Steganalysis of Spread Spectrum Data Hiding Exploiting Cover Memory

Kenneth Sullivan, Upamanyu Madhow, Shivkumar Chandrasekaran, and B.S. Manjunath

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106-9560, USA

ABSTRACT

In this paper we study steganalysis, the detection of hidden data. Specifically we focus on detecting data hidden in grayscale images with spread spectrum hiding. To accomplish this we use a statistical model of images and estimate the detectability of a few basic spread spectrum methods. To verify the results of these findings, we create a tool to discriminate between natural "cover" images and "stego" images (containing hidden data) taken from a diverse database. Existing steganalysis schemes that exploit the spatial memory found in natural images are particularly effective. Motivated by this, we include inter-pixel dependencies in our model of image pixel probabilities and use an appropriate statistical measure for the security of a steganography system subject to optimal hypothesis testing. Using this analysis as a guide, we design a tool for detecting hiding on various spread spectrum methods. Depending on the method and power of the hidden message, we correctly detect the presences of hidden data in about 95% of images.

Keywords: steganalysis, steganography, data hiding, spread spectrum

1. INTRODUCTION

For the past several years spread spectrum (SS) data hiding has enjoyed a wide popularity in the data hiding community, particularly for watermarking. Additionally SS hiding can be used for covert communication, i.e. steganography, which inevitably leads to others searching to detect the presence of this secret communication. In this paper we focus on steganalysis, the detecting of hidden data, in images undergoing spread spectrum data hiding.

There have been many steganalysis methods put forth over the years.[1,2] For spread spectrum hiding, Harmsen and Pearlman[3] analyze the general case of additive noise modeling, which some varieties of SS can be modeled as, using statistics across the RGB color planes. Wang and Moulin[4] consider additive SS steganalysis for the case of Gaussian covers. For other schemes or general steganalysis, many have taken approaches that model images as realizations of random processes, and leverage statistical hypothesis testing theory to characterize optimal detection limits and design approaches for both detection and evading detection.[4–8] Most of these (Wang and Moulin[4] consider general Gaussian random vectors) employ an independent and identically distributed (i.i.d.) random process for analysis. Images however are known not to be i.i.d., since each pixel is statistically dependent on its neighbors. Steganalysis methods[9–11] that exploit this cover memory well are very effective, suggesting that statistical models need to include inter-pixel dependencies.

In Section 2 we present our Markov chain image model, statistical analysis of spread spectrum hiding under this model, and estimations of the detectability of various adaptations of SS hiding. In Section 3, we detail steganalysis experiments performed on a database of images and present the results. Finally Section 4 summarizes our conclusions.

http://vision.ece.ucsb.edu, Send correspondence to K. Sullivan: sullivak@ece.ucsb.edu

2. SPREAD SPECTRUM HIDING IN CORRELATED DATA

There are many flavors and adaptations of spread spectrum data hiding. However many of these improvements are designed to increase the robustness from attacks, and the statistical effect is generally the same as the most basic schemes. We therefore consider the general approach presented by Cox et al,[12] of adding a noise-like message bearing signal to the cover medium. The essence of this technique is straightforward to model. The hidden message is a pseudo-randomly generated sequence approximating a zero mean, unit variance Gaussian, $\mathcal{N}(0,1)$. The message sequence is scaled and added to the cover; typically the scaling is proportional to the cover. The scaling can be performed globally or locally, i.e. the scale factor can be the same for all coefficients or vary as a function of the cover coefficient.

In addition to choosing between a globally and locally adaptive scheme, a hider can choose to embed directly in the spatial domain or in a transform space such as the DFT or DCT. We however concentrate our steganalysis on the spatial domain for a number of reasons:

- The transform of a Gaussian message sequence with linear transforms such as DCT and DFT is also Gaussian, and by the superposition property of these transforms, adding a Gaussian signal in the frequency domain is equivalent to adding in the spatial domain. This does not mean that spread spectrum data hiding in the transform domain is exactly equivalent to hiding in intensity values, because of the effect of rounding and clipping. However we have observed that in practice the deviation is small.
- The sample space of spatial domain values is small and static. In practical terms, the pixel values are fixed bit-depth integers, whereas the transform coefficients are floating point values. There are a number of difficulties that arise when estimating the distribution of floating point values that do not exist with a small set of integers.
- Spatial analysis is generic. We can use the same general framework, and fine tune to specific scheme differences at the final stage.

Finally, we concentrate on intensity (gray-value) images; typically using 8-bit images though any bit-depth can be used with this approach. We point out that we examine the correlation between pixels in the intensity plane, ignoring the relationship across color planes. An advantage of this approach is that most SS hiding schemes are designed for hiding in gray-scale images, and correlation is known to be strong in the intensity plane. If, however, hiding is done across all color planes, important information for detection[3] is ignored. In the future we plan the straightforward extension of incorporating color plane correlation information with inter-pixel correlation.

2.1. Judging Detectability in Correlated Data

In the context of independent and identically distributed (i.i.d) data, Cachin[5] suggested the Kullback-Leibler (K-L) divergence, also known as relative entropy, between the cover and stego distributions as a benchmark for the detectability of a steganography system. Image cover data however is not i.i.d. We can abandon the i.i.d. assumption, however this raises the question of an appropriate benchmark. For the i.i.d. case, the error probabilities for an ideal hypothesis test decrease exponentially as the K-L divergence increases,[13] and for zero K-L divergence an ideal hypothesis test can do no better than random guessing. An appropriate benchmark for correlated data should have the same property. If we model the cover data as a Markov chain[14] to capture inter-dependence, then such a function exists.[15] The function is essentially a distance between the empirical matrices of the two hypotheses: cover and stego. The empirical matrix of a Markov chain realization is comprised of the counts of the numbers of transitions from one state to another. So for a received vector of pixel values, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, if we let $n_{ij}(\mathbf{y})$ be the number of transitions from value i to value j in $\mathbf{y}$, we have the empirical matrix $\mathbf{M}(\mathbf{y}) = (n_{ij}(\mathbf{y})/n)$. Here the empirical matrix (or co-occurrence matrix) can be viewed as an estimate of the joint probability mass function (PMF), and we will generally view it as such. We use the following divergence function between two empirical matrices $\mathbf{P}$ and $\mathbf{Q}$:

$$D(\mathbf{P},\mathbf{Q}) = \sum_{i,j} p_{ij} \log\left(\frac{p_{ij}}{\sum_j p_{ij}} \frac{\sum_j q_{ij}}{q_{ij}}\right)$$

For a constant false alarm rate, the minimal achievable missed detection rate approaches $e^{-nD(\mathbf{P},\mathbf{Q})}$ as n, the number of samples, goes to infinity,[15] just as in the i.i.d. case with K-L divergence. That is, under the assumption of a Markov chain model, the performance of any possible detector is exponentially bounded by this measure. We have then a measure of the information gained from one sample of a Markov chain. We can use this value to compare the detectability of SS hiding in a Markov chains of varying parameters. We can also compare with a steganalyzer assuming i.i.d. data, to evaluate if the increased complexity of a detector assuming correlations is worthwhile.

2.2. Statistical Effect of Spread Spectrum Hiding

For globally adaptive hiding, the same scale factor is used for all coefficients, so the function to embed a zero mean, unit variance, Gaussian message signal $\mathbf{D}$ into the cover medium $\mathbf{X}$ is $S_k = X_k + \alpha D_k$ where $\mathbf{S}$ is the resulting stego sequence and α is the scaling factor. This is essentially the method used by Marvel, Boncelet, and Retter in spread spectrum image steganography (SSIS).[16] The additive "noise" is $\sim \mathcal{N}(0, \alpha^2)$. The effect is additive and the statistical effect is a convolution of the message and cover distributions.[3] The consequence of this convolution is easily seen in the joint PMFs, see Figure 1

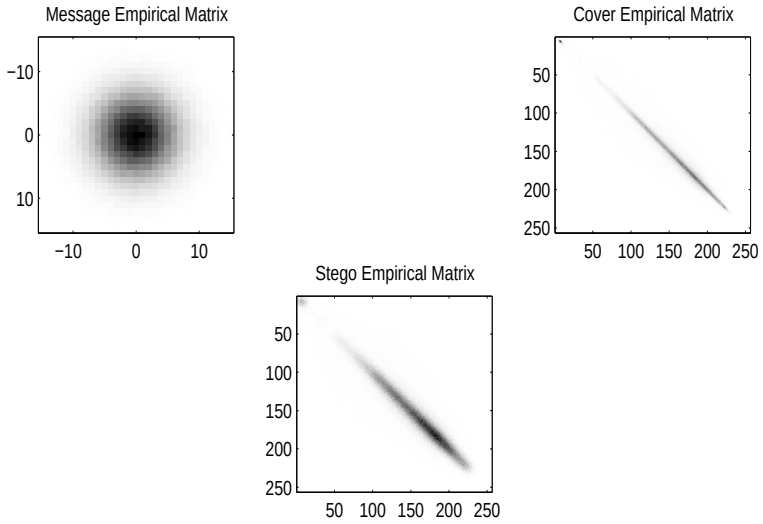

Figure 1. These plots show the empirical matrix of the cover, message, and stego. Black represents a large number of transitions, white represents few or none. Additive hiding convolves a white Gaussian message with the cover, weakening the dependencies in the cover image, seen as a spreading from the main diagonal.

The spread away from the center line means the probability of two consecutive coefficients having the same or close values decreases. Intuitively, by adding independent message data, the cover data becomes less dependent.

For locally adaptive hiding, the embedding function is $S_k = X_k + \alpha X_k D_k$. This additive description however defines the effect of hiding as a function of the cover samples, X_k. To avoid this we can alternatively view the hiding as multiplying X_k by the random variable $B = (1 + \alpha D_k) \sim \mathcal{N}(1, \alpha^2)$.

The joint distribution of the resulting unquantized stego variables S' is

$$F_{S'_1,S'_2}(s'_1, s'_2) = \int_0^\infty \int_0^\infty f(b_1) f(b_2) \left[\sum_{x_1=0}^{\min(\lfloor s'_1/b_1 \rfloor, 255)} \sum_{x_2=0}^{\min(\lfloor s'_2/b_2 \rfloor, 255)} P(x_1, x_2) \right] \mathrm{d}b_1 \mathrm{d}b_1$$

and the density is

$$f_{S'_1,S'_2}(s'_1, s'_2) = \frac{\partial^2 F_{S'_1,S'_2}(s'_1, s'_2)}{\partial s'_1 \partial s'_2}$$

For a given cover probability mass function (PMF), $P(x_1, x_2)$, these expressions can be evaluated numerically. Here we have assumed that the probability density function $f(b_1) = f(b_2)$ is essentially Gaussian, but with zero probability at values less than or equal to zero. This assumption is warranted by the typical α values used in hiding, which are chosen small enough to avoid visual distortion to images. To fit standard image formats, the S' are rounded and clipped to become S. After rounding, the PMF is

$$P(s_1, s_2) = \int_{s_1-.5}^{s_1+.5} \int_{s_2-.5}^{s_2+.5} f_{S'_1,S'_2}(s'_1, s'_2) \mathrm{d}s'_1 \mathrm{d}s'_2$$

To clip the values to fit within [0,255], we can simply not embed in those coefficients, i.e. the clipping process is:

$$S = \begin{cases} Q(S') & Q(S') \in [0, 255] \\ X & \text{else} \end{cases}$$

Where $Q(\cdot)$ is a rounding function. This slightly reduces the actual strength of hiding, making the task of steganalysis more difficult, but typically the effect is negligible.

The net statistical result of the hiding process is not obvious from the equations, however as seen in Figure 2 the consequence of hiding is again a "spreading" of probability away from the center line. However the effect is less strong. For locally adaptive hiding, if we view the embedding as additive noise, the white Gaussian noise is colored by scaling with the cover. We would then expect the de-correlating effect to be less pronounced then adding white noise. This intuition is borne out by measurements in the following section.

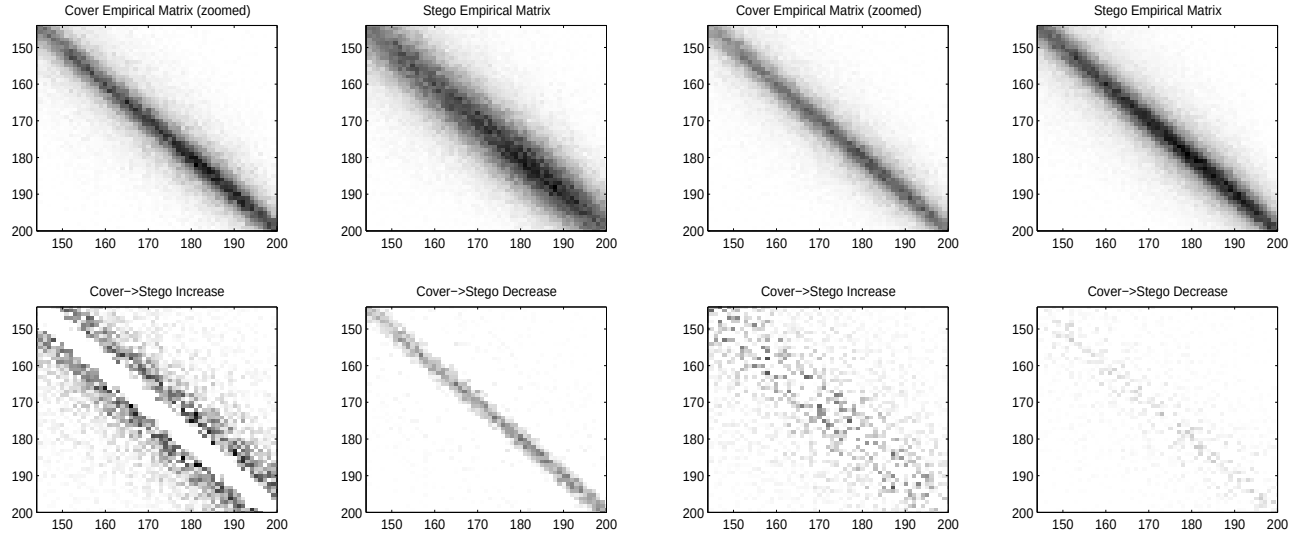

Figure 2. Though both global (left) and local (right) adaptive schemes weaken the dependencies in the host, the effect of the locally adaptive scheme is less pronounced. That is, both decrease the number of transitions near the main diagonal (where dependency is strong) while increasing away from the main diagnoal, however the magnitude of change is less with locally adaptive hiding.

2.3. Measuring Detectability of Hiding

For a given cover joint PMF, we know the effect of data hiding. However, cover statistics change from image to image. To analyze the detectability of spread spectrum hiding on images, we analyze statistics taken from a variety of real images. Random message data is hidden into synthetic cover data generated from these real cover statistics, and the resulting K-L divergence is measured. Over several images, an estimate of the divergence introduced by SS hiding can be found.

We examine global and local adapting schemes embedded into both the spatial domain and DCT domain. Two of these four cases correspond to tests run by others: SSIS[16] is a globally adaptive spatial embedding scheme

and Cox et al[12] use a locally adaptive DCT scheme in their experiments. We added SS data over a range of different message to cover ratios (MCR), to quantify the important effect of message power. For the globally adaptive scheme we can hold the MCR constant and adjust the total message power for each image. In the adaptive scheme, we hold the scale factor α constant. In this case MCR varies from image to image; we note the average value along with the divergences. Using a non-i.i.d. model adds complexity to the detector. To verify that this added complexity yields an increase in detection rates, we compare the normalized one-dimensional histogram divergence to the empirical matrix divergence.

In Table 1 and Table 2 we show the different divergences of empirical matrices (labeled 2-d div.) averaged over many images for a range of embedding powers. To show the gain in accounting for correlation, we show the average K-L divergence of the one-dimensional normalized histograms (labeled 1-d div.). The greater the divergence, the better an ideal detector will perform. These findings imply that SS hiding is indeed detectable if the detector has the several hundred samples available in image data. We note that in this ideal case, the cover statistics are assumed to be known, which is not the case in a realistic scenario.

Globally adaptive, Spatial			
MCR	-23	-20	-17
Mean 2-d div.	39.78	52.12	64.69
Mean 1-d div.	3.438	4.553	5.472
Mean ratio	20.45	20.23	20.27

Globally adaptive, DCT			
MCR	-23	-20	-17
Mean 2-d div.	40.14	52.22	64.52
Mean 1-d div.	3.514	4.395	5.526
Mean ratio	20.31	20.20	20.14

Table 1. Divergences for globally adaptive hiding in two domains, all divergence values are $\times 10^{-2}$. As expected the divergence, and thus detectability, increases with added message power. Additionally there is a gain to examining the cover dependencies, as indicated by the greater divergence per sample. For example, in both cases at MCR = -20 dB, the divergence considering correlation is about 20.2 times greater than assuming i.i.d. data. This means 95% fewer samples are required to achieve the same detection results.

Locally adaptive, Spatial			
Alpha	0.05	0.1	0.5
Mean MCR	-20.34	-14.63	-3.17
Mean 2-d div.	40.14	56.12	93.22
Mean 1-d div.	2.514	4.188	12.25
Mean ratio	26.06	22.32	9.613

Locally adaptive, DCT			
Alpha	0.05	0.1	0.5
Mean MCR	-26.07	-20.23	-7.205
Mean 2-d div.	2.649	3.897	14.69
Mean 1-d div.	2.499	4.344	9.735
Mean ratio	2.997	2.151	2.433

Table 2. Divergence values for locally adaptive hiding (again $\times 10^{-2}$). As expected from the intuition mentioned above, locally adaptive hiding introduces less divergence (for equal MCR) between cover and stego. Also, because the added message is a function of the cover medium, the locally adaptive divergence depends on the domain of embedding.

These findings motivate further pursuit of SS steganalysis, since theoretically, detection is possible. In the next section we present experiments performed on real image data, in which the statistics are not known beforehand.

3. DETECTION EXPERIMENTS

If we know the cover statistics, we can calculate the stego statistics and use a likelihood ratio test[17] to classify any unknown image as either cover or stego. For the realistic blind case, we do not know the cover statistics. We have several options. We can attempt to estimate the cover statistics from the received data. A problem with this, especially with model fitting estimates, is that the estimation may be too coarse, resulting in an estimate that is further from the cover than even the stego. It is in general difficult to find an estimation function that will return cover statistics from both stego data and cover data. Another approach is to attempt to classify between all possible cover and stego images. Specifically, a learning system can be trained on both cover and stego images, to learn to discriminate between the two. Learning has been successfully used for steganalysis[3, 18–20] and Martin et al[21] have shown many methods of steganography to be "unnatural." We employ this approach to detect spread spectrum hiding.

3.1. Test Details

3.1.1. Hiding Parameters

We test the detection of the varieties of spread spectrum hiding examined in Section 2, a globally adaptive spatial domain scheme as in Marvel et al's SSIS,[16] a locally adaptive DCT-domain scheme as in Cox et al's experiments,[12] as well as a locally adaptive spatial scheme and a globally adaptive DCT scheme. For SSIS, we used a constant message to cover ratio (MCR) of -23 dB. This is just below the minimum hiding power reported by Marvel et al, so we chose this as a worst-case for detection. For the adaptive DCT tests, we had a constant α of 0.1, the same as the experiments reported in the paper. We found the mean MCR hiding with this α to be -20.71 dB. For the locally adaptive spatial scheme we choose an α that tended to have a similar mean MCR. Finally for the globally adaptive DCT scheme, we also used a constant MCR of -23. Unfortunately it is difficult to state the effective hiding rate in bits per pixel (bpp) for these tests. SSIS uses image restoration, varying message power, and error correction codes to overcome the noise due to the cover in decoding the data. The reported payload is therefore different from image to image. The maximum payload reported is 0.1667 bpp, so we estimate that the effective hiding rate of our test images is less than or equal to this. Cox et al present their hiding method as a watermarking application, which is effectively one bit per image, though more data than this could almost certainly be sent with the method.

3.1.2. Classifier

As a classifier, we use Joachim's support vector machine (SVM) implementation,[22] SVM^{light}. The machine is trained on roughly 700 images and the same amount is used for testing. Each set, training and testing, consists of half cover images and half (distinct) stego images. A linear kernel is used; we found other kernels perform only slightly differently.

3.1.3. Image Database

In an attempt to capture the variance of real world images, we use images from a database comprised of four separate sources:

1. digital camera images, partitioned into smaller sub-images
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.

3.1.4. Feature Vector

We use a learning machine as a substitute for a likelihood ratio test, to attempt to learn unknown statistics. Therefore we should choose the empirical matrices, or joint PMFs, as the feature to learn. However this would result in an enormous feature vector. For example, an 8-bit image has 256 values, so the joint PMF has 65,636 features. To reduce this to a more manageable size, we first observe that these matrices are very sparse, because it is very rare to have a large jump in values from one pixel to another. We can ignore regions that have zero or very low probabilities of occurring. Additionally, we noted in Section 2.2 that hiding tends to spread values away from the main diagonal. To concentrate on this effect we examine the region immediately surrounding the main diagonal, specifically high probability regions where changes are greater in magnitude. With this in mind, first the 6 highest probabilities on the main diagonal ($P(x_1 = m, x_2 = m)$) are chosen, and for each of these the following 10 nearest differences are picked:

$$\{P(x_1 = m, x_2 = m),\ P(x_1 = m, x_2 = m - 1),\ P(x_1 = m, x_2 = m - 2), \ldots, P(x_1 = m, x_2 = m - 10)\}$$

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

$$\{P(x_1 = 1, x_2 = 1),\ P(x_1 = 5, x_2 = 5),\ P(x_1 = 9, x_2 = 9), \ldots, P(x_1 = 253, x_2 = 253)\}$$

The resulting total feature vector is 129-dimensional, a manageable size that still captures much of the hiding effect. An example of the feature vector before and after hiding is shown in Figure 3. For comparison to tests on first-order statistics, the histogram of the image under the same hiding is shown.

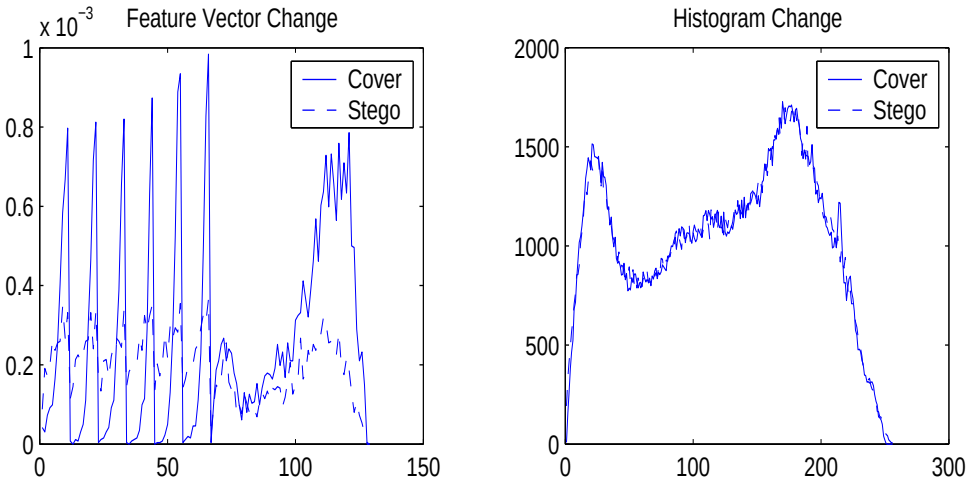

Figure 3. The feature vector (left) of an image changes dramatically after hiding, whereas the histogram (right) hardly changes.

3.2. Results

We summarize the results for globally adaptive hiding in Table 3 and locally adaptive hiding in Table 4. We include both the detection benchmarks: probability of false alarm and probability of missed detection, and learning benchmarks: recall and precision, for convenience. Here globally adaptive schemes are much more susceptible to detection than locally adaptive schemes, and it seems a wise steganographer would exclusively use a locally adaptive technique. However there are a couple of important points to mention:

- Despite an attempt to calibrate the tests by maintaining a benchmark, the message to cover power ratio (MCR), throughout all the tests, there is much more variance in this measure for locally adaptive hiding than globally adaptive hiding. Globally adaptive hiding by definition matches the MCR (with a nominal error due to rounding and clipping), whereas locally adaptive hiding can only choose α such that the chosen MCR is achieved *on average*. The variance from image to image certainly contributes to the difficulties the classifier has in discriminating between clean and stego.

- Ideally the tests would be calibrated to the same message hiding rate, bits per pixel. However it is not necessarily clear that MCR is directly related to the hiding rate. We note that for SSIS, Marvel et al used the very fact that the image could be restored, that is, the message removed, to recover the hidden data. This would certainly be more difficult, if not impossible, with locally adaptive hiding. Indeed, their hiding rate decreased in highly textured images. In these images the cover dependencies are less strong, so the cover statistics are closer to the message statistics, just as in locally adaptive hiding.

Recalling Section 2.3, the divergences for both DCT and spatial hiding are similar and large. Spatial, locally adaptive divergence is smaller, and DCT, locally adaptive is much smaller. The results of our SVM tests follow

Globally adaptive, spatial (SSIS)	
Pr(false alarm)	0.027
Pr(missed detection)	0.017
Recall	0.982
Precision	0.974

Globally adaptive, DCT	
Pr(false alarm)	0.050
Pr(missed detection)	0.006
Recall	0.994
Precision	0.952

Table 3. Globally adaptive hiding, whether in spatial or DCT domain, is properly detected about 95% of the time.

Locally adaptive, Spatial	
Pr(false alarm)	0.118
Pr(missed detection)	0.015
Recall	0.985
Precision	0.893

Locally adaptive, DCT (Cox experiments)	
Pr(false alarm)	0.284
Pr(missed detection)	0.107
Recall	0.893
Precision	0.759

Table 4. As suggested by the Markov random chain tests, locally adaptive hiding is more difficult for a steganalyzer to detect, though spatial is easier than DCT. However, though the average message to cover power ratio is slightly greater than the global hiding tests, it is not clear if this means a similar amount of data can be hidden.

these divergences. That is, the divergence measures on Markov random chains match the relative detectablities of these tests, reinforcing the idea of the divergence as a benchmark. We do however note that our SVM tests are not optimal, and an optimal detector may deviate from these findings.

We note that our results detecting SSIS are very near to Harmsen and Pearlman,[3] despite using different information. I.e. we look at neighboring pixels in the intensity plane, and they look at the same pixel across different color planes. RGB planes are certainly not independent, so this similarity is not surprising.

4. CONCLUSION

We have examined the use of dependency information for the detection of spread spectrum data hiding in grayscale images. A Markov random chain was used to model the correlation between pixels in an image. A detection-theoretic benchmark was used to find the detectability of a few simple methods of spread spectrum hiding in Markov chains, and so to estimate the vulnerability of such hiding to steganalysis. Additionally a gain is found to including dependency information in detection.

To compare the findings for Markov random chains to detection of real image data, we used a supervised learning machine to classify between cover and stego images. The learning machine was trained on a subset of the estimated distributions of a diverse database of images. For detecting SSIS and a similar DCT embedder, the learning machine is able to correctly detect the presence of hidden data about 95% of the time. The performance on other spread spectrum variants is not as powerful, however this is expected from the divergence measured on Markov chains.

This framework can be used to estimate the detectability of other data hiding methods in correlated data, and to design methods of detection. Additionally, more orders of dependence can be modeled. For example, Markov random fields[23] of all neighboring pixels are often used in image modeling. Though this quickly increases the complexity, it may more closely match real images.

Whatever model is used for steganalysis, there is no doubt that dependencies aid detection. We note that with some exceptions,[24] most steganographers do not consider these dependencies, greatly increasing the risk of detection.

ACKNOWLEDGMENTS

This research was supported in part by a grant from ONR #N0014-01-1-0380.

REFERENCES

1. R. Chandramouli, M. Kharrazi, and N. Memon, "Image steganography and steganalysis: Concepts and practices," in *Proceedings of Second Int'l Workshop on Digital Watermarking*, pp. 35–49, 2003.
2. J. Fridrich and M. Golijan, "Practical steganalysis of digital images - state of the art," in *Proceedings of SPIE*, **4675**, 2002.
3. J. J. Harmsen and W. A. Pearlman, "Steganalysis of additive noise modelable information hiding," in *Proceedings of SPIE*, (San Jose, CA), 2003.
4. Y. Wang and P. Moulin, "Steganalysis of block-structured stegotext," in *Proceedings of SPIE*, (San Jose, CA), 2004.
5. C. Cachin, "An information theoretic model for steganography," *Lecture Notes in Computer Science: 2nd Int'l Workshop on Information Hiding* **1525**, pp. 306–318, 1998.
6. P. Moulin and Y. Wang, "New results on steganographic capacity," in *Proceedings of CISS*, 2004.
7. O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath, "Detection of hiding in the least significant bit," *IEEE Trans. on Signal Processing, Supplement on Secure Media I* **52**(10), pp. 3046–3058, 2004.
8. P. Sallee, "Model-based steganography," in *Proceedings of Second Int'l Workshop on Digital Watermarking*, pp. 154–167, 2003.
9. S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," in *Lecture notes in computer science: 5th International Workshop on Information Hiding*, **2578**, 2002.
10. J. Fridrich and M. Goljan, "On estimation of secret message length in LSB steganography in spatial domain," in *Proceedings of SPIE*, (San Jose, CA), 2004.
11. S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB steganography via sample pair analysis," *IEEE Trans. on Signal Processing* **51**(7), pp. 1995–2007, 2003.
12. I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing* **6**(12), pp. 1673–1687, 1997.
13. R. Blahut, *Principles and practice of information theory*, Addison Wesley, 1987.
14. K. L. Chung, *Markov Chains with stationary transition probabilities*, Spring-Verlag, 1960.
15. S. Natarajan, "Large deviations, hypothesis testing, and source coding for finite Markov chains," *IEEE Trans. on Information Theory* **31**(3), pp. 360–365, 1985.
16. L. Marvel, C. G. Boncelet Jr., and C. T. Retter, "Spread spectrum image steganography," *IEEE Trans. on Image Processing* **8**(8), pp. 1075–1083, 1999.
17. V. Poor, *An introduction to signal detection and estimation*, Springer, NY, 1994.
18. S. Lyu and H. Farid, "Steganalysis using color wavelet statistics and one-class support vector machines," in *Proceedings of SPIE*, (San Jose, CA), 2004.
19. K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, "Steganalysis of quantization index modulation data hiding," in *Proceedings of ICIP*, pp. 1165–1168, (Singapore), Oct 2004.
20. I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," in *Proceedings of SPIE*, (San Jose, CA), 2001.
21. A. Martin, G. Sapiro, and G. Seroussi, "Is image steganography natural?." HP labs tech report HPL-2004-39, 7 March 2004.
22. T. Joachims, "Making large-Scale SVM Learning Practical," in *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf, C. Burges, and A. Smola, eds., MIT Press, 1999.
23. A. Rangarajan and R. Chellappa, "Markov random field models in image processing," in *The handbook of brain theory and neural networks*, pp. 564–567, 1995.
24. E. Franz, "Steganography preserving statistical properties," in *5th International Working Conference on Communication and Multimedia Security*, 2002.