Information-theoretic Analysis of Steganalysis in Real Images *

Oleksiy Koval
CVML,CUI - University of Geneva
24, rue General Dufour
1211, Geneva, Switzerland
koval@cui.unige.ch

Svyatoslav Voloshynovskiy
CVML,CUI - University of Geneva
24, rue General Dufour
1211, Geneva, Switzerland
svolos@cui.unige.ch

Taras Holotyak
Department of Radioelectronic devices and systems
Lviv Polytechnic National University
12, Bandery St, Lviv, 79013, Ukraine
t.holotyak@polynet.lviv.ua

Thierry Pun
CVML,CUI - University of Geneva
24, rue General Dufour
1211, Geneva, Switzerland
pun@cui.unige.ch

ABSTRACT

In this paper we consider the problem of performance improvement of non-blind statistical steganalysis of additive steganography in real images. The proposed approach differs from the existing solutions in two main aspects: (a) a locally non-stationary Gaussian model is introduced via source splitting to represent the statistics of the cover image and (b) the detection of the hidden information is performed not from all but from those channels that allow to perform it with the required accuracy. We analyze the theoretically attainable bounds in such a framework and compare them to the corresponding limits of the existing state-of-the-art frameworks. The performed analysis demonstrates the superiority of the proposed approach.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous.

General Terms

Performance.

*(Further information: contact S. Voloshynovskiy, e-mail: svolos@cui.unige.ch)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM&Sec'06, September 26–27, 2006, Geneva, Switzerland.
Copyright 2006 ACM 1-59593-493-6/06/0009 ...$5.00.

Keywords

Steganography, steganalysis, parallel source splitting.

1. INTRODUCTION

Steganography is a branch of information hiding that deals with secret communications. The main goal of steganography is to provide the means of perceptually and statistically undetectable information transmission over some covert channel. According to the "Prisoners' Problem" model proposed in [18] one party (Alice, the steganographer) is attemting to communicate a message to another party (Bob, the decoder). She hides some information into a cover text producing a stego text. The third party (Wendy, the warden) observing the data sent by the steganographer is trying to decide if the sent data are the cover text or stego text. Depending on the passive or active strategy operated by warden in the case the analyzed data are approved to not contain any suspicious information, these data can in the original or corrupted form be transfered to the decoder.

There are two possible attacking strategies that the warden or steganalyser can apply to identify the presense of steganographic information in the communicated data. First, information-theoretic techniques might be exploited for this purpose that requires the availability of the statistics of the cover and stego data to the steganalyzer [3, 20]. The second option consists in the application of various constructive techniques mostly operated without this knowledge and based on the signal processing fundamentals [7, 12].

In this paper we focus on the steganalysis of real image steganography in some transform domain. In particular, we are targeting establishing theoretical performance limits for the steganalyser assuming some non-Gaussian global statistical model of the cover data and availability of some extra side information to the warden and will refer to this technique as to semi-blind steganalysis.

The paper is organized as follows. In Section 2 we will formulate the problem under consideration and briefly review the existing results bounding the information-detection performance of steganalysis systems. Section 3 contains some aspects of stochastic modeling of real images. The analysis of undetectability of additive steganographic systems is performed in Section 4. Finally, Section 5 concludes the paper.

Notations. Capital letters are used to denote scalar random variables X and regular letters x designate their realizations. We use $X \sim f_X(x)$ or simply $X \sim f(x)$ ($X \sim p_X(x)$ or $X \sim p(x)$) to indicate that a continuous (discrete) random variable X is distributed according to $f_X(x)(p_X(x))$. The variance of X is denoted as σ_X^2. I_N stays for an identity matrix of cardinality $N \times N$. The set of integers is denoted by $\mathbb{Z}$.

2. PROBLEM FORMULATION AND PRIOR ART

The problem of hidden information detection on the side of steganalyser might be considered from the information-detection perspectives. First, this approach was adopted by Maurer [13] in authentication application and afterward extended to steganography by Cachin [3] and further elaborated by Wang and Moulin [20].

Accordingly to the assumed setup (Figure 1), the encoder Φ based on the knowledge of the secret key $K = k$ uniformly distributed over the set $\mathcal{K} = \{1, 2, ..., |\mathcal{K}|\}$, a secret message $m \in \{1, 2, .., |\mathcal{M}|\}$ and a realization of the cover data $x^N \in \mathcal{X}^N, X^N \sim f(x^N)$ produces a stego data $Y^N, y^N \in \mathcal{Y}^N$ as $\Phi^N : \mathcal{X}^N \times \mathcal{K} \longrightarrow \mathcal{Y}^N$ in such a way that the embedding distortion constraint $\frac{1}{N} E \,||\, X^N - Y^N \,||^2 \leq P$ is satisfied. One should point out that in more general formulation of a steganographic system design one should satisfy the ϵ-security constraint [3]:

$$D(p(x^N) \,||\, p(y^N)) \leq \epsilon, \tag{1}$$

where $D(f(x^N) \,||\, f(y^N)) = \Sigma_{y^N \in \mathcal{Y}^N} f_{Y^N}(y^N) \frac{f_{Y^N}(y^N)}{f_{X^N}(y^N)}$ denotes the relative entropy or Kullback-Leibler divergence between X^N and Y^N [5]. However, this generalized consideration falls outside of the scope of this paper.

The encoder output is forwarded to the steganalyser Υ that should decide if Y^N is a cover or a stego data. In the scope of this paper following the Kerckhoff's principle [9], we assume that the steganalyser has an access to the embedding function and to $f(x^N)$ but not the secret key k and message m used by the encoder.

Knowing $f(x^N)$, the steganographer performs a statistically optimal binary hypothesis test trying to discriminate between the following two hypotheses:

$$\begin{array}{ll} H_0, & Y^N \sim f(x^N), \\ H_1, & Y^N \sim f(y^N), \end{array} \tag{2}$$

where hypothesis H_0 (H_1) corresponds to the case when no hidden information is communicated via the cover data (a fact of secret communication is approved).

Independently of a particular statistical test used by the steganalyser (Bayessian, minimax, Neyman-Pearson), two types of error might occur. The first one (error of type I or false alarm) corresponds to the situation when an innocuous document is declared to be a stego. In contrast, the error of the second kind (error of type II or miss) takes place when the incorrect decision considering true stego data is made.

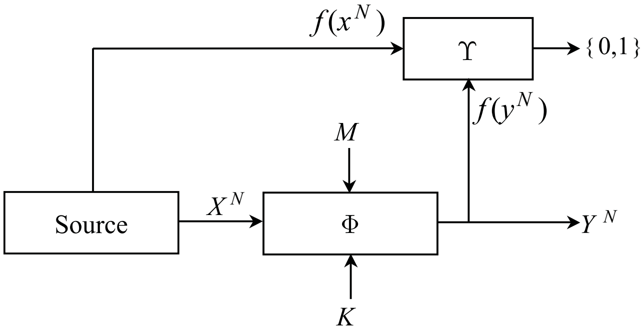

Figure 1: Information-theoretic model of steganography.

Accordingly to the Neyman-Pearson theorem [2] the optimal decision rule for a given maximal tolerable probability of type I error, type II error can be minimized by assuming hypothesis H_0 iff:

$$\ell \triangleq \log_2 \frac{f_{Y^N}(y^N)}{f_{X^N}(x^N)} \tag{3}$$

satisfies:

$$\ell \geq T, \tag{4}$$

for certain threshold T.

Defining error probabilities of type I and II as $P_I \triangleq Pr[\ell > T|H_0]$ and $P_{II} \triangleq Pr[\ell < T|H_1]$, respectively, the following inequality is valid [2]:

$$P_{II} \log \frac{P_{II}}{1 - P_I} + (1 - P_{II}) \log \frac{1 - P_{II}}{P_I}, \\ \leq D(f(y^N) \,||\, f(x^N)). \tag{5}$$

Fixing in (5) $P_{II} = 0$, one obtains a lower bound on P_I that increases with decrease of $D(f(y^N) \,||\, f(x^N))$:

$$P_{II} \leq 2^{-D(f(y^N)||f(x^N))}. \tag{6}$$

In the case, $D(f(y^N) \,||\, f(x^N)) = 0$, the sum probability of error $P_I + P_{II} = 1$ and perfect stegosystem undetectability is attained [3].

In order to achieve perfect undetectability of secret communications in terms of the Bayessian probability of error, $P_E = \pi_I P_I + \pi_{II} P_{II}$, where π_I, π_{II} stay for costs of making the error of type I and II, respectively, the so-called J-divergence, $J = D(f(y^N) \,||\, f(x^N)) + D(f(x^N) \,||\, f(y^N))$ should be minimized [17].

Thus, the main task of the steganographer consists in the minimization of the relative entropy in order to guarantee secret communications undetectability. Oppositely, a large value of $D(f(y^N) \,||\, f(x^N))$ will significantly simplify the task for the steganalyser.

In [20] Moulin and Wang considered the problem of steganographic system detectability quantification. In particular, for several types of embedding strategies that include spread spectrum and modifications of QIM [4], they demonstrated the way of approaching the perfect undetectability by means of optimization of system parameters.

The results derived in [20] were obtained under the Gaussian assumption about the statistics of both the covertext data and embedded signal. Being fundamental, the obtained bounds should be redetermined in any case a statistical assumption deviate from Gaussian. Another constraint of their applicability in practice is closely related to the limitation of real data modeling using Gaussian distribution: it is a known fact that, for instance, real images have highly non-Gaussian statistics in both coordinate and transform domain [8, 11, 14].

Therefore, in order to reveal the real level of steganographic system undetectability that operates with real images, in the sequel Sections we will extend the results obtained in [20] to some realistic stochastic models of the cover data and suggest a new steganalysis strategy.

3. STOCHASTIC MODELING OF REAL IMAGES

The development of accurate and tractable stochastic image models is a very important and challenging problem. The most simple and widely used class of stochastic image models that represents the global stochastic behavior of image coefficients in some transform domains (such as discrete cosine transform or discrete wavelet transform domain) is an i.i.d. Generalized Gaussian (GG) pdf. Laplacian pdf represents a particular case of this model obtained when the GG pdf shape parameter is equal to 1 [22]. A lot of practical image coders and denoisers are designed based on the Laplacian model [15, 22]. However, an even more significant gain can be achieved when the coefficients are considered on the local level. The corresponding procedure of local image coefficients classification based on their statistical properties is known as a *source splitting* [6] that establishes as well the mathematical relationship between local and global stochastic models.

This link can be developed based on the analysis of joint distribution $f(x, \sigma_X^2)$. Using chain rule for probability one obtains:

$$f_{X,\Sigma_X^2}(x, \sigma_X^2) = f_{\Sigma_X^2}(\sigma_X^2) f_{X|\Sigma_X^2}(x|\sigma_X^2), \qquad (7)$$

where $f_{\Sigma_X^2}(\sigma_X^2)$ represents the marginal variance distribution and the conditional distribution $f_{X|\Sigma_X^2}(x|\sigma_X^2)$ is supposed to capture local data statistical behavior. The global data statistics correspond to the marginal distribution:

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X,\Sigma_X^2}(x, \sigma_X^2) d\sigma_X^2 \\ &= \int_0^\infty f_{X|\Sigma_X^2}(x|\sigma_X^2) f_{\Sigma_X^2}(\sigma_X^2) d\sigma_X^2. \qquad (8) \end{aligned}$$

A particular case of interest is given by the *infinite Gaussian mixture model* [21]. According to this model, $f_{X|\Sigma_X^2}(x|\sigma_X^2)$ takes the form of zero-mean conditional Gaussian distribution, i.e., $f_{X|\Sigma_X^2}(x|\sigma_X^2) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{x^2}{2\sigma_X^2}}$. In this case, local image statistics are modeled as Gaussian while the global distribution $f_X(x)$ is obtained by specifying the variance distribution $f_{\Sigma_X^2}(\sigma_X^2)$. In a particular case of the zero-mean Laplacian distribution, $\mathcal{L}(0, \lambda)$ the global Laplacian pdf is obtained as a weighted infinite mixture of zero-mean conditionally Gaussian pdfs given exponentially distributed local variance $f_{\Sigma_X^2}(\sigma_X^2) = \lambda_1 e^{-\lambda_1 \sigma_X^2}$, where λ_1 is a scale parameter of the exponential distribution, and:

$$\begin{aligned} f_X(x) &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{x^2}{2\sigma_X^2}} \lambda_1 e^{-\lambda_1 \sigma_X^2} d\sigma_X^2. \\ &= \sqrt{\frac{\lambda_1}{2}} e^{-\sqrt{2\lambda_1}|x|}, \qquad (9) \end{aligned}$$

where $\lambda = \sqrt{2\lambda_1}$. This simple relationship provides a fundamental link between the global and local statistics of image coefficients. Therefore, the same data can be considered to be locally zero-mean Gaussian with the variance distributed according to the exponential pdf and, simultaneously, having the Laplacian global statistics. The Gaussian mixture model is also the basis for the *parallel channel decomposition* of stochastic image sources that makes it possible to use the simple relationship for the Gaussian statistics. Moreover, properly selecting the variance distribution, one can obtain a general class of GG distributions $f_X(x)$ [1]. However, in the following we will concentrate only on the Laplacian case.

In practice, the number of Gaussian channels is limited by L instead of an infinite number. Therefore, the source X^N is split into L classes according to its variance in the intervals $[0; \sigma_{X_1}^2), [\sigma_{X_1}^2; \sigma_{X_2}^2), \cdots, [\sigma_{X_{L-1}}^2; +\infty)$. The parallel Laplacian source decomposition into L Gaussian channels, $L < \infty$, is schematically explained in Figure 2.

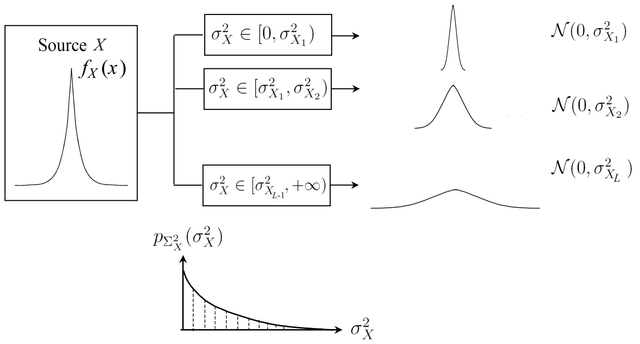

Figure 2: Source splitting as L parallel Gaussian channels.

It is important to underline that one of the state-of-the-art image compression algorithms known as *estimation-quantization* codec [11] as well as its enhanced version [10] exploit this model.[1] In fact, omitting the practical details of side information communications between the encoder and the decoder, Hjorungnes, Lervik and Ramstad [6] were the first who theoretically demonstrated that the rate gain between Laplacian and infinite Gaussian mixture models can be as much as 0.312 bits/sample for high-rate regime. This is achieved by the proper design of entropy coders for each subclass of coefficients that have the same statistics in assumption of the common uniform quantizer.

[1]In fact, practical codec implementations assume that local variances of real image data in the transform domain have locally correlated structure. However, we keep i.i.d. assumption about their distribution in order to simplify our analysis in the scope of this paper.

4. UNDETECTABILITY ANALYSIS OF ADDITIVE STEGANOGRAPHY IN REAL IMAGES

The main goal of this Section consists in the validation of the undetectability of additive steganography in real images. In our analysis we assume that information hiding is performed using spread-spectrum embedding principle:

$$Y^N = X^N + W^N, \tag{10}$$

where the hidden signal is zero-mean Gaussian with variance P, $W^N \sim \mathcal{N}(\mathbf{0}, PI_N)$, and X^N is an i.i.d. vector with its components distributed according to the Laplace distribution $\mathcal{L}(0, \lambda)$ with variance $2/\lambda^2 = N$.

4.1 Non-adaptive steganalysis

In this particular setup it is supposed that embedding of secret information into the cover data is performed globally as well as the detection of secret information on the steganalyser side is based on the global statistics. The relatives entropies between the cover and stego data are given by:

$$D(f(y^N)\,||\,f(x^N)) = \int_{-\infty}^{\infty} f(y^N)\log_2\left(\frac{f(y^N)}{f(x^N)}\right); \tag{11}$$

$$D(f(x^N)\,||\,f(y^N)) = \int_{-\infty}^{\infty} f(x^N)\log_2\left(\frac{f(x^N)}{f(y^N)}\right), \tag{12}$$

where $f(y) = \frac{\lambda}{2}\exp\left(\frac{\lambda^2 P}{2}\right)\left(\exp(-\lambda y)Q\left(\frac{-y+\lambda P}{\sqrt{P}}\right) + \exp(\lambda y)Q\left(\frac{y+\lambda P}{\sqrt{P}}\right)\right)$ is obtained as a convolution of Gaussian and Laplace pdfs of the secret payload and host data, respectively. Unfortunately, the corresponding integrals cannot be computed analytically and divergences $D(f(y^N)\,||\,f(x^N))$ and $D(f(x^N)\,||\,f(y^N))$ were evaluated numerically. In particular, we assumed that $\sigma_X^2 = 2/\lambda^2 = 1$, $P = 0.1$ that give $D(f(y^N)\,||\,f(x^N)) = 0.0108N$ and $D(f(x^N)\,||\,f(y^N)) = 0.0113N$.

Assuming a Bayessian detection framework, equal importance of error probabilities of the I and II kind ($\pi_I = \pi_{II} = 0.5$), the lower bound on the average probability of error for $N = 1000$ is given by [17]:

$$P_E > \frac{1}{4}\exp(-0.5(D(f(y^N)\,||\,f(x^N)) + D(f(x^N)\,||\,f(y^N)))) \approx 3.97 \cdot 10^{-6}. \tag{13}$$

Thus, one can conclude that a steganalyser by means of the optimal binary statistical test can detect the fact of hidden communications performed by the steganographer with the performance that enhances with the length of communicated data N.

4.2 Adaptive steganalysis

Assume that information embedding into the cover image is still performed based on the spread-spectrum principle while the steganalyser has an access to the parameters of the splitting of the Laplace source (a vector of local variances σ_X^{2N}). Having access to the local statistics of the host data, the stegananalyser might exploit this knowledge in the statistical test design. In this case the corresponding expected average probability of error is given by:

$$P_E > \int_0^{\infty} \frac{1}{4}\exp(-0.5(D(f(y^N\,||\,x^N)) + D(f(x^N\,||\,y^N))) \times f_{\Sigma_X^2}(\sigma_X^2)d(\sigma_X^2), \tag{14}$$

where $D(f(y^N)\,||\,f(x^N)) = \frac{N}{2}\left(\frac{P}{\sigma_X^2} - \ln\left(1 + \frac{P}{\sigma_X^2}\right)\right)$, $D(f(x^N)\,||\,f(y^N)) = \frac{N}{2}\left(-\frac{P}{\sigma_X^2+P} + \ln\left(1 + \frac{P}{\sigma_X^2}\right)\right)$ [20] and $f_{\Sigma_X^2}(\sigma_X^2)d(\sigma_X^2) = \frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2\sigma_X^2}{2}\right)$ (Fig. 3). Using numerical integration one obtains for $\sigma_X^2 = 1$, $P = 0.1$ and $N = 1000$:

$$P_E > 0.004 \tag{15}$$

that is significantly higher than in the non-adaptive steganalysis case.

The reason for such a result can be justified by an influence of a significant number of high-variance components in the splitting of Laplace source model that have very low individual relative entropy for the assumed information hiding setup (the value of P). In order to improve the performance, the steganographer might apply an idea that is very close to those one proposed in [16, 19]: not all but only "reliable" parallel channels (with the variance that satisfy $\sigma_X^2 \approx P$) should be used for the detection. In this case, modifying the upper integration limit and renormalizing (14) one obtains:

$$P_E > 5.68 \cdot 10^{-12} \tag{16}$$

for $\sigma_X^2 \leq 0.3$ (Fig. 4).

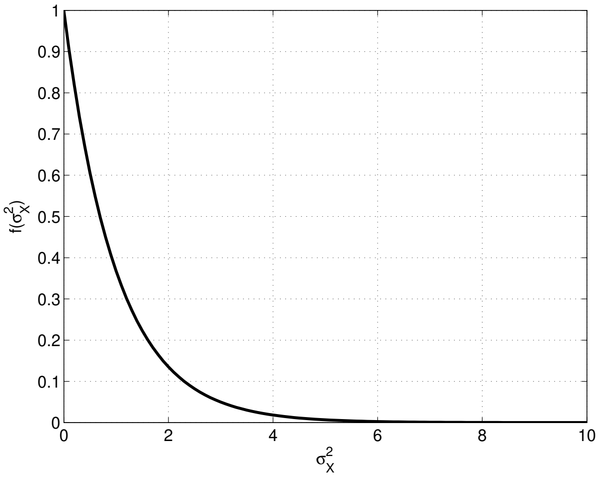

Figure 3: Adaptive steganalysis: local variance probability distribution for the case of detection from all parallel channels.

As an evident consequence driven from the obtained experimental results one can immediately see that the performance of the steganalyser can be enhanced in the case if the detection is performed from the image regions with low variance (flat areas).

Such a strategy immediatly suggests a countermeasure allowing the steganographer to mask the fact of secret com-

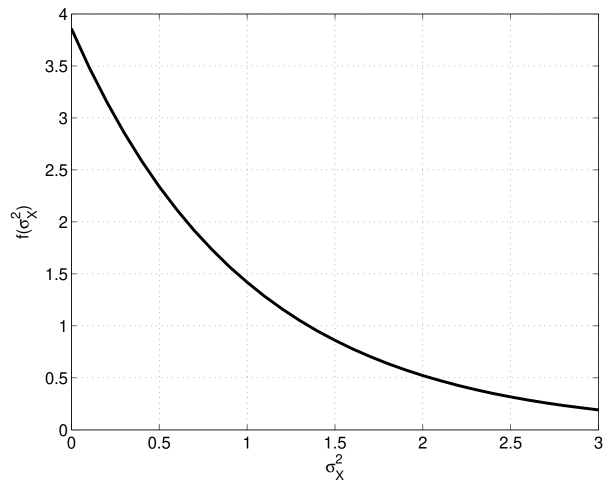

Figure 4: Adaptive steganalysis: local variance probability distribution for the case of detection from "reliable" parallel channels.

munications from the steganalyser by hiding the information into the regions with high variance (edges and textures).

5. CONCLUSIONS

In this paper we considered the problem of steganalysis in real images. In contrast to the existing approaches that mainly analyze the Gaussian setup, we investigated a protocol with the cover data distributed accordingly to the global i.i.d. Laplace pdf or locally non-stationary Gaussian pdf with exponentially distributed local variance. For both statistical assumptions we derived the lower bound on the probability of error for the steganalyser assuming spread-spectrum based embedding of secret information and realized that in the global analysis case the performance is higher. We showed that in order to improve the efficiency of the local analysis, the steganalyser might perform the binary hypothesis test based not on all parallel channels but rather on "relaible" ones with variance commensurable with the variance of the communicated hidden signal.

As possible future research lines we would like to consider the problem of perfect undetectability of the considered class of steganographic systems as well as to extend the performed analysis to the GG host data.

6. ACKNOWLEDGMENT

This paper was partially supported by Swiss National Scientific Foundation Professeur Boursier grant PP002–68653, by the European Commission through the IST Programme under contract IST-2002-507932-ECRYPT and Swiss IM2 projects. The authors thank Jessica Fridrich for a number of stimulating and fruitful discussions. The authors are also thankful to the members of SIP group for many helpful and interesting meetings and seminars.

The information in this document reflects only the authors views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

7. REFERENCES

[1] T. I. Alecu, S. Voloshynovskiy, and T. Pun. The Gaussian transform. In *EUSIPCO 2005, 13th European Signal Processing Conference*, Antalya, Turkey, September 4-8 2005.

[2] R. E. Blahut. *Principles and practice of information theory*. Addison Wesley Publ. Co., 1987.

[3] C. Cachin. An information-theoretic model for steganography. *Lecture Notes in Computer Science*, 1525:306–318, April 1998.

[4] B. Chen and G. Wornell. Dither modulation: a new approach to digital watermarking and information embedding. In *IS&T/SPIE's 11th Annual Symposium, Electronic Imaging 1999: Security and Watermarking of Multimedia Content I SPIE: Security and Watermarking of Multimedia Contents*, volume 3657, pages 342–353, San Jose, California, USA, January 1999.

[5] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley and Sons, New York, 1991.

[6] A. Hjorungnes, J. Lervik, and T. Ramstad. Entropy coding of composite sources modeled by infinite gaussian mixture distributions. In *IEEE Digital Signal Processing Workshop*, pages 235–238, 20-24 January 1996.

[7] T. Holotyak, J. Fridrich, and S. Voloshynovskiy. Blind statistical steganalysis of additive steganography using wavelet higher order statistics. In *9th IFIP TC-6 TC-11 Conference on Communications and Multimedia Security*, volume 2578, Salzburg, Austria, September 19-21 2005.

[8] A. L. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.

[9] A. Kerckhoffs. La cryptographie militaire. *Journal des sciences militaires*, pages 161–191, February 1883.

[10] X. Li. On exploiting geometrical constrants of image wavelet coefficients. *IEEE Trans. on Image Processing*, 12(11):1378–1387, November 2003.

[11] S. LoPresto, K. Ramchandran, and M. Orhard. Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework. In *Data Compression Conference 97*, pages 221–230, Snowbird, Utah, USA, 1997.

[12] S. Lyu and H. Farid. Detecting hidden messages using higher-order statistics and support vector machines. *Lecture Notes in Computer Science*, 2578, April 2002.

[13] U. Maurer. A unified and generalized treatment of authentication theory. In *Proc. 13th Symp. on Theoretical Aspects of Computer Science (STACS'96)*, volume 1046 of *Lecture Notes in Computer Science*, pages 387–398. Springer-Verlag, Feb. 1996.

[14] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, 6(12):300–303, December 1999.

[15] P. Moulin and J. Liu. Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors. In *IEEE Trans. on Information Theory*, volume 45, pages 909–919, Apr. 1999.

[16] P. Moulin and M. K. Mihcak. A framework for evaluating the data-hiding capacity of image sources.

IEEE Trans. on Image Processing, 11(9):1029–1042, September 2002.

[17] H. V. Poor. *An Introduction to Detection and Estimation Theory*. Springer-Verlag, 1994.

[18] G. Simmons. The prisoners' problem and the subliminal channel. In *D. Chaum (Ed.): Advances in Cryptology: Proc. of Crypto'83*, pages 51–67. Plenum Press, 1984.

[19] S. Voloshynovskiy, O. Koval, M. K. Mihcak, and T. Pun. The edge process model and its application to information hiding capacity analysis. *IEEE Trans. on Signal Processing*, page Accepted for publication, May 2004.

[20] Y. Wang and P. Moulin. Steganalysis of block-structured stegotext. In *SPIE: Security and Watermarking of Multimedia Contents*, volume 5306, San Jose, California, January 2004.

[21] R. Wilson. Mgmm: Multiresolution Gaussian mixture models for computer vision. In *ICPR*, pages 1212–1215, 2000.

[22] Y. Yoo, A. Ortega, and B. Yu. Image subband coding using context-based classification and adaptive quantization. *IEEE Trans. on Image Processing*, 8:1702–1215, August 1999.