Steganalysis for Markov Cover Data With Applications to Images

Kenneth Sullivan, *Member, IEEE*, Upamanyu Madhow, *Fellow, IEEE*, Shivkumar Chandrasekaran, and B. S. Manjunath, *Fellow, IEEE*

***Abstract*—The difficult task of steganalysis, or the detection of the presence of hidden data, can be greatly aided by exploiting the correlations inherent in typical host or cover signals. In particular, several effective image steganalysis techniques are based on the strong interpixel dependencies exhibited by natural images. Thus, existing theoretical benchmarks based on independent and identically distributed (i.i.d.) models for the cover data underestimate attainable steganalysis performance and, hence, overestimate the security of the steganography technique used for hiding the data. In this paper, we investigate detection-theoretic performance benchmarks for steganalysis when the cover data are modeled as a Markov chain. The main application explored here is steganalysis of data hidden in images. While the Markov chain model does not completely capture the spatial dependencies, it provides an analytically tractable framework whose predictions are consistent with the performance of practical steganalysis algorithms that account for spatial dependencies. Numerical results are provided for image steganalysis of spread-spectrum and perturbed quantization data hiding.**

***Index Terms*—Data hiding, Markov chain, steganalysis, steganography.**

I. Introduction

RESEARCH in data hiding into multimedia objects, such as music, image, and video, has advanced considerably over the past decade [1]. Much of this work has been focused on protecting the ownership rights [2] of digital media. In addition, the use of digital data hiding for covert communication has a long history [3], [4]. As the state of the art of steganography progresses, there is increased interest in steganalysis, or detection of the presence of hidden data. A review of steganalysis [5], [6] shows many effective methods. In particular, while steganalysis is a difficult task, its performance can be greatly enhanced by exploiting the correlations inherent in typical multimedia host or cover data. For example, several effective image steganalysis techniques [7]–[12] are based on the strong interpixel dependencies exhibited by natural images. However, existing theoretical benchmarks for steganalysis are based on modeling the cover data as independent and identically distributed (i.i.d.) and, therefore, underestimate attainable steganalysis performance.

Our objective in this paper is to derive detection-theoretic performance benchmarks for steganalysis, accounting for dependencies in the cover data. It is important for both steganalysts and steganographers that such benchmarks are reasonably close to the performance attainable by practical steganalysis techniques. The benchmarks inform the steganalyst when the steganalysis algorithm being considered is "good enough" that no further effort needs to be expended in design, while they provide the steganographer with a measure of the security (in terms of resisting steganalysis) of the data hiding scheme employed. Existing theoretical benchmarks, such as Cachins's ϵ-secure measure [13], guarantee detector limits only if the cover data are i.i.d. [14] and, therefore, underestimate the attainable steganalysis performance. In this paper, we take the logical next step toward computing a more accurate performance benchmark, modeling the cover data as a Markov chain (MC). The Markov model has the advantage of analytical tractability, in that performance benchmarks governing detection performance can be characterized and computed explicitly. In our examples and numerical results, we focus on images as cover data, using a Markov model in which statistical dependency is limited to an adjacent pixel. Clearly, this model does not completely capture interpixel dependencies. However, we find that the performance benchmarks we compute are consistent with the performance of a number of image steganalysis techniques that exploit spatial correlations.

Our main results are as follows.

- We derive a detection-theoretic benchmark for steganalysis in sources with memory by employing an MC model for the statistics of the host signal. Specifically, a first-order Markov approximation is used to model interpixel dependencies in images, which are the focus of this paper. This benchmark gives to the steganographer a measure of security, and for the steganalyst gauges practical detection relative to a theoretical bound.
- We use this benchmark to gauge the inherent detectability of spread-spectrum (SS) and perturbed quantization hiding schemes. These estimates are found to be consistent with practical detection.
- As SS hiding is judged to be detectable, we devise a method to detect various flavors of SS hiding in real images with detection error rates as low as 4%.

Performance analysis incorporating dependencies in the host yields immediate benefits over current i.i.d. analysis [13], [15]–[18] for quantifying the security of various methods for hiding data in images. For example, although SS hiding can be detected with reasonable accuracy using current steganalysis

Manuscript received July 27, 2005; revised February 17, 2006. This work was supported in part by ONR Grants N0014-01-1-0380 and N0014-05-1-0816. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jessica Fridrich.

The authors are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: sullivak@ece.ucsb.edu).

Digital Object Identifier 10.1109/TIFS.2006.873595

1556-6013/$20.00 © 2006 IEEE

techniques [19]–[21], security tests derived from the i.i.d analysis determine that SS hiding is, in fact, safe from detection. MC analysis, on the other hand, correctly determines SS hiding to be at risk from steganalysis. MC analysis also provides guidance for steganographers seeking to evade steganalysis. Prior efforts in this direction have generally focused on matching the one-dimensional (1-D) histogram [17], [22], [23] or other specific steganalysis statistics [24], [25] which provides no guarantee that a future steganalysis scheme will not be able to detect the hiding by using a different statistic. Security measures including dependency have been considered previously by Chandramouli *et al.* [5], [26] but their γ_D security measure applies for a specific detector, and does not provide a performance estimate for other detectors. On the other hand, our analysis will predict optimal steganalysis based on an MC assumption. Although the MC model does not completely characterize image statistics, practical constraints on the ability of the steganalyst to estimate more complex statistical models have limited efforts to date and may continue to restrict the complexity of detectors.

The rest of the paper is organized as follows: In Section II, we review the detection theory approach to steganalysis to date and show how an MC model of images allows interpixel dependencies to be included in this approach. We then show how this model relates to current steganalysis. In Section III, we use this model to analyze SS and perturbed quantization hiding schemes and show how the MC analysis relates to practical findings. Finally, we present our conclusions in Section IV.

II. Steganalysis, Detection Theory, and Statistically Dependent Data

The general steganalysis problem is inherently difficult as little information is available to the steganalyst. The original cover image is not available and the steganographer can choose from a wide variety of data hiding methods with an array of parameters for each method, all with differing effects. However, due to the importance of the problem, many attempts have been made to solve the problem within a limited context; for example, for a given hiding scheme or a defined model of cover data. In addition to designing tools for detection, much theoretical analysis has been done, specifically applying hypothesis testing theory to the problem. Here, we review this approach and introduce our MC approach, which allows for the analysis of dependent covers.

A. Optimal Hypothesis Testing and Steganalysis

A natural approach to steganalysis is to model an image as a realization of a random process and leverage detection theory to determine optimal solutions and estimate performance. This approach has been widely used [13], [15]–[18], [27] to analyze steganalysis and guide detection efforts. The advantage of this model is the availability of results prescribing optimal (error minimizing) detection methods as well as providing estimates of the results of optimal detection. The essence of this approach is to determine which random process generated an unknown image under scrutiny. It is assumed that the statistics of cover images (also known as source or host images) are different than the statistics of a stego image, an image with data hidden in it. The statistics of a discrete valued random process are described by a probability mass function (PMF), from which the probability of any event can be evaluated. Given the PMFs for cover and stego images, detection theory describes the optimal test of the image under scrutiny to decide whether it is generated from the cover process or the stego process. An optimal detector will minimize the chance of choosing incorrectly. In the Neyman–Pearson sense, this means minimizing the probability of missed detection subject to a given probability of false alarm. For steganalysis, a missed detection is to declare an image under scrutiny to be a cover image when, in fact, it is stego. A false alarm is deciding stego when cover should have been chosen. If $\mathbf{y}$ is the received vector (e.g., the image under scrutiny), and $P_X(\cdot)$ and $P_S(\cdot)$ are the PMFs of cover and stego, respectively, the optimal detector is known to be the likelihood ratio test

$$\frac{P_X(\mathbf{y})}{P_S(\mathbf{y})} \underset{S}{\overset{X}{\gtrless}} \tau(\alpha)$$

where τ is a threshold chosen to achieve a set false alarm probability α.

Typically, for the steganalysis problem, it is assumed that the data samples (elements of $\mathbf{y}$) are i.i.d. Under this simplifying assumption, the probability of a received vector is the product of the marginal probabilities $P(\mathbf{y}) = \prod_{k=1}^{L} P(y_k)$. In this case, the likelihood ratio test is equivalent to choosing the hypothesis with the smallest Kullback–Leibler (K–L) divergence between an estimate of the received PMF and the hypothesis PMF [14], where the K–L divergence (sometimes called relative entropy) between two PMFs is given as

$$D(P_X||P_S) = \sum_{m\in\mathcal{Y}} P_X(m) \log \frac{P_X(m)}{P_S(m)} \tag{1}$$

where $\mathcal{Y}$ is the set of possible events m (e.g., pixel values). The estimate of the received PMF is a normalized histogram (or type) formed by counting the number of occurrences of different events (pixel values, transform coefficients, etc.) and dividing by the total number of samples. Therefore, the K-L divergence is a measure of "closeness" of histograms in a way that is compatible with optimal hypothesis testing. Of greater interest than providing an alternative expression to the likelihood ratio test, the error probabilities for an optimal hypothesis test decrease exponentially as the K–L divergence between the two hypothesis PMFs increases [14]. In other words, the K–L divergence provides a convenient means of gauging how easy it is to discriminate between cover and stego. Because of this property, Cachin suggested [13] using the K–L divergence as a benchmark of the inherent detectability of a steganographic system.

Typical cover data, however, are not i.i.d. For example, both pixels and audio samples are known to be highly correlated. To more accurately characterize optimal hypothesis testing in steganalysis, a model employing dependency must be used.

B. Detection-Theoretic Divergence Measure for Markov Chains

To include interpixel dependency in our analysis, we employ an MC [28] model of image data. A MC is a random sequence indexed by n, subject to the following condition: $P(Y_n|Y_{n-1},Y_{n-2},\ldots,Y_1)=P(Y_n|Y_{n-1})$. Under this model, the probability of a given pixel is dependent on an immediately adjacent pixel. We have used this model to analyze and detect SS hiding [21], and Sidorov used MC and Markov random field analysis for detecting least-significant bit (LSB) hiding [29]. There are a number of reasons to use an MC model. First, the MC model accounts for dependency, yet it is very general and flexible. Second, while an MC model is more complex than an i.i.d model, it is the least complex model incorporating dependencies. Though many have used Markov random fields [30] to model images accounting for a larger neighborhood of dependency than one adjacent pixel, for the steganalyst, there is a practical drawback to increasing the levels of dependency. As the model complexity increases, the number of samples required to make an accurate estimate of the statistics also increases. However, the number of received samples depends on the image size and cannot be increased by the steganalyst. Thus, although the complexity for the steganalyst increases quickly, the benefit does not. For more on the difficulties of multivariate density estimation, see [31]. The MC model, on the other hand, is simple enough to make realistic statistical estimates. This is analogous to an nth order DPCM coding system, in which the benefit of an increase in n, the number of pixels used for prediction, has been shown to quickly drop after 2 or 3 [32]. Finally, a divergence metric, which measures the performance of optimal detection, analogous to the K–L divergence for i.i.d sources, exists [33] for MCs and is examined below.

First, we clarify our notation. Let $\{Y_n, n=1,2,\ldots,L\}$ be an MC on the finite set $\mathcal{Y}$. In our context, Y_n are the n-indexed set of pixels obtained by a row or column scanning and $\mathcal{Y}$ are all possible gray scale values (e.g., for an 8-b image $\mathcal{Y}=\{0,1,\ldots,255\}$). A MC source is defined by a transition matrix $\mathbf{T}_{ij}\triangleq P(Y_n=i|Y_{n-1}=j)$, and marginal probabilities $p_i\triangleq P(Y_n=i)$. For a realization $\mathbf{y}=(y_1,y_2,\ldots,y_L)^T$, let $\eta_{ij}(\mathbf{y})$ be the number of transitions from value i to value j in $\mathbf{y}$. The empirical matrix is $\mathbf{M}(\mathbf{y})\triangleq(\eta_{ij}(\mathbf{y})/(L-1))$. That is, the i,jth element represents the proportion of spatially adjacent pixel pairs with a grayscale value of i followed by j and, therefore, provides an estimate of the probability $P(Y_n=i,Y_{n-1}=j)$. The empirical matrix thus provides an estimate of the transition matrix and marginal probabilities $\mathbf{T}_{ij}=P(Y_n=i,Y_{n-1}=j)/P(Y_{n-1}=j)$; $P(Y_{n-1}=i)=P(Y_n=i)=\sum_j P(Y_n=i,Y_{n-1}=j)$. The empirical matrix, similar to the cooccurrence matrix (see citations in [34]), can be recognized as a matrix form of the two-dimensional (2-D) normalized histogram (or type) used to estimate the joint PMF of an arbitrary source. Intuitively, for sources that are strongly correlated, such as pixels, we expect the probability of two adjacent samples having equal, or nearly equal, value to be high. Therefore, in the empirical matrix, we expect the mass to be more concentrated near the main diagonal (all elements such that $i=j$) in a correlated source then we would expect for an i.i.d source; see the examples in Fig. 1.

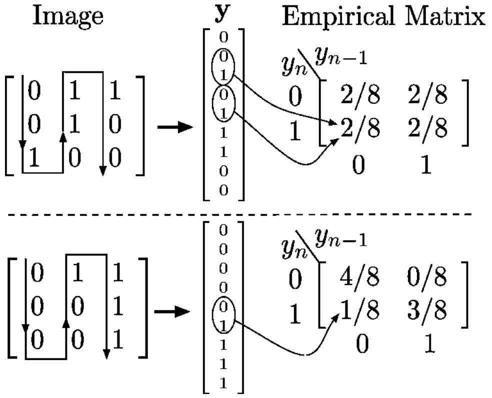


Fig. 1. Simple example of empirical matrices, here we have two binary (i.e., $\mathcal{Y}=\{0,1\}$) 3×3 images. From each image, a vector is created by scanning, and an empirical matrix is computed. The top image has no obvious interpixel dependence, which is reflected in a uniform empirical matrix. The second image has dependency between pixels, as seen in the homogenous regions and so its empirical matrix has a probability concentrated along the main diagonal. Though, in this contrived example, the method of scanning (horizontal, vertical, zig-zag) has a large effect on the empirical matrix, we find the effect of the scanning method on real images to be small.

The divergence measure we employ to quantify the statistical changes introduced by steganography is essentially a distance between the empirical matrices $\mathbf{M}^{(X)}$ and $\mathbf{M}^{(S)}$ of the two hypotheses, cover and stego

$$D\left(\mathbf{M}^{(X)},\mathbf{M}^{(S)}\right)=\sum_{i,j\in\mathcal{Y}}\mathbf{M}_{ij}^{(X)}\log\left(\frac{\mathbf{M}_{ij}^{(X)}}{\sum_j\mathbf{M}_{ij}^{(X)}}\frac{\sum_j\mathbf{M}_{ij}^{(S)}}{\mathbf{M}_{ij}^{(S)}}\right). \tag{2}$$

This divergence has many useful properties for the study of steganalysis in sources with memory, from the point of view of both the steganographer and the steganalyst.

For a constant false alarm rate, the minimal achievable missed detection rate approaches $e^{-LD(\mathbf{M}^{(X)},\mathbf{M}^{(S)})}$ as L, the number of samples, goes to infinity [33], [35], just as in the i.i.d case with K–L divergence. In other words, under the assumption of an MC model, the performance of the best possible steganalysis is exponentially bounded by this measure.

It can be seen then that $D(\mathbf{M}^{(X)},\mathbf{M}^{(S)})$ provides a measure to the steganographer of the inherent detectability of a steganographic scheme, given an assumption on the complexity of the detector. Other similar measures have been proposed. Cachin suggested [13] ϵ-secure steganography, in which the acceptable K–L divergence between cover and stego marginal PMFs is bounded. The advantage of using a bound on the matrix divergence is the addition of dependency to the model. In other words, if the detector, in fact, uses dependency, an ϵ-secure hiding scheme will overestimate the secrecy of hiding. To prevent this problem, Chandramouli *et al.*, [5], [26] suggest the $\gamma_{\mathcal{D}}$ metric. Here, the measure of detectability of a steganography method is a bound on the allowed probabilities of false alarm and missed detection for a given detector $\mathcal{D}$. While this certainly avoids the problem of underestimating the power of detectors employing dependency, it is only valid with respect to a given detector. If a different detector is employed, or invented, the security is unknown. The matrix divergence, however, bounds the false alarm and missed detection probabilities of the best possible detector using one level of dependency. The detector does not have to be known or even exist. In practice, the steganographer can choose a scheme that minimizes the divergence for a given cover joint distribution model (e.g., Gaussian, Laplacian, etc.). Alternatively, given a scheme, the steganographer can choose to use only images that exhibit a small divergence after hiding.

For the steganalyst, $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$ measures the amount of information gained for each additional sample received, just as with the K–L divergence for independent samples. The detector can use this to decide if there is enough gain to justify using a more complex detector. We note that $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$ is equal to the K–L divergence if the samples are indeed independent $(\mathbf{M}_{ij} = p_i p_j)$

$$\begin{aligned}
&\sum_{i,j\in\mathcal{Y}} \mathbf{M}_{ij}^{(X)} \log\left(\frac{\mathbf{M}_{ij}^{(X)}}{\sum_j \mathbf{M}_{ij}^{(X)}} \frac{\sum_j \mathbf{M}_{ij}^{(S)}}{\mathbf{M}_{ij}^{(S)}}\right)\\
&\quad= \sum_{i,j} \mathbf{M}_{ij}^{(X)} \log\left(\frac{\mathbf{M}_{ij}^{(X)}}{p_i^{(X)}} \frac{p_i^{(S)}}{\mathbf{M}_{ij}^{(S)}}\right)\\
&\quad= \sum_{i,j} p_i^{(X)} p_j^{(X)} \log\left(\frac{p_i^{(X)} p_j^{(X)}}{p_i^{(X)}} \frac{p_i^{(S)}}{p_i^{(S)} p_j^{(S)}}\right)\\
&\quad= \sum_i p_i^{(X)} \sum_j p_j^{(X)} \log \frac{p_j^{(X)}}{p_j^{(S)}}\\
&\quad= \sum_j p_j^{(X)} \log \frac{p_j^{(X)}}{p_j^{(S)}}.
\end{aligned}$$

Let R be the ratio of the matrix divergence measure to the K–L divergence. R represents the gain of employing the more complex model. For example, to achieve the same detector power (i.e., same probabilities of miss and false alarms) requires R times as many samples if the detector uses an i.i.d cover model versus an MC model. In the case of independent data, R will be one, reflecting the lack of gain of a detector that uses statistics beyond a 1-D histogram.

C. Relation to Existing Steganalysis Methods

As mentioned above, using the MC model is analogous to assuming a complexity constraint on the detector. Since dependency is limited to one adjacent pixel, empirical matrices provide sufficient statistics for optimal detection. However, even 2-D joint statistics are difficult to use practically. For practical applications, it is useful to use a subset, or function, of the empirical matrix. Often, these subsets or functions are chosen to match a specific hiding scheme and, if done correctly, do not sacrifice much detection power. However, they certainly cannot improve detection. We now show that many ongoing efforts in steganalysis use such a subset or function.

Many steganalysis schemes and analysis [16], [18], [36]–[38] use a histogram, or estimate the 1-D PMF, to discriminate between cover and stego. A 1-D histogram is simply the row sums of the empirical matrix $P(i) = \sum_j \mathbf{M}_{ij}$.

To capture the effect of hiding on interpixel dependencies, some [9], [39] have used difference histograms, that is, instead of a histogram of sample values, a histogram of the difference of values between samples is used. As pixels are strongly correlated, the difference between pixels is small and the histogram is concentrated toward zero. Typically, hiding disrupts this concentration and, with appropriate calibration, the hiding can be detected. The difference histogram is formed by the sums of the diagonals of the empirical matrix. That is, the difference histogram is $P(x) = \sum_i \sum_{i-j=x} \mathbf{M}_{ij}$. The concentration at zero in the difference histogram corresponds to the concentration along the main diagonal of the 2-D histogram.

To detect LSB hiding, the RS scheme (so named because of the use of sets called regular and singular) [40] and related sample-pair analysis [8] also use counts of differences between pixel values. Though sample-pair analysis is not limited to adjacent pixels, the authors note the estimate is improved in practice for spatially adjacent samples. In [10], Roue *et al.* use the empirical matrix directly to improve the effectiveness of sample-pair analysis.

Also for LSB detection, Sidorov, explicitly using an MC model [29], [41], uses an entropy-like measure based on ratios of values near the main diagonal of the empirical matrix

$$\begin{aligned}
\lambda = -2\sum_{i=2} \Bigg(&\mathbf{M}_{i,i-1} \log \frac{\mathbf{M}_{i,i-1}\mathbf{M}_{i,i+1}}{2\mathbf{M}_{i,i-1}}\\
&+\mathbf{M}_{i,i+1} \log \frac{\mathbf{M}_{i,i-1}+\mathbf{M}_{i,i+1}}{2\mathbf{M}_{i,i+1}}\Bigg).
\end{aligned}$$

Recently, Ker [42], [43] has approached the detection of LSB matching, a variant of standard LSB hiding that is not detected by standard LSB steganalysis. To improve the results of a detection method introduced by Harmsen and Pearlman [19] that employs a histogram, an adjacency histogram is used instead. The use of the adjacency histogram, which is exactly equivalent to the empirical matrix, substantially improves the detection results.

In [11], Fridrich *et al.* use a calibrated blockiness measure to detect Outguess 0.2b [22]. This blockiness measure is the expected value of the absolute difference of border pairs and can be rewritten in terms of the empirical matrix generated from pixels straddling 8 × 8 block boundaries

$$B = \sum_{x=0,1,\ldots} x \left(\sum_i \sum_{|i-j|=x} \mathbf{M}_{ij}\right).$$

For blind detection, in which a hiding scheme is not assumed, Fridrich *et al.* [44], [45] use a combination of features, histograms, and co-occurrence matrices [of bands of discrete cosine transform (DCT) coefficients]. The co-occurrence matrix is essentially the same as the empirical matrix. Though our experiments focus on the joint statistics of pixels rather than DCT coefficients, the analysis is generic to any Markov data. Though the DCT is known to significantly reduce dependencies, it does not create completely independent data. The effectiveness of their detection shows that steganalysis can be improved by including these dependencies.

In [46], Ambalavanan *et al.* use a Markov random field (MRF) to study active steganalysis, that is, extracting the message from a known stego image. As we note in Section II-B, the MRF is an extension to the MC model, including more dependency. Though the bit extraction works well at low hiding rates, the increased complexity of the model has its drawbacks as can be seen in the deterioration of results for higher rates of embedding.

In [12], Avcibas *et al.* use image-quality metrics to measure the effect of hiding. Though these metrics are not easily related to the empirical matrix, it is notable that the metrics are evaluated between the image under scrutiny and a low-pass-filtered version of the image. To generate the filtered image, each pixel is replaced with a weighted sum of a 3 × 3 neighborhood surrounding the pixel. In other words, it is assumed that the measurable difference between a given image and the same image with artificially enhanced interpixel dependencies is different for stego images and cover images.

We have argued that analysis using an MC model provides meaningful results under the condition of a steganalyst incorporating one level of dependency for detection. We have seen here that many existing detection methods indeed implicitly employ such a model.

III. Measured Divergence of Steganographic Schemes

Due to the lack of information available to the steganalyst, practical detection is inevitably suboptimal. However, we still expect some relationship between the calculated divergence and the efficacy of state-of-the-art steganalysis. We now examine the divergence measure of some existing steganographic schemes and compare them with current detection methods to test this assumption. We focus on interpixel correlation and always perform measurements in the spatial domain. Additionally, we compare the calculated divergence under an assumption of independence (1) to the divergence assuming dependency (2) to evaluate the value to the steganalyst in incorporating a more complex statistical model.

A. Spread Spectrum

SS data hiding [47] is an established embedding method, often used for watermarking, but also applicable for steganography [48]. Here, we measure and study the statistical effect of hiding on the empirical matrix and relate this to detection experiments we performed.

1) Measuring Detectability of Hiding: In SS data hiding, the message data modulates a noise sequence to create a message-bearing signal, which is then added to the cover data. Since its introduction, many variants of SS have been proposed, typically in the context of watermarking. The major goal in watermarking is robustness to malicious attacks, rather than statistical invisibility. We therefore focus on basic models of hiding suggested by Cox *et al.* [47], shown here for reference. Let $\{D_k, k \geq 0\}$ be a zero mean, unit variance, Gaussian message bearing signal, and $\{X_k, \geq 0\}$ be the cover samples. Two methods of generating stego data S_k are

$$S_k = X_k + \alpha D_k \tag{3a}$$

$$S_k = X_k + (\alpha X_k) D_k = X_k(1 + \alpha D_k) \tag{3b}$$

where α is a scaling parameter used to adjust the hiding power. This adjustment allows the data hider to adapt the hiding to the cover in order to control the perceptual distortion, the robustness of the message, and security from steganalysis. In the first method, the adaptation is done globally (i.e., a constant hiding power is used for all cover samples). We refer to this as globally adaptive hiding. In the second method, the hiding power adapts to each cover sample, so we characterize this as locally adaptive. We also note that often the cover image is transformed before data are hidden; for example, Cox *et al.* [47] use a whole image DCT. We measure the divergence of four variants of SS hiding: local and globally adaptive hiding in both the spatial and DCT domains. We have seen globally adaptive spatial SS hiding by Marvel *et al.* in SS image steganography (SSIS) [48] and, more recently, by Fridrich *et al.* in a variant of stochastic modulation [49]. The latter allows for a higher number of bits to be successfully decoded and the embedding rate is a function of the message signal power. The experiments presented by Cox *et al.* [47] are locally adaptive DCT hiding.

For each variant, we calculated the divergence over a range of message signal power. For globally adaptive hiding, we hold the message-to-cover power ratio (MCR) constant. In the locally adaptive case, we assume the sender and receiver have a shared constant scale factor α. The MCR will vary from image to image; we record the average value with the data.

To generalize our approach, we would like to simplify the divergence measurement by eliminating the need to derive the stego empirical matrix. Instead of using statistical analysis of the hiding scheme to generate an expected stego empirical matrix, Monte Carlo simulations of data hiding in several images may provide an accurate means of estimating divergence. To do this, several synthetic images are generated from the empirical matrix of a cover image. Data are hidden in these synthetic images and stego empirical matrices are calculated from the resulting images. The average divergence between these empirical matrices and the original cover matrix represent an estimate of the divergence introduced by hiding.

We note here an important practical consideration when using empirical matrices. The matrix divergence results we are using require absolute continuity of $\mathbf{M}^{(X)}$ with respect to $\mathbf{M}^{(S)}(\mathbf{M}^{(X)} \ll \mathbf{M}^{(S)})$ [35]. That is, $\mathbf{M}^{(X)}$ cannot be nonzero at any point where $\mathbf{M}^{(S)}$ is zero. We will see that for the steganographic methods we look at, the distribution of stego coefficients tends to be more spread out. Generally then, $\mathbf{M}^{(S)}$ is more likely to have values in bins where $\mathbf{M}^{(X)}$ does not, rather than the opposite. However, this is not guaranteed and typically there is a nominal violation of absolute continuity, that is, some percentage of $\mathbf{M}^{(X)}$ is nonzero where $\mathbf{M}^{(S)}$ is zero. For our estimate of the divergence introduced by hiding, if the violation is negligible, we use an approximate form of $\mathbf{M}^{(X)}$ in which nonzero values violating the constraint are set to zero and the matrix is re-normalized. Generally, experiments that produce non-negligible violations (>0.1%) are pathological cases of little interest. For example, in these cases, the experimental setup in practice may yield grossly distorted images. These experiments are not reported here.

As mentioned in Section II-B, the divergence measure provides to the steganographer a benchmark of the inherent detectability of hiding. Additionally, it allows the steganalyst to compare the information gained from each new sample by ex-

TABLE I
DIVERGENCE MEASUREMENTS OF SPREAD-SPECTRUM HIDING (ALL DIVERGENCE VALUES ARE MULTIPLIED BY 100). AS EXPECTED, THE EFFECT OF TRANSFORM AND SPATIAL HIDING IS SIMILAR. THERE IS A CLEAR GAIN HERE FOR THE DETECTOR TO USE DEPENDENCY. A FACTOR OF 20 MEANS THE DETECTOR CAN USE 95% LESS SAMPLES TO ACHIEVE THE SAME DETECTION RATES

Globally adaptive, Spatial			
MCR	-23	-20	-17
Mean $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$	30.20	36.82	43.43
Mean $D(p^{(X)} \| p^{(S)})$	2.48	3.27	4.10
Mean ratio	24.23	22.24	20.57

Globally adaptive, DCT			
MCR	-23	-20	-17
Mean $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$	30.38	36.98	43.46
Mean $D(p^{(X)} \| p^{(S)})$	2.49	3.26	4.06
Mean ratio	24.45	22.24	20.56

TABLE II
FOR SS LOCALLY ADAPTIVE HIDING, THE CALCULATED DIVERGENCE IS RELATED TO THE COVER MEDIUM, WITH DCT HIDING BEING MUCH LOWER. ADDITIONALLY, THE DETECTOR GAIN IS SMALLER FOR DCT HIDING

Locally adaptive, Spatial			
α	0.375	0.05	0.1
Mean MCR	-22.74	-20.33	-14.63
Mean $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$	27.92	32.17	42.88
Mean $D(p^{(X)} \| p^{(S)})$	1.48	1.93	3.34
Mean ratio	33.49	30.31	22.45

Locally adaptive, DCT			
α	0.375	0.05	0.1
Mean MCR	-28.52	-26.13	-20.19
Mean $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$	5.61	5.87	6.52
Mean $D(p^{(X)} \| p^{(S)})$	1.39	1.78	3.41
Mean ratio	13.10	10.21	5.90

ploiting dependency to the information gained using only first-order statistics. We present both divergence measures: between empirical matrices (2) and between marginal histograms (1), and the average ratio of these two to show the gain by using dependency at the detector. These measurements are summarized in Tables I and II. From stochastic modulation [49] (a variant of globally adaptive SS hiding), we have a means of relating message signal power to a realizable embedding rate. The average hiding rates for MCRs -23, -20, and -17 are 0.91, 0.94, and 0.96 bits per pixel (bpp), respectively.

From these data, we can see many trends. Not surprisingly, the divergence measure always increases with the MCR; the more powerful a message (and, subsequently, a higher hiding rate), the more obvious the hiding becomes. Additionally, though the measured divergence introduced by globally adaptive hiding is roughly the same for both spatial hiding and transform hiding, locally adaptive divergence changes depending on the hiding domain. Locally adaptive spatial hiding is slightly less divergent than globally adaptive hiding (for similar MCR), however, locally adaptive DCT is much less divergent. We expect from these divergence measurements that detection will be more difficult for locally adaptive hiding, particularly DCT, than for the other cases. Finally, in all cases, there is an advantage to including dependencies in detection. In the best case, about 95% fewer samples can be used to achieve the same performance. Even in locally adaptive DCT hiding, where the advantage is the least, a gain of 5.9 means only about a sixth of the samples are required. Below, we analyze the underlying statistical changes caused by hiding in order to explain these findings.

2) Statistical Effect of SS Hiding: Globally adaptive hiding is analogous to inserting zero mean additive white Gaussian noise (AWGN) with power α^2. The net statistical effect is a convolution of the message signal distribution $\mathcal{N}(0, \alpha^2)$, with the cover distribution [19]. Deriving the exact empirical matrix of the stego signal is complicated somewhat by the necessity of quantization and clipping as a final step to return to the same sample space $\mathcal{Y}$ as the source. For example, in hiding in pixels, the stego values must be rounded to integers between 0 and 255. When necessary to prevent ambiguity, we delineate the unquantized stego signal as S'. The probability density function (pdf) of the stego signal before quantization is

$$\begin{aligned} f_{S'}\left(s'_1, s'_2\right) =& \iint \Bigg(\sum_k \sum_l \mathbf{M}^{(X)}_{kl} \delta(t_1 - k, t_2 - l) \frac{1}{2\pi\alpha^2} \exp \\ & - \left\{ \frac{(s'_1 - t_1)^2 + (s'_2 - t_2)^2}{2\alpha^2} \right\} dt_1 dt_2 \Bigg) \\ =& \frac{1}{2\pi\alpha^2} \Bigg(\sum_k \sum_l \mathbf{M}^{(X)}_{kl} \exp \\ & - \left\{ \frac{(s'_1 - k)^2 + (s'_2 - l)^2}{2\alpha^2} \right\} \Bigg). \end{aligned}$$

In other words, there is a white joint Gaussian pdf centered at each point in the cover empirical matrix and scaled by the empirical matrix value. This can be seen as a blur of the cover empirical matrix and is directly analogous to spatial lowpass filtering of images by convolution with a Gaussian function [50, Sec. 4.3.2]. After rounding to pixel values, the empirical matrix of the stego signal is

$$\mathbf{M}^{(S)}_{ij} = \int_{i-1/2}^{i+1/2} \int_{j-1/2}^{j+1/2} f_{S'}\left(s'_1, s'_2\right) ds'_1 ds'_2. \tag{4}$$

Since the message signal is white, or uncorrelated, its empirical matrix is spread evenly and there is no greater probability for values near the main diagonal. Hiding weakens the dependencies between the cover samples, which causes spreading from the main diagonal of the empirical matrix, as seen in Fig. 2.

Locally adaptive hiding can also be viewed as zero mean AWGN; however, it is nonstationary, since the noise power $(\alpha X_k)^2$ depends on X_k. Instead, we view it as multiplicative noise with mean of one. Let $B_k \sim \mathcal{N}(1, \alpha^2)$ be a multiplicative

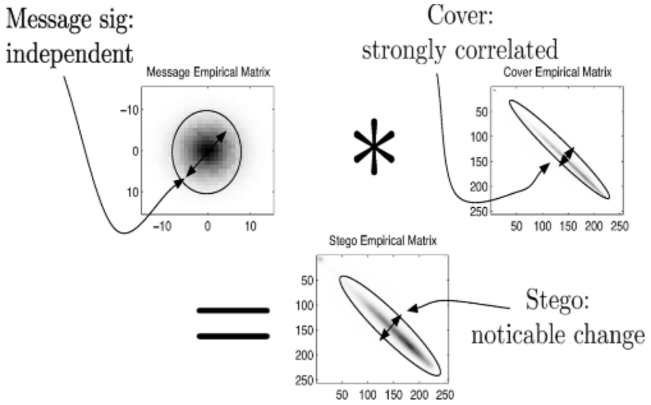


Fig. 2. Empirical matrices of SS globally adaptive hiding. The convolution of a white Gaussian empirical matrix (bell shaped) with an image empirical matrix (concentrated at the main diagonal) results in a new stego matrix less concentrated along the main diagonal. In other words, the hiding weakens dependencies.

message signal, then $S_k = X_k B_k$, the cumulative distribution function (cdf) of (prequantized) S is

$$F_{S_1',S_2'}(s_1',s_2')=\int_0^\infty\int_0^\infty\left[\sum_{i=0}^{\min\left(\lfloor s_1'/b_1\rfloor,255\right)}\sum_{j=0}^{\min\left(\lfloor s_2'/b_2\rfloor,255\right)}\mathbf{M}_{ij}^{(X)}\right]\times\exp-\left\{\frac{(b_1-1)^2+(b_2-1)^2}{2\alpha^2}\right\}db_1db_1$$

and the pdf is

$$f_{S_1',S_2'}(s_1',s_2')=\frac{\partial^2 F_{S_1',S_2'}(s_1',s_2')}{\partial s_1'\partial s_2'}$$

and the empirical matrix of the quantized S can be found with (4). To simplify the expressions, we have assumed α is such that the probability of b_1 and b_2 at values less than zero are negligible. This assumption is warranted by the typical α values used in hiding, which are chosen to be small enough to avoid visual distortion. For a given cover empirical matrix $\mathbf{M}^{(X)}$, these expressions can be evaluated numerically.

From the equations, the statistical effect is not obvious. However, as seen in Fig. 3, hiding still blurs the cover matrix, shifting probability away from the main diagonal. The effect, however, is less strong.

We can now summarize the statistical effect of SS hiding and relate this to our findings in Section III-A-1. In a general sense, SS data hiding adds an i.i.d. message signal to a non-i.i.d cover. It is not surprising then that the statistical effect is a decrease in the dependence of the cover. For globally adaptive hiding, this effect is very clearly seen in a shift of probability away from the main diagonal. For locally adaptive hiding, the adaptation causes the additive message sequence to become dependent on the cover. Effectively, the message sequence is de-whitened, that is, correlations are introduced and the effect is weakened. This can be seen to explain the smaller divergence measured in for locally adaptive hiding compared to global.

For the linear transformations typically used, such as DCT and discrete Fourier transform (DFT), the addition of a Gaussian message signal in the transform domain is equivalent to adding a Gaussian message signal in the spatial domain. Therefore, globally adaptive hiding in DCT coefficients statistically has the same effect as hiding in pixels. However, locally adaptive hiding, which can be seen as a multiplicative Gaussian signal, is not equivalent in both domains. This helps explain why the calculated divergence was nearly equal for globally adaptive hiding in either domain, but differed greatly for locally adaptive hiding.

Finally, we found the most noticeable effect of SS hiding to be spreading from the main diagonal of the empirical matrix. Since the histogram is simply the collection of sums of each row of the empirical matrix, this effect will be missed by studying only marginal statistics. That is, the spreading along each row will not be visible when the row is collapsed into a single point. This explains the gain of using dependency in detection.

3) Experiments: We now compare the measurements of the detectability of optimal SS hiding to experiments using a practical detector and find that the practical experiments follow the estimates above. We also compare experiments for a detector using dependencies with a simpler detector to judge the expected gain in detection.

To achieve optimal detection of data hiding, the detection-theoretic prescription is to calculate the empirical matrix of a suspected image and calculate the divergence between this and the empirical matrices of both the cover and the stego. Whichever is "closer" (i.e., has a smaller divergence measurement) is the optimal decision. From the analysis above, we can evaluate the stego empirical matrix given by the cover matrix. The cover statistics, however, will not be known in a practical scenario. To overcome this, we can attempt to estimate the cover statistics for each received image or estimate the cover statistics for all images. Some steganalysis has been able to estimate the statistics on an image by image basis [16], [37]; however, there is no general prescription for making such an estimate. There has also been some success with classifying between the set of all cover images and all stego images, typically through the use of supervised learning techniques [19], [38], [51], [52]. The idea is to train a machine with several examples of both cover and stego to discriminate between the two classes. For our experiments, we choose to employ supervised learning.

For the experiments, we need an image database, a learning algorithm, and a feature vector to train the machine. In the image database, we want to represent the vast variety of real images as well as possible. We use an image set comprised of a mix of four separate sources:

1) digital camera images, partitioned into smaller subimages;
2) scanned photographs;
3) scanned, downsampled, and cropped photographs;
4) images from the Corel volume scenic sites.

All images are converted losslessly to PNG format and color images are converted to grayscale. The entire database contains approximately 1400 images. Half of these are used for training and half for testing. Within both the training and testing sets, half are cover images and half are (distinct) stego images. And so there are four sets of distinct images (no image is in two sets): cover training, stego training, cover testing, and stego testing.

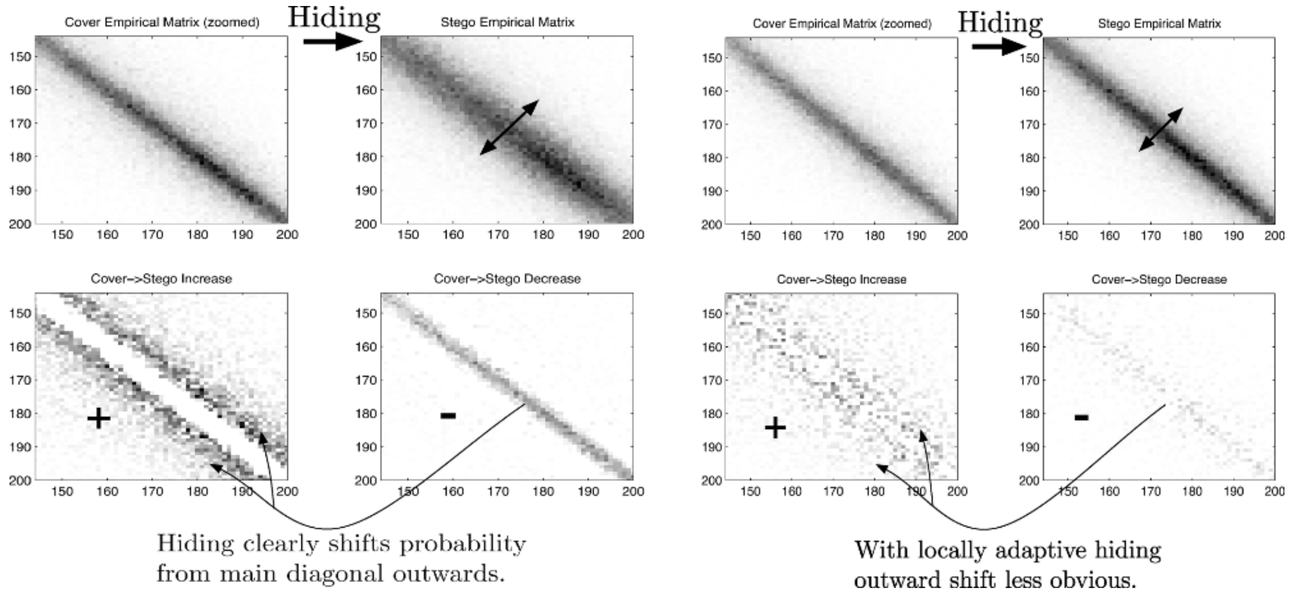


Fig. 3. Global (left) and local (right) hiding both have similar effects, a weakening of dependencies as seen as a shift out from the main diagonal. However, the effect is more pronounced with globally adaptive hiding.

For a classifier, we use Joachim's support vector machine (SVM) implementation [53], $\mathrm{SVM}^{\mathrm{light}}$. A linear kernel is used; we found other kernels perform only slightly differently. The SVM classifier is shown to be an effective classifier for steganalysis by Lyu *et al.* [51] and, more recently, by Pevný *et al.* [45].

Since the optimal hypothesis test finds the minimum divergence between PMF estimates, we are motivated to use PMF estimates to train the SVM. For our experiments with a detector not using dependency, we can use the appropriate PMF estimate: the normalized histogram of pixel values, a 256-dimensional feature vector. Unfortunately for the detector using dependency, the empirical matrix is too large (256^2 dimensions) to use directly. As with the other steganalysis schemes mentioned in Section II-C, we use a reduced version of the empirical matrix for a classification statistic. We have noted that image empirical matrices are very concentrated toward the main diagonal, and that hiding tends to spread the density away from this line. To capture this effect, the feature vector should then include the region immediately surrounding the main diagonal.

To generate the empirical matrix, we need a method of generating a 1-D chain from an image (i.e., a scan). We first use a vertical scanning, as in Fig. 1, for the experiments. We recognize that images have anisotropic dependencies not captured by vertical scanning, and so we also explore different feature vectors that combine horizontal, vertical, and diagonal pairs, in order to more accurately characterize pixel dependencies.

For an empirical matrix $\mathbf{M}$ calculated from an image, the six highest probabilities on the main diagonal $(\mathbf{M}_{ii})$ are chosen first, and for each of these, the following ten nearest differences are chosen:

$$\{\mathbf{M}_{i,i}, \mathbf{M}_{i,i-1}, \mathbf{M}_{i,i-2}, \ldots, \mathbf{M}_{i,i-10}\}.$$

Altogether, this gives a 66-dimensional vector. We also wish to capture changes along the center line. To do this, we subsample the remaining main diagonal values by four

$$\{\mathbf{M}_{1,1}, \mathbf{M}_{5,5}, \mathbf{M}_{9,9}, \ldots, \mathbf{M}_{253\,253}\}$$

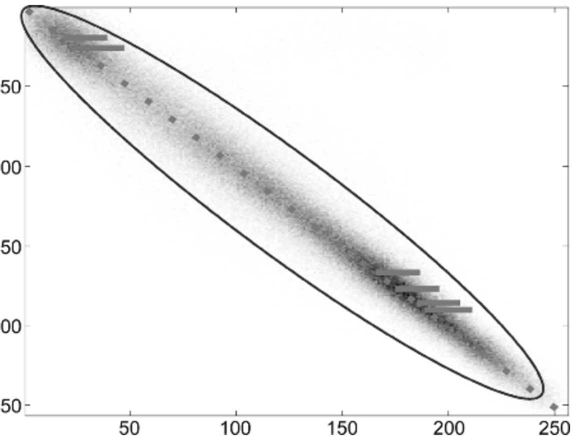


Fig. 4. Example of the feature vector extraction from an empirical matrix (not to scale). Most of the probability is concentrated in the circled region. Six row segments are taken at high probabilities along the main diagonal and the main diagonal itself is subsampled.

(Fig. 4). The resulting total feature vector has 129-dimensions, a manageable size that still captures much of the hiding effect. A comparison of the feature vectors used to evaluate the performance of both detectors, using and not using dependencies, is shown in Fig. 5. In addition to generating an empirical matrix based on adjacent pixels, we experimented with an empirical matrix generated from a pixel and an average of its four nearest neighbors. This is done in an attempt to capture a possible gain to using a more complex model, while still falling into our framework.

We tested the same four SS variants as in the previous sections. To relate these experiments to other work done, we based our hiding power on that reported in the literature. SS image steganography (SSIS) [48] is an implementation of globally adaptive hiding. In the experiments done by Marvel *et al.*, the MCR reported is always greater than −23 dB, so we choose this as a worst case. For the locally adaptive DCT scheme, we look to the experiments performed by Cox *et al.*, [47], and

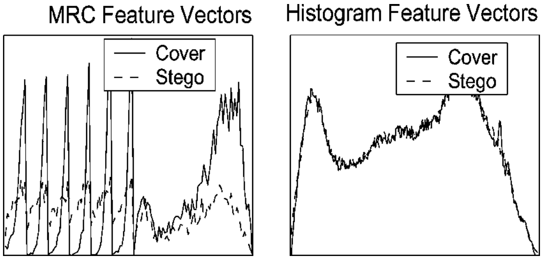


Fig. 5. Feature vector on the left is derived from the empirical matrix and captures the changes to interdependencies caused by SS data hiding. The feature vector on the right is the normalized histogram and only captures changes to first-order statistics, which are negligible.

use $\alpha = 0.1$, which gives an MCR of roughly -21 dB. In the spatial domain, we choose α to achieve a similar MCR.

Our results are summarized in the receiver operating characteristics (ROC) curves in Fig. 6 for the detector based on the empirical matrix and the histogram, respectively. In the ROCs, the best detector will reach the origin (0 false alarms and misses), and the worst detector is on the line connecting the upper-left corner to the lower right. The probabilities of false alarm and missed detection are defined as

$$\begin{aligned}&\Pr(\text{false alarm})\\&\quad=\frac{\text{number of clean images identified as stego}}{\text{total number of clean images}}\\&\Pr(\text{miss})\\&\quad=\frac{\text{number of stego images identified as clean}}{\text{total number of stego images}}.\end{aligned}$$

Since the vertical scan method will not capture all directions of image dependency, we explore different features that incorporate different aspects of dependency. First, we look at generating the same feature vector as in the above experiments instead of scanning the image into a vector using horizontal or zig–zag scanning (as is done for DCT coefficients in JPEG compression [54]).

In Fig. 7, we compare the ROCs of three detectors based on vertical, horizontal, and zigzag scans on locally adaptive transform hiding, the hiding scenario with the weakest detector performance. All methods perform approximately the same, with horizontal scan being slightly better than the other two. We find this same trend for locally adaptive spatial hiding as well as global hiding in either domain. Moreover, we tried several methods of combining different scan information. Generally, these perform about the same as a single directional scan.

We also compare the results of the detector based on the empirical matrix based on one adjacent pixel, and that generated from an average of four adjacent pixels. In Fig. 8, we compare the results of both detectors for locally adaptive DCT hiding; the results for the other three variants are similar. The detectors perform closely, suggesting the simple MC model captures the important changes introduced by hiding.

We note that our results for detecting spatial globally adaptive hiding, error rates on the order of 1%–5%, are similar to those of Harmsen and Pearlman in [19] detecting SSIS in color images. For detection, they used a statistic based on color plane statistics. Though the detection tests are not directly analogous since our tests are strictly on grayscale images, it is likely that a similar effect to the weakening of dependencies between pixels occurs between color planes. Celik *et al.* [20] perform the detection of stochastic modulation, statistically the same as spatial globally adaptive hiding. Stochastic modulation allows a greater embedding rate for a smaller MCR (or larger peak signal-to-noise ratio PSNR). As such, Celik *et al.* tested with a lower MCR, and so, although their detection rates are not as high as ours, it is difficult to directly compare.

In [55], Chandramouli studies the different but related problem of active steganalysis of SS hiding. In active steganalysis, the goal is to extract information about the message (or the message itself) from a known stego signal or, in this case, two known stego signals. Both correlations between the signals and within the signals themselves are exploited to estimate the original message. Of particular note in relation to our work, Chandramouli has exploited the fact that cover signals are not white Gaussians in order to identify message bits.

We find the results of a practical detector matches that which our divergence measurements and analysis lead us to expect. For the steganographer, it may first seem that locally adaptive DCT hiding is the superior choice for hiding. However, there are two important points to mention: First, unlike globally adaptive hiding, locally adaptive hiding only meets a target MCR on average. The MCR for each image varies and this may make detection more difficult. If the hider chooses instead to keep the MCR constant, rather than α, detection rates may increase. Second, although the currently realizable globally adaptive hiding rate is a function of the message signal power, the locally adaptive hiding rate is not readily available and may, in fact, be less than globally adaptive hiding for a given MCR. In all cases, there is clearly a gain for the steganalyst to use a model of dependency for detection. In the following section, we perform a similar analysis to a hiding scheme specifically designed to evade detection.

B. *Double-Compressed JPEG Perturbation Quantization*

Recently, Fridrich *et al.* [56] introduced an implementation of their perturbation quantization hiding method that creates stego images that mimic a double-compressed clean image. We measure the divergence of this method and show these are related to practical detection results presented by Kharrazi *et al.* [57]. As with SS hiding, we study the statistical effect of hiding to explain these findings.

1) Detectability of JPEG PQ Hiding: As the name implies, perturbation quantization is a variant of quantization index modulation (QIM) [58]. Standard QIM hiding in JPEG images has a distinctive statistical effect, and can be detected [18], [38]. Double-compressed PQ, however, is specifically contrived to minimize the statistical difference between the stego image, and an image that has simply been compressed twice. This is achieved by embedding in coefficients that have the same distribution after a second compression as they do after data hiding.

In [57], Kharrazi *et al.* measure the detection rates for three blind methods of steganalysis used on a variety of steganography schemes. The term "cover" is somewhat ambiguous for PQ JPEG hiding. The original source, from which the stego image is generated, is a once-compressed image. However,

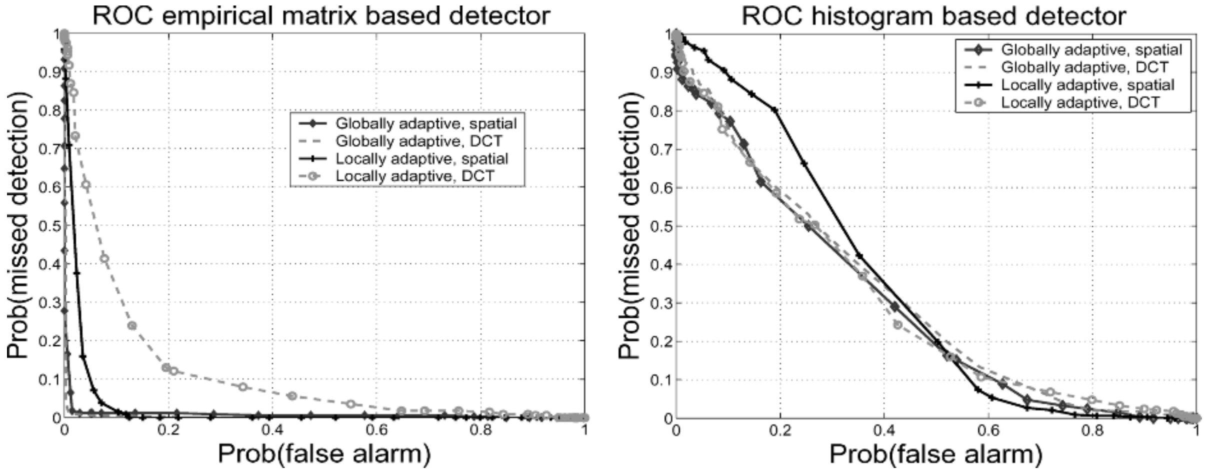


Fig. 6. ROCs of SS detectors based on empirical matrices (left) and 1-D histograms (right). In all cases, detection is much better for the detector including dependency. For this detector (left), the globally adaptive schemes can be seen to be more easily detected than locally adaptive schemes. Additionally, spatial and DCT hiding rates are nearly identical for globally adaptive hiding, but differ greatly for locally adaptive hiding. In all cases, detection is better than random guessing. The globally adaptive schemes achieve best error rates of about 2%–3% for Pr(false alarm) and P(miss).

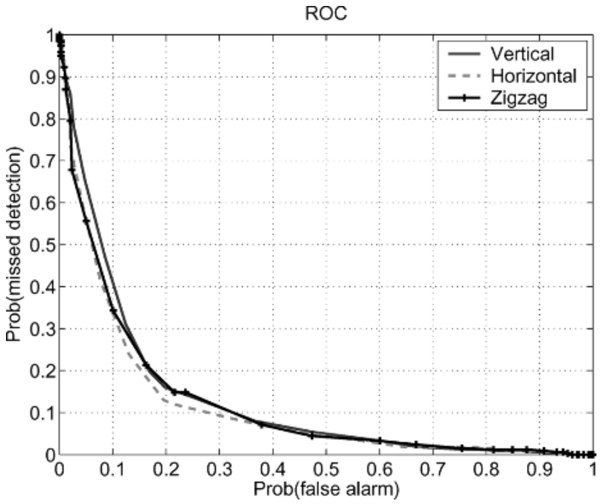


Fig. 7. Detecting locally adaptive DCT hiding with three different supervised learning detectors. The feature vectors are derived from empirical matrices calculated from three separate scanning methods: vertical, horizontal, and zigzag. All perform roughly the same.

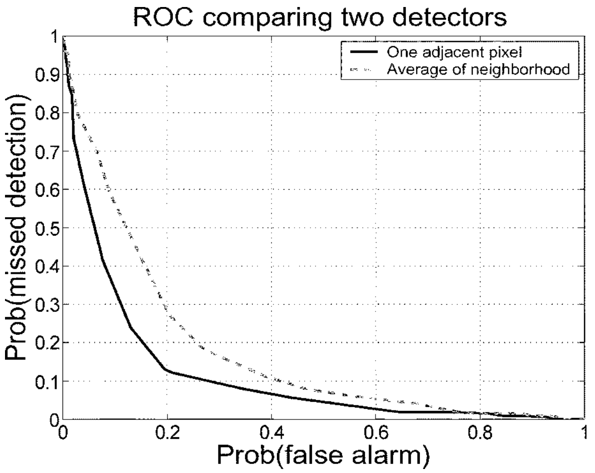


Fig. 8. Comparison of detectors for locally adaptive DCT SS hiding. The two empirical matrix detectors, one using one adjacent pixel and the other using an average of a neighborhood around each pixel, perform similarly.

PQ is designed to mimic twice-compressed images, which the authors argue occur naturally [56]. Because of this ambiguity, Kharrazi *et al.* measure the detection rates of two cases: comparing with the source (single-compressed) images, and comparing with re-compressed (i.e., double-compressed) images. For the first case, detection is found to be possible, but by no means certain. For example, in one case, the sum of errors (false alarm and missed detection) is about 0.3. For the second case, the detection rates are essentially random. In other words, guessing or flipping a coin will be just as effective for steganalysis. For details, please see their paper [57]. We note that the detection schemes are blind to the method, and one would expect better results from a scheme specifically designed to detect PQ JPEG. However, these results provide an idea of the detectability of this scheme. As with SS above (Section III-A-1), we measure the divergence introduced by PQ JPEG hiding. In Table III, we summarize the results. Q_1 and Q_2 are the JPEG quality levels used for the first and second compressions. Both of these cases correspond to a large number of embeddable coefficients, and all available coefficients are used. For the (75, 50) trial, the average embedding rate is 0.11 bits per pixel (bpp), 0.38 bits per nonzero DCT coefficient (bpnz-DCT). In the second trial (88, 76), the average rate is 0.13 bpp and 0.35 bpnz-DCT.

As in the SS case, we found that the measure of theoretically optimal detection of data hiding in Markov random chains corresponds to experiments in the nonidealized case. This again suggests that the model is a useful tool in judging the inherent detectability of a steganographic method. Additionally, there is a gain for a steganalyst to use dependency for detection, up to 7.5 times gain in this example. We now explore how this low divergence is obtained.

2) Statistical Effect of Double JPEG Compressed PQ: As mentioned above, the source for data hiding is an image that has undergone JPEG compression. During JPEG compression, the image is broken into small blocks, each of which undergoes a 2-d DCT. These DCT coefficients are then quantized to reduce the number of bits used to store or transmit the image (for details, see [54]). An inverse DCT of these coefficients reproduces

TABLE III

DIVERGENCE MEASURES OF PQ HIDING (ALL VALUES ARE MULTIPLIED BY 100). NOT SURPRISINGLY, THE DIVERGENCE IS GREATER COMPARED TO A TWICE-COMPRESSED COVER THAN A SINGLE-COMPRESSED COVER, MATCHING THE FINDINGS OF KHARRAZI ET AL. THE DIVERGENCE MEASURES ON THE RIGHT (COMPARING TO A DOUBLE-COMPRESSED COVER) ARE ABOUT HALF THAT OF THE LOCALLY ADAPTIVE DCT SS CASE IN WHICH DETECTION WAS DIFFICULT, HELPING TO EXPLAIN THE POOR DETECTION RESULTS

Single-compressed cover		
Q_1, Q_2	75,50	88,76
Mean MCR	-15.63	-17.89
Mean $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$	14.64	13.18
Mean $D(p^{(X)} \| p^{(S)})$	4.66	3.03
Mean ratio	4.23	6.66

Double-compressed cover		
Q_1, Q_2	75,50	88,76
Mean MCR	-19.26	-21.43
Mean $D(\mathbf{M}^{(X)}, \mathbf{M}^{(S)})$	3.04	3.89
Mean $D(p^{(X)} \| p^{(S)})$	0.63	0.63
Mean ratio	5.28	7.46

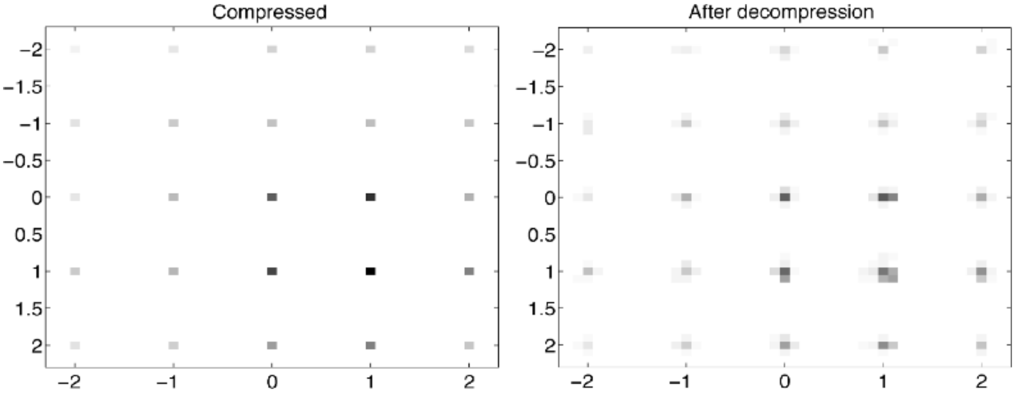


Fig. 9. On the left is an empirical matrix of DCT coefficients after quantization. When decompressed to the spatial domain and rounded to pixel values, right, the DCT coefficients are randomly distributed around the quantization points.

a spatial domain image. However, the spatial domain (pixel) values will no longer be integers, due to the quantization in the DCT domain. To display or otherwise use the image in the spatial domain, the pixel values are rounded to the nearest integer in the bit depth range (e.g., $\{0, 1, \ldots, 255\}$). Now, the DCT coefficients (of the pixel image) will no longer be exactly quantized, but instead randomly spread around the quantized value. In summary, if an image is compressed, then decompressed, the DCT values will be randomly distributed around their quantized values as seen in Fig. 9. Asymptotically, this density is a white Gaussian centered at the quantized value [56].

If the image is recompressed with a different quality level (i.e., different quantization step size), these blurred coefficients will be rounded to the nearest new quantizer output. In some special cases, the first quantizer output value lies halfway between two output levels of the new quantizer. For example, if the first quantizer used a step size of 21, and the second quantizer uses 24, then $4 \times 21 = 84$ is straddled by $3 \times 24 = 72$ and $4 \times 24 = 96$. Since it is assumed that the distribution is white Gaussian and, therefore, symmetric, it is expected that under normal quantization, roughly half of the coefficients originally quantized to 84 will become 72, and half 96. For pairs of coefficients, a quarter of pairs originally at (84,84) will become (72,72), a quarter (72,96), a quarter (96,72) and a quarter (96,96) (Fig. 10). Fridrich *et al.*, propose changing the quantization of these values to add hidden data. If instead a value originally at 84 becomes 72 to represent a zero, and becomes 96 to represent one, the statistics are not expected to change.

This statistical equivalency will only fail if the density blurring is not, in fact, symmetric about the original quantization point. Though asymptotically it is expected to be, each realization will be slightly asymmetric, as can be seen in Fig. 9. We have found the asymmetry to be small; however, the calculated divergence between a double-compressed cover image and a PQ stego image will be greater than zero. The net effect, however, is minimal and the divergence and detection results above are not surprising. Again, we see a match between analysis, divergence measurements, and practical detection.

IV. CONCLUSION

Our Markov model for cover data permits explicit computation of a detection-theoretic divergence measure that characterizes the susceptibility of a steganographic scheme to detection by an optimal classifier. This measure has advantages over other steganographic security benchmarks. It provides a more accurate security measure than Cachin's ϵ-secure [13] metric, as dependencies between samples are accounted for. Additionally, it is a more general metric than that given by Chandramouli *et al.* [5], [26], which is measured for a given detector. The divergence measure also provides a quick estimate of the performance benefits of using dependency in steganalysis: the ratio of the divergence for the Markov model to the divergence between marginal PMFs represents the factor by which the use of dependency reduces the number of samples required for a given performance relative to steganalysis based on 1-D histograms.

The application we have focused on in examples and numerical results is image steganalysis. While the Markov model does not completely capture interpixel dependencies in images, we have shown it to be consistent with many image steganalysis schemes exploiting memory, which typically use a function of the statistics used to optimally discriminate between Markov source models. Furthermore, the detection-theoretic benchmarks computed using the Markov model are close to the performance attained by practical image steganalysis tech-

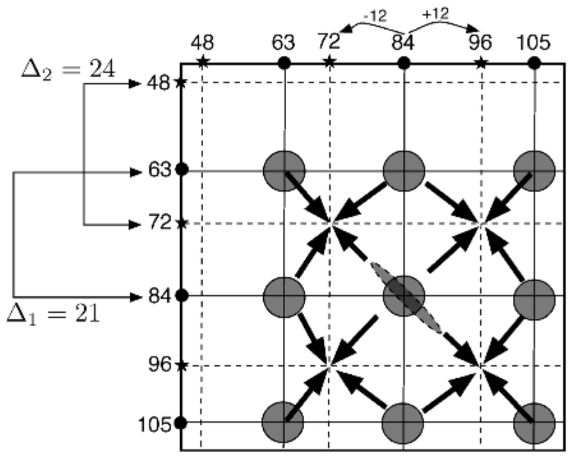


Fig. 10. Simplified example of second compression on an empirical matrix. Solid lines are the first quantizer intervals, dotted lines are the second. The arrows represent the result of the second quantization. The density blurring after decompression is represented by the circles centered at the quantization points. For the density at (84,84), if the density is symmetric, the values will be evenly distributed to the surrounding pairs. If, however, there is an asymmetry, such as the dotted ellipse, the new density will favor some pairs over others (e.g., (72,72), (96,96) over (72,96), (96,72)). The effect is similar for other splits such as (63,84) to (72,72) and (72,96).

niques. However, further research is needed into whether more complex statistical models can yield better image steganalysis techniques and how to compute performance benchmarks for such techniques. Improved models for images could include more degrees of dependency, as well as some model of non- or piecewise-stationarity. However, the parameters of more complex models are also more difficult to estimate, and variations from image to image may make it difficult to calibrate steganalysis techniques based on such models. Thus, much work remains to be done on the fundamental problem of understanding how the complexity of the model for the cover data impacts the accuracy of estimating the model parameters, computational complexity, and performance of steganalysis based on the model.

Acknowledgment

The authors would like to thank the reviewers for their careful reading and insightful comments. We also express appreciation to D. Fedorov for translation assistance.

References

[1] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. San Mateo, CA: Morgan Kauffman, 2002.

[2] B. Macq, J. Dittmann, and E. J. Delp, “Benchmarking of image watermarking algorithms for digital rights managment,” *Proc. IEEE*, vol. 92, no. 6, pp. 971–983, Jun. 2004.

[3] E. T. Lin and E. J. Delp, “A review of data of hiding in digital images,” in *Proc. Image Processing, Image Quality, Image Capture Systems Conference (PICS ’99)*, Savannah, GA, 1999, pp. 274–278.

[4] T. Aura, “Practical invisibility in digital communication,” in *Lecture Notes Comput. Sci.: 1st Int. Workshop Inform. Hiding*, 1996, vol. 1174, pp. 265–278.

[5] R. Chandramouli, M. Kharrazi, and N. Memon, “Image steganography and steganalysis: concepts and practices,” in *Proc. 2nd Int. Workshop Digital Watermarking*, Seoul, Korea, Jan. 2003, pp. 35–49.

[6] J. Fridrich and M. Goljan, “Practical steganalysis of digital images—state of the art,” in *Proc. IST/SPIE 14th Annu. Symp. Electronc Imaging Science Technology*, 2002, vol. 4675.

[7] ——, “On estimation of secret message length in LSB steganography in spatial domain,” in *Proc. IST/SPIE 16th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2004.

[8] S. Dumitrescu, X. Wu, and Z. Zhe Wang, “Detection of LSB steganography via sample pair analysis,” *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1995–2007, Jul. 2003.

[9] T. Zhang and X. Ping, “A new approach to reliable detection of LSB steganography in natural images,” *Signal Process.*, vol. 83, no. 10, pp. 2085–2093, Oct. 2003.

[10] B. Roue, P. Bas, and J.-M. Chassery, “Improving LSB steganalysis using marginal and joint probabilistic distributions,” in *Proc. ACM Multimedia Security Workshop*, Magdeburg, Germany, Sep. 2004, pp. 75–80.

[11] J. Fridrich, M. Goljan, and D. Hogea, “Attacking the OutGuess,” in *Proc. ACM Workshop Multimedia Security*, Juan-Pins, France, Dec. 2002.

[12] I. Avcibas, N. Memon, and B. Sankur, “Steganalysis using image quality metrics,” *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 221–229, Feb. 2003.

[13] C. Cachin, “An information theoretic model for steganography,” in *Proc. Int. Workshop Information Hiding*, Portland , OR, Apr. 1998, vol. 1525, pp. 306–318.

[14] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[15] P. Moulin and Y. Wang, “New results on steganographic capacity,” in *Proc. Conf. Information Sciences Systems (CISS)*, Princeton, NJ, Mar. 2004.

[16] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, “Detection of hiding in the least significant bit,” *IEEE Trans. Signal Processing, Supplement on Secure Media I*, vol. 52, no. 10, pp. 3046–3058, Oct. 2004.

[17] P. Sallee, “Model-based steganography,” in *Proc. 2nd Int. Workshop Digital Watermarking*, Seoul, Korea, 2003, pp. 154–167.

[18] M. T. Hogan, N. J. Hurley, G. C. M. Silvestre, F. Balado, and K. M. Whelan, “ML detection of steganography,” in *Proc. SPIE Symp. EIS&T*, San Jose, CA, Jan. 2005, pp. 16–27.

[19] J. J. Harmsen and W. A. Pearlman, “Steganalysis of additive noise modelable information hiding,” in *Proc. IST/SPIE 15th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2003, pp. 21–24.

[20] M. U. Celik, G. Sharma, and A. M. Tekalp, “Universal image steganalysis using rate-distortion curves,” in *Proc. IST/SPIE 16th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2004, pp. 19–22.

[21] K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, “Steganalysis of spread spectrum data hiding exploiting cover memory,” in *Proc. IST/SPIE 17th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2005, pp. 38–46.

[22] N. Provos, “Defending against statistical steganalysis,” in *Proc. 10th USENIX Security Symp.*, Washington, D.C., Aug. 2001, pp. 323–336.

[23] J. J. Eggers, R. Bauml, and B. Girod, “A communications approach to image steganography,” in *Proc. IST/SPIE 14th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2002, pp. 26–37.

[24] P. Sallee, “Model-based methods for steganography and steganalysis,” *Int. J. Image Graphics*, vol. 5, no. 1, pp. 167–190, Jan. 2005.

[25] E. Franz, “Steganography preserving statistical properties,” in *Proc. 5th Int. Working Conf. Communication Multimedia Security*, Oct. 2002, pp. 278–294.

[26] R. Chandramouli and N. Memon, E. J. Delp, III and P. W. Wong, Eds. , “Steganography capacity: a steganalysis perspective,” *Security and Watermarking of Multimedia Contents V*, vol. 5020, Jan. 2003.

[27] Y. Wang and P. Moulin, “Steganalysis of block-structured stegotext,” in *Proc. IST/SPIE’s 16th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2004, pp. 477–488.

[28] K. L. Chung, *Markov Chains With Stationary Transition Probabilities*. New York: Springer-Verlag, 1960.

[29] M. Sidorov, “Hidden Markov models and steganalysis,” in *Proc. ACM Multimedia Security Workshop*, Magdeburg, Germany, Sep. 2004, pp. 63–67.

[30] A. Rangarajan and R. Chellappa, *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press, 1995, pp. 564–567.

[31] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, U.K.: Chapman & Hall, 1995.

[32] A. Habibi, “Comparison of n-th order DPCM encoder with linear transformations and block quantization techniques,” *IEEE Trans. Commun. Technol.*, vol. COM-19, no. 6, pp. 948–956, Dec. 1971.

[33] S. Natarajan, “Large deviations, hypothesis testing, and source coding for finite Markov chains,” *IEEE Trans. Inform. Theory*, vol. IT-31, no. 3, pp. 360–365, May 1985.

[34] R. F. Walker, P. Jackway, and I. D. Longstaff, "Improving co-occurrence matrix feature discrimination," in *Proc. Digital Image Computing: Techniques Applications*, Brisbane, Australia, Dec. 1995, pp. 643–648.

[35] V. Anantharam, "A large deviations approach to error exponents in source coding and hypothesis testing," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 938–943, Jul. 1990.

[36] N. Provos and P. Honeyman, "Detecting steganographic content on the internet," in *Proc. ISOC NDSS*, San Diego, CA, Feb. 2002, Online Available: http://www.isoc.org/isoc/conferences/ndss/02/proceedings/.

[37] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of JPEG images: breaking the F5 algorithm," in *Lecture Notes Comput. Sci.: 5th Int. Workshop Inform. Hiding*, 2002, vol. 2578, pp. 310–323.

[38] K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, "Steganalysis of quantization index modulation data hiding," in *Proc. ICIP*, Singapore, Oct. 2004, pp. 1165–1168.

[39] Y. Wang and P. Moulin, "Steganalysis of block-DCT steganography," in *Proc. IEEE Workshop Statistical Signal Processing*, St. Louis, MO, Sep. 2003, pp. 339–342.

[40] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of LSB steganography in color and grayscale images," in *Proc. ACM Workshop Multimedia Security*, Ottawa, ON, Canada, 2001, pp. 27–30.

[41] M. Sidorov, "A statistical steganalysis for digital images," in *Proc. Int. I and S Workshop*, Moscow, Russia, Jan. 2004, pp. 34–36.

[42] A. Ker, "Resampling and the detection of LSB matching in color bitmaps," in *Proc. IST/SPIE 17th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2005, pp. 1–15.

[43] ——, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Process. Lett.*, vol. 12, no. 6, pp. 441–444, Jun. 2005.

[44] J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," in *Proc. 6th Information Hiding Workshop*, ON, Canada, May 2004, pp. 67–81.

[45] T. Pevny and J. Fridrich, "Toward multi-class blind steganalyzer for JPEG images," in *Lecture Notes Comput. Sci.: Proc. Int. Workshop Digital Watermarking*, 2005, vol. 3710, pp. 39–53.

[46] A. Ambalavanan and R. Chandramouli, "A Bayesian image steganalysis approach to estimate the embedded secret message length," in *Proc. ACM Multimedia Security Workshop*, New York, Aug. 2005, pp. 33–38.

[47] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[48] L. Marvel, C. G. Boncelet, Jr, and C. T. Retter, "Spread spectrum image steganography," *IEEE Trans. Image Process.*, vol. 8, no. 8, pp. 1075–1083, Aug. 1999.

[49] J. Fridrich and M. Goljan, "Digital image steganography using stochastic modulation," in *Proc. IST/SPIE's 15th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2003, pp. 191–202.

[50] R. C. Gonzalez and R. E. Woods*, Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.

[51] S. Lyu and H. Farid, "Steganalysis using color wavelet statistics and one-class support vector machines," in *Proc. IST/SPIE 16th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2004, pp. 35–45.

[52] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 221–229, Feb. 2003.

[53] T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods—Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1999.

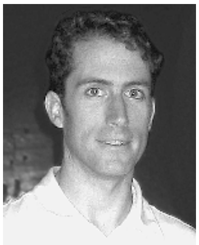
[54] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[55] R. Chandramouli, "A mathematical framework for active steganalysis," *Multimedia Syst.*, vol. 9, pp. 303–311, Sep. 2005.

[56] J. Fridrich, M. Goljan, and D. Soukal, "Perturbed quantization steganography with wet paper codes," in *Proc. ACM Multimedia Security Workshop*, Magdeburg, Germany, Sep. 2004, pp. 4–15.

[57] M. Kharrazi, H. T. Sencar, and N. Memon, "Benchmarking steganographic and steganalysis techniques," in *Proc. IST/SPIE 17th Annu. Symp. Electronic Imaging Science Technology*, San Jose, CA, Jan. 2005, pp. 252–263.

[58] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.



Kenneth Sullivan (M'95) received the B.S. degree in electrical engineering from the University of California at San Diego in 1998, and the M.S. and Ph.D. degrees from the University of California at Santa Barbara in 2002 and 2005, respectively.

His research interests include image processing, data hiding, and multimedia databases.

Upamanyu Madhow (F'05) received the B.Sc. degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana-Champaign (UIUC), in 1987 and 1990, respectively.

From 1990 to 1991, he was a Visiting Assistant Professor at UIUC. From 1991 to 1994, he was a Research Scientist with Bell Communications Research, Morristown, NJ. From 1994 to 1999, he was on the faculty of the Department of Electrical and Computer Engineering with the University of Illinois. Currently, he is a Professor with the Department of Electrical and Computer Engineering, University of California, Santa Barbara. His research interests are in communication systems and networking, with current emphasis on wireless communication, sensor networks, and multimedia security. He was Associate Editor for Spread Spectrum for the IEEE TRANSACTIONS ON COMMUNICATIONS, and was Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Dr. Madhow is a recipient of the National Science Foundation CAREER award.

Shivkumar Chandrasekaran received the M.Sc. degree in physics (Hons.) from B.I.T.S., Pilani, India, in 1987, and the Ph.D. degree in computer science from Yale University, New Haven, CT, in 1994.

He was a Visiting Instructor at North Carolina State University, Raleigh, in the mathematics department, before joining the University of California at Santa Barbara in the electrical and computer engineering department, where he is currently an Associate Professor. His research interests are in computational mathematics.

B. S. Manjunath (F'05) received the B.E. degree in electronics (Hons.) from the Bangalore University, Bangalore, India, in 1985, the M.E. degree in systems science and automation (Hons.) from the Indian Institute of Science, Bangalore, in 1987, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1991.

Currently, he is a Professor of Electrical Computer Engineering and Director of the Center for Bio-Image Informatics at the University of California, Santa Barbara. His research interests include image processing, data hiding, multimedia databases, and bio-image informatics. He is a co-editor of the book on *Introduction to MPEG-7* (Wiley, 2002). He was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and is currently an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and the IEEE TRANSACTIONS ON MULTIMEDIA.

Dr. Manjunath was a recipient of the national merit scholarship from 1978 to 1985 and was awarded the university gold medal for the best graduating student in electronics engineering from Bangalore University in 1985.