al [2] have defined "steganographic capacity" as follows - for a host signal, it is the maximal message length that can be embedded without producing perceptually or statistically detectable distortions. It generally depends on the hiding method. Chandramouli et al [3] have analyzed capacity estimation for Least Significant Bit based image steganography, where the cover is assumed to follow a zero mean Gaussian distribution.

In [4], we had obtained a secure hiding rate for the quantization index modulation (QIM) scheme [5], with the cover signals being generated from Gaussian distributions. The statistical restoration method that we had described in [6, 7] was used for steganography. The framework is general enough to be used for other hiding methods. In this paper, we consider odd-even based embedding in the block-based quantized discrete cosine transform (DCT) domain. We present the analysis for the optimum hiding fraction for the first order histogram matching case and then show its generalization for higher orders of co-occurrence statistics.

In [6, 7], we had presented a steganographic scheme where the first order probability mass function (PMF) of the block-based quantized DCT coefficients lying in a certain frequency band and with magnitudes less than a certain threshold was statistically restored. We call this method - total compensation. We have generalized this method to allow for compensation along individual frequency components, the hiding rates being such that individual 1-D PMFs are restored. We then show that after performing total compensation based steganography and using individual coefficients for steganalysis, detection becomes easier only for certain frequency coefficients.

2. PROBLEM FORMULATION

Let the input feature set available for hiding be X. We divide it into two disjoint sets - H for hiding and C for compensation, as in (1). We call the hiding fraction λ, which equals $\frac{|H|}{|X|}$, where $|E|$ denotes the cardinality of a given set E. Let the feature set obtained after hiding and compensation be Y, as in (1); the part available for hiding, H, is changed to $\hat{H}$ and the compensation part C is changed to $\hat{C}$ for histogram matching. We divide the feature set into bins and find their respective bin-counts (number of terms per bin). The normalized bin-count is regarded as the PMF. The aim is to find the maximum hiding fraction λ_{opt} (5), which maximizes $|H|$, subject to the constraint that enough terms are left for compensation so that the PMF of the feature set, before and after hiding, denoted by P_X and P_Y, respectively, remains the same. Let $B_X(i)$ denote the number of elements which gets mapped to the i^{th} bin of X.

$$X = H \cup C,\ Y = \hat{H} \cup \hat{C},\ H \cap C = \phi,\ \hat{H} \cap \hat{C} = \phi \quad (1)$$

$$\hat{H} \cap \hat{C} = \phi \Rightarrow B_Y(i) = B_{\hat{H}}(i) + B_{\hat{C}}(i),\ \forall\, i \quad (2)$$

$$\text{To obtain } P_Y = P_X, \text{ we need } B_Y(i) = B_X(i),\ \forall\, i \quad (3)$$

$$\Rightarrow B_{\hat{C}}(i) = \{B_X(i) - B_{\hat{H}}(i)\} \geq 0,\ \forall\, i \quad (4)$$

$$\lambda_{opt} = \underset{\lambda=\frac{|H|}{|X|}}{\arg\max} \{|H| = |\hat{H}| :\ B_X(i) - B_{\hat{H}}(i) \geq 0,\ \forall\, i\} \quad (5)$$

For a dataset X, $B_X(i)$ is known; after data hiding and changing H to $\hat{H}$, $B_{\hat{H}}(i)$ can be found - thus, $B_{\hat{C}}(i)$ can be computed using (4). As shown in (4), perfect restoration is possible only if the required number of terms in every bin of $\hat{C}$ is non-negative. As λ increases, the distance between the two PMFs P_X and P_Y increases and there are less number of terms available for compensation.

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

3. ODD-EVEN BASED HIDING FRAMEWORK

The luminance part of the image is used for hiding. We divide the luminance image into 8×8 blocks, perform block-wise DCT, divide element-wise by a certain quality factor matrix and then select a certain frequency band for hiding. The DCT coefficients thus selected are rounded off to produce the quantized DCT (QDCT) based dataset X. For hiding, we use odd/even embedding (a simple version of QIM) to convert the terms to their nearest odd or even integer, depending on whether the input bit is 1 or 0, respectively. Suppose, a QDCT term is 4 and we wish to embed 0 - then the QDCT term gets mapped to the nearest even number, which is 4. For embedding 1, we use a dither sequence, with numbers in the range [-0.5,0.5] which are produced by a pseudorandom generator, to decide whether to map 4 to 3 or 5.

$$\text{To embed } 1 \rightarrow q = \text{round}(p + 1 - \text{mod}(p - \delta, 2)), \quad (6)$$
$$\text{to embed } 0 \rightarrow q = \text{round}(p + 1 - \text{mod}(p + 1 - \delta, 2)) \quad (7)$$

where p, the original QDCT term, is mapped to q, δ denotes the corresponding number obtained from the dither sequence, "mod(p,2)" is the remainder obtained after dividing p by 2 and "round" denotes the rounding off operation. If p is an even(odd) number and 1(0) is to be embedded, it is mapped to $(p-1)$ or $(p+1)$ depending on whether δ belongs to the range (0,0.5] or [-0.5,0], respectively.

Let λ be the common hiding fraction for all bins. Let $X(i)$ and $\hat{H}(i)$ denote the elements mapped to the i^{th} bins of X and $\hat{H}$, respectively. Now, assuming an equal number of 0's and 1's in the input message that affects the elements in $X(i)$, $\frac{\lambda}{4}$ fraction of coefficients from $X(i)$ gets transferred to both $\hat{H}(i+1)$ and $\hat{H}(i-1)$. Also, $\frac{\lambda}{2}$ fraction of coefficients is moved to $\hat{H}(i)$. Explanation - let the value of the input QDCT coefficient be i, an even number, and if the input bit is 0, the output term, obtained using (7), is i itself. Since about half the bits in the input sequence are 0, about $\frac{\lambda}{2}$ terms in $X(i)$ are moved to $\hat{H}(i)$. If the input bit is 1, the output term gets mapped to the nearest odd number, which can be $(i-1)$ or $(i+1)$, depending on whether the dither value (δ in (6)) is positive or negative. By a similar logic, $\frac{\lambda}{4}$ fraction of terms from bins $X(i-1)$ and $X(i+1)$ will be shifted to $\hat{H}(i)$. Thus, based on this analysis, the number of terms in $\hat{H}(i)$ is as follows:

$$B_{\hat{H}}(i) \approx \frac{\lambda B_X(i)}{2} + \frac{\lambda B_X(i-1)}{4} + \frac{\lambda B_X(i+1)}{4} \quad (8)$$

To reiterate, the main assumptions behind this analysis are : the input message has equal number of 0's and 1's and the dither values are equally likely to be positive or negative. The assumptions are valid only if both the message and the dither sequence are long enough (minimum image size considered is 256×256). The goodness of this assumption is experimentally verified in Sec. 7.

4. COMPUTING THE OPTIMAL HIDING FRACTION AND RATE FOR 1-D HISTOGRAM BASED COMPENSATION

While computing the 1-D histograms for QDCT coefficients, we only consider those with magnitude less than a certain threshold T. Since the distribution of the QDCT coefficients is very peaky near 0 and falls off sharply for higher values, higher valued terms may be ignored in PMF estimation. For a given T, there are $(2T+1)$ bins from $[-T, T]$, and we optimally hide in all the bins, except the two extreme ones. For the $(-T)^{th}$ and T^{th} bins, perfect compensation may not be possible as we consider neighboring bins at one side only. From (4) and (8), considering the i^{th} bin, the hiding fraction λ needs to satisfy:

$$B_{\hat{H}}(i) \leq B_X(i) \Rightarrow \lambda \leq \left\{ \frac{B_X(i)}{\frac{B_X(i-1)}{4} + \frac{B_X(i)}{2} + \frac{B_X(i+1)}{4}} \right\} \quad (9)$$

For ease of notation, we define

$$\lambda_i = \left\{ \frac{B_X(i)}{\frac{B_X(i-1)}{4} + \frac{B_X(i)}{2} + \frac{B_X(i+1)}{4}} \right\} \quad (10)$$

It is to be noted that the whole analysis, especially, the expression for $B_{\hat{H}}(i)$ (8) as was derived in Sec. 3, assumed an equal hiding fraction for all the bins. For the i^{th} bin, λ_i can be viewed as $\frac{B_X(i)}{B_{\hat{H}}(i)}$ where $B_{\hat{H}}(i)$ is computed using a hiding fraction of unity. In Sec. 5, we shall be using this notation for $B_{\hat{H}}$ for the higher order cases. The effective hiding fraction $\lambda^\star(T)$, for a given T, is the minimum of all these λ_i terms (since the hiding fraction $\lambda \leq \lambda_i, \forall i$, using (9) and (10)).

$$\lambda^\star(T) \quad = \quad \min_{-T<i<T} \{\lambda_i : \lambda_i > 0\}. \quad (11)$$

The condition $(\lambda_i > 0)$ in (11) ensures that the hiding fraction will not be reduced to zero for bins with no elements. This may lead to PMF mismatches in bins with no elements before hiding but the mismatch is unlikely to be statistically and steganalytically significant, and hence not too useful for detection. Also, just as for the equal number of 0's and 1's assumption in Sec. 3, the experimental results in Sec. 7 indicate that it is a valid assumption for the first order histogram matching case.

Once we select a certain frequency band for hiding after performing block-wise DCT, the maximum fraction of the terms which can actually be used for hiding at a given threshold under the statistical restoration constraint is called the "rate" for that threshold. Let $G(T)$ denote the fraction of terms available for hiding at threshold T, while the hiding rate corresponding to a threshold T is $R(T)$.

$$G(T) \quad = \quad \sum_{-T<i<T} P_X(i) \quad (12)$$
$$R(T) \quad = \quad \lambda^\star(T).G(T) \quad (13)$$

where P_X is the PMF of X. As T increases, $G(T)$ increases while $\lambda^\star(T)$ decreases, since we are finding the minimum value over a larger set of T's (11). We vary the thresholds and select the threshold T_{opt} for which the rate is maximized.

$$T_{opt} = \arg\max_T R(T) \quad (14)$$

Thus, the maximum attainable rate for the QDCT based feature set using odd-even embedding and first-order compensation is $R(T_{opt})$, computed using (10)-(14).

5. EXTENSION FOR THE HIGHER ORDER STATISTICS

In the odd-even based hiding scheme, let us consider two coefficients at a time (2-D co-occurrence scenario). Let the two terms have values i and j respectively. If we call this pair as (i,j), then owing to an incoming bit, the new coefficient pair can be (i',j') where $i' \in \{i-1, i, i+1\}$ and $j' \in \{j-1, j, j+1\}$. We now obtain the optimum hiding fraction, given a certain threshold T, in a manner identical to the 1-D steganography case. Let $B_X(i,j)$ denote the bin-count in the $(i,j)^{th}$ bin of X.

$$B_{\hat{H}}(i,j) = \sum_{(i',j') \in D_8 \setminus D_4} \frac{B_X(i',j')}{16} + \sum_{(i',j') \in D_4} \frac{B_X(i',j')}{8} + \frac{B_X(i,j)}{4} \quad (15)$$

$$\lambda_{i,j}(T) = \frac{B_X(i,j)}{B_{\hat{H}}(i,j)} \quad (16)$$

$$\lambda^\star(T) = \min_{-T<i,j<T} \{\lambda_{i,j}(T) : \lambda_{i,j}(T) > 0\} \quad (17)$$

The $B_{\hat{H}}$ term in (15) is the bin-count for the $(i,j)^{th}$ bin of $\hat{H}$ computed using a hiding fraction of 1. Then, (16) and (17) are just the 2-D versions of (10) and (11), respectively. In (15), the set of the four nearest neighbors of the current 2-D point (i,j) is called D_4 while the set of D_4 and the four diagonal neighbors is called D_8.

We now provide a generalization for the n^{th} order co-occurrence statistic. A single bin will consist of n elements, say $(i_1, i_2, ..., i_n)$. Since we perform odd-even based hiding, the i_1 component can be mapped to i_1, $(i_1 - 1)$ or $(i_1 + 1)$ with probability $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{4}$, (valid under the same two assumptions as in Sec. 3) respectively. Thus, $(i_1, i_2, ..., i_n)$ can be mapped to $(i_1 + \delta_1, i_2 + \delta_2, ..., i_n + \delta_n)$, where $\delta_j \in \{-1, 0, 1\}, 1 \leq j \leq n$.

$$f(0) = \frac{1}{2}, f(1) = \frac{1}{4}, f(-1) = \frac{1}{4} \quad (18)$$

$$B_{\hat{H}}(i_1, i_2, ..., i_n) = \sum_{\delta_1} \sum_{\delta_2} ... \sum_{\delta_n} [f(\delta_1) f(\delta_2) ... f(\delta_n)] \times B_X(i_1 + \delta_1, i_2 + \delta_2, ..., i_n + \delta_n) \quad (19)$$

For the co-occurrence order n = 1 and 2 in (19), we compute $B_{\hat{H}}$ using (8) and (15), respectively. The optimal hiding fraction, $\lambda^\star(T)$ can be computed as in (16) and (17) for the 2-D case, by taking the ratio of the B_X and the $B_{\hat{H}}$ terms, and finding the minimum over a range specified by T. The hiding rate and optimal threshold estimates, $R(T)$ and T_{opt}, can then be obtained, using (13) and (14), respectively.

6. VARIATION OF OPTIMUM HIDING PARAMETERS WITH ORDER OF CO-OCCURRENCE STATISTICS

As we proceed from 1-D to 2-D co-occurrence statistic for the QDCT coefficients, the number of coefficients per bin decreases - the total number of coefficients remains the same but the number of bins in the 2-D case is the square of the number of bins in the 1-D case. Thus, there are many empty 2-D bins. We put a tolerance limit (p%) on the number of bins in X with zero elements. We gradually increased the threshold T from 1 till we found there were more than p% bins which had zero elements. Let the threshold corresponding to p% bins having zero elements be T_p. We now vary the threshold from 1 to T_p and find the threshold T_{opt} (14) at which the hiding rate $R(T)$ (13) is maximum. We use $p = 5$ in the experiments (Table 1).

For generating the QDCT terms for the luminance part of an image, we use a quality factor of 75. We consider the first 19 AC DCT coefficients that occur during zigzag scan, for a 8×8 block, for hiding and compensation. We limit the range of allowed threshold values to 30. For every image, after computing the QDCT terms X, $B_X(i)$ is computed for all the bins for a certain threshold T ($i \in [-T, T]$). The hiding fraction $\lambda^\star(T)$ is obtained using (11). The maximum attainable rate $R(T_{opt})$ is then computed using (14). We repeat this process for higher orders of co-occurrence statistics. In Table 1, we quantify the steganographic capacity in 3 ways: firstly, the maximum attainable rate $R(T_{opt})$, next, the bits hidden per pixel in the image and lastly, the total number of bits embedded in the image. The experiment is performed on 4500 images and the optimal hiding parameters, averaged over the entire set, are reported.

Table 1. Variation of the optimum hiding threshold, fraction and capacities with the order of co-occurrence, being averaged over 4500 images - since the threshold can assume only integer values, T_{opt}, after averaging, is changed to the nearest integer higher than it.

Order	T_{opt}	$\lambda^\star(T_{opt})$%	$R(T_{opt})$	Bits/pixel	Bits hidden ($\times 10^3$)
1	27	48.434	0.502	0.141	25.120
2	6	29.895	0.264	0.074	13.242
3	3	8.253	0.057	0.016	2.895

The maximum allowed hiding fraction expectedly decreases as we compensate for higher orders of co-occurrence. The amount of data that we need to hide decides the maximum order of co-occurrence upto which we need to compensate.

7. HIDING AND COMPENSATION - PER INDIVIDUAL QUANTIZED DCT COEFFICIENT

The steganographer may consider a certain band of QDCT terms for hiding in the "total compensation" procedure - mentioned in Sec. 1. The steganalyst is however free to choose a feature of his choice - e.g. he may consider first order histograms corresponding to each individual frequency coefficient stream. For each individual stream, there may be PMF mismatches between a cover and the corresponding stego image. However, machine-learning based steganalysis will be able to detect the hiding only if the mismatches are consistent enough across images. Here, for the total and individual compensation cases, we embed the maximum possible data while ensuring that first order restoration is still possible for the entire frequency band based histogram and for each individual frequency band based histogram, respectively, using (11).

We use 4500 images for the experiments - half for training and the other half for testing. Both the training and testing sets have half the images as cover and the other half as stego. During the training phase, we develop separate support vector machine (SVM) classifiers trained on each individual QDCT stream. The SVM classifiers are then used to distinguish between cover and stego images in the testing phase. After hiding, we compensate using firstly, the total compensation scheme and secondly, the individual compensation method. We hide data in those QDCT coefficients whose magnitude is less than 30. The QDCT features are generated using the same parameters (19 AC DCT terms) as in Sec. 6. We compute the probabilities of missed detection (P_{miss}) and false alarm (P_{FA}) after performing histogram-based steganalysis using the individual QDCT streams. For undetectable hiding, the total detection error $P_{total} = (P_{FA} + P_{miss})$ should be close to 1.

The P_{total} term is found to be close to 1 for all the 19 AC DCT streams for individual compensation - indicating undetectable hiding. The fact that the optimal hiding fraction based scheme turns out to be undetectable shows that the assumptions (mentioned in Sec. 3) based on which the hiding parameters were derived are practically justified. After performing total compensation and using individual QDCT streams for detection, we found that the P_{total} term varies for different QDCT streams. We compute the difference between the P_{total} values, averaged over all the test images, for the individual and total compensation cases, for each of the 19 streams. Let QDCT(i, j) denote the QDCT stream corresponding to the i^{th} row and j^{th} column of the 8×8 QDCT matrix ($1\leq\, i, j\, \leq 8$). The difference in P_{total} is significantly greater than 0 for certain streams (0.21-QDCT$(1, 2)$, 0.10-QDCT$(1, 3)$, 0.15-QDCT$(1, 5)$, 0.44-QDCT$(1, 6)$) while it is very close to 0 for the remaining 15 terms. In Fig. 1, a high difference, as in (a) and (b), indicates that individual stream based steganalysis is able to detect total compensation based hiding for these streams while a small difference, as in (c) and (d), suggests that explicit compensation is not needed for these streams to avoid detection. Thus, we observe that total compensation based hiding is most detectable when QDCT$(1, 6)$ is used for steganalysis.

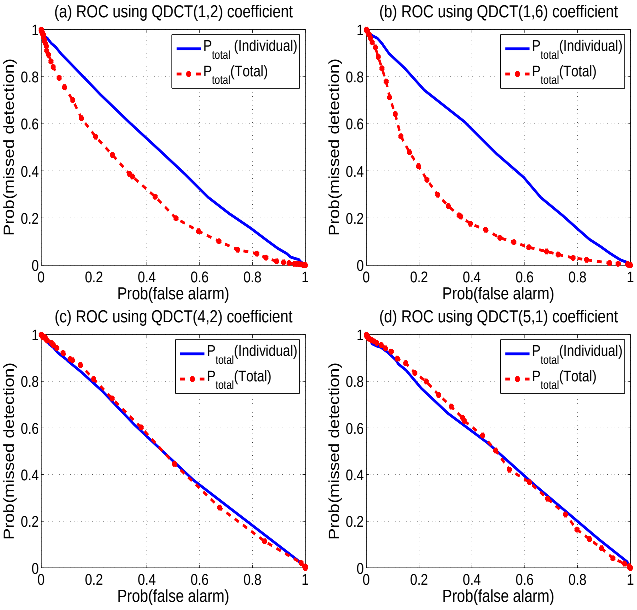

Fig. 1. Average detection error (P_{total}), averaged over all the test images, computed for different QDCT streams after statistical compensation: "Individual" and "Total" refer to the individual and total compensation schemes, respectively.

To explain the superior performance of QDCT$(1, 6)$ over other streams, we compute the PMF difference between the original image's QDCT stream and the data-embedded (and compensated) QDCT stream, for each individual stream and for both the compensation methods. We present examples of the average PMF difference (averaged over the test stego images) in Fig. 2. We observe that whenever the PMF difference is consistently high for the low magnitude bins, i.e. $\{-1, 0, 1\}$, it reflects in increased detection accuracy (Fig. 2(b)) - this occurs due to the peaky nature of the PMF near 0. While the peak PMF difference is as low as 3×10^{-3} for QDCT$(5, 1)$ (Fig. 2(a)), it is as high as 0.10 for QDCT$(1, 6)$ (Fig. 2(b)). The PMF of QDCT$(1, 6)$ is found to be much more peaky compared to that of other frequency streams. As the PMF of a certain frequency coefficient X becomes more peaky near 0, $B_X(0)$ becomes much greater than $B_X(1)$ and $B_X(-1)$. Using (10), the hiding fraction λ_i for $i = \{-1, 1\}$ becomes very small due to the dominance of $B_X(0)$. The effective hiding fraction $\lambda^{\star}(T)$ (11) for QDCT$(1, 6)$ will therefore be much smaller compared to other frequency streams, whose PMF is less peaky. When "total compensation" is performed, we consistently hide much more data in QDCT$(1, 6)$ than is permitted by $\lambda^{\star}(T)$ (11); hence, statistical restoration cannot compensate for the mismatched PMF and this leads to enhanced detection.

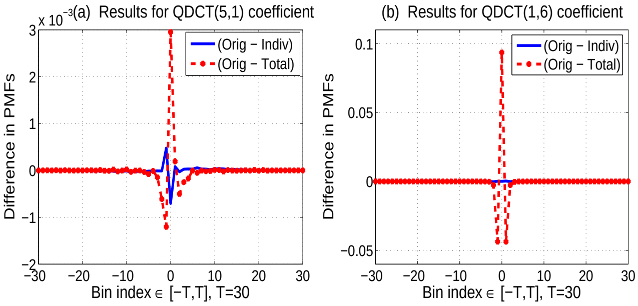

Fig. 2. Comparison of PMF differences, after statistical restoration, for individual and total compensation based methods, for different QDCT streams: here, "Orig", "Indiv" and "Total" refer to the original PMF, PMF after hiding and individual compensation, and PMF after hiding and total compensation, respectively.

8. CONCLUSION

Here, we have demonstrated a method to compute the maximum hiding fraction and hiding rate for odd-even based hiding for quantized DCT coefficients such that the hiding remains undetectable after first order statistical restoration. The method has been generalized for higher orders of co-occurrence statistics. From a steganalyst's perspective, we have looked at first order histograms of individual frequency streams. We have observed that a certain quantized DCT stream (pertaining to the 1^{st} row and 6^{th} column per 8×8 block) is particularly effective for first order steganalysis. There is a direct relationship between the peaky nature of the PMF of an individual quantized DCT stream and its usefulness in first order steganalysis.

References

[1] C. Cachin, "An information theoretic model for steganography," *LNCS: 2nd Int'l Workshop on Info. Hiding*, vol. 1525, pp. 306–318, 1998.

[2] J. Fridrich, M. Goljan, D. Hogea, and D. Soukal, "Quantitive steganalysis of digital images: Estimating the secret message length," *ACM Multimedia Systems Journal, Special issue on Multimedia Security*, vol. 9, no. 3, pp. 288–302, 2003.

[3] R. Chandramouli and N. Memon, "Analysis of LSB based image steganography techniques," in *Proc. ICIP*, Oct. 2001, pp. III–1019–22.

[4] K. Sullivan, K. Solanki, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Determing achievable rates for secure, zero-divergence, steganography," in *Proc. ICIP*, 2006, pp. 121–124.

[5] 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*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[6] K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Statistical restoration for robust and secure steganography," in *Proc. ICIP*, 2005, pp. II–1118–21.

[7] K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Provably secure steganography: Achieving zero K-L divergence using statistical restoration," in *Proc. ICIP*, 2006, pp. 125–128.