Neighboring Joint Density-Based JPEG Steganalysis

QINGZHONG LIU, Sam Houston State University
ANDREW H. SUNG and MENGYU QIAO, New Mexico Tech

The threat posed by hackers, spies, terrorists, and criminals, etc. using steganography for stealthy communications and other illegal purposes is a serious concern of cyber security. Several steganographic systems that have been developed and made readily available utilize JPEG images as carriers. Due to the popularity of JPEG images on the Internet, effective steganalysis techniques are called for to counter the threat of JPEG steganography. In this article, we propose a new approach based on feature mining on the discrete cosine transform (DCT) domain and machine learning for steganalysis of JPEG images. First, neighboring joint density features on both intra-block and inter-block are extracted from the DCT coefficient array and the absolute array, respectively; then a support vector machine (SVM) is applied to the features for detection. An evolving neural-fuzzy inference system is employed to predict the hiding amount in JPEG steganograms. We also adopt a feature selection method of support vector machine recursive feature elimination to reduce the number of features. Experimental results show that, in detecting several JPEG-based steganographic systems, our method prominently outperforms the well-known Markov-process based approach.

Categories and Subject Descriptors: I 5.2 [**Pattern Recognition**]: Design Methodology—*Feature evaluation and selection*; I.4.9 [**Image Processing and Computer Vision**]: Applications; K.6.m [**Miscellaneous**]: Security

General Terms: Algorithms, Design

Additional Key Words and Phrases: JPEG, steganography, steganalysis, neighboring joint density, SVM, SVMRFE, nuero-fuzzy, classification

ACM Reference Format:
Liu, Q., Sung, A. H., and Qiao, M. 2011. Neighboring joint density-based JPEG steganalysis. ACM Trans. Intell. Syst. Technol. 2, 2, Article 16 (February 2011), 16 pages.
DOI = 10.1145/1899412.1899420 http://doi.acm.org/10.1145/1899412.1899420

1. INTRODUCTION

Steganography is the secret embedding of information into digital objects such as images, audios, videos, documents, network packets, etc. The innocuous digital media or files are called carriers or covers, the covers embedded with hidden data are called steganograms. As there is little or no perceptible difference between the cover and the steganogram, the potential of exploiting steganography for covert dissemination of malicious software, mobile code, illegal material or protected information is great.

The support for this study from the Institute for Complex Additive Systems Analysis, a division of New Mexico Tech, is greatly appreciated. Q. Liu is also grateful to the Computer Science Department and the Center of Digital Forensics, and to the College of Arts and Sciences, of the Sam Houston State University for their support.

Authors' addresses: Q. Liu, Department of Computer Science, Sam Houston State University, Huntsville, TX 77341; email: liu@shsu.edu; A. H. Sung, Department of Computer Science and the Institute for Complex Additive Systems, New Mexico Tech, Socorro, NM 87801; email: sung@cs.nmt.edu; M. Qiao, Department of Computer Science, New Mexico Tech, Socorro, NM 87801; email: myuqiao@cs.nmt.edu.

DOI 10.1145/1899412.1899420 http://doi.acm.org/10.1145/1899412.1899420

Recent media reports[1] and court documents released by the U.S. Justice Department[2,3] have revealed the first confirmed use of steganography for espionage. Therefore, there is a heightened need for implementing effective countermeasures for steganography, and developing steganalysis techniques with improved efficiency and reliability will be crucial.

Image steganography can roughly be divided into two types: space-hiding directly embeds data in pixel values, for instance, LSB matching [Sharp 2001], improved LSB matching steganography [Mielikainen 2006], and highly undetectable steganography [Pevny et al. 2010]; transform-hiding hides data in the transform coefficients, such as compressed DCT compressed domain [Westfeld 2001; Chang et al. 2007]. Some other information hiding techniques include spread spectrum steganography [Marvel et al. 1999], statistical steganography, distortion, and cover generation steganography [Katzenbeisser and Petitcolas 2000], etc.

Seganalysis generally employs techniques of signal processing, feature mining, and pattern recognition, and aims at detecting the existence of hidden material. To this date, a few popular steganographic systems such as LSB embedding, LSB matching, spread spectrum steganography, etc., have been successfully steganalyzed [Harmsen and Pearlman 2003; Ker 2005; Lyu and Farid 2005; Liu and Sung 2007; Liu et al. 2008a, 2008b, 2008c, 2009a, 2009b, 2010a, 2010b].

JPEG image is one of the most popular media and several steganographic systems for hiding data in JPEG images are available on the Internet, including commercial hiding software CryptoBola, well-known hiding algorithms/tools Outguess [Provos 2001], F5 [Westfeld 2001], steghide [Hetzl and Mutzel 2005], YASS [Solanki et al. 2007], and the improved versions [Sarkar et al. 2008a, 2008b], minimal distortion steganography by using modified matrix encoding (MME) [Kim et al. 2006] and the improved version based on BCH syndrome coding [Sachnev et al. 2009], etc.

To detect the information-hiding in JPEG images using the steganographic algorithm F5, Fridrich et al. [2002] designed a method to estimate the cover-image histogram from the steganogram. To break the recently designed JPEG steganography YASS, Li et al. [2009] proposed a method by extracting the statistical features from locations that are possible to hold embedding host blocks. Yu and Babguchi [2008] performed the detection via pixel and DCT coefficient analysis; Kodovsky et al. [2010] conducted the study by using different feature sets. To design a universal steganalysis method, Harmsen and Pearlman [2004] implemented a detection scheme using only the indices of the quantized DCT coefficients in JPEG images. Fridrich [2004] presented a feature-based steganalytic method for JPEG images. Shi et al. [2007] applied a DCT intra-block-based Markov approach to effectively attacking JPEG steganography, which demonstrated a remarkable advantage over the prior methods. Inspired by Shi et al.'s [2007] work, Pevny and Fridrich [2007] merged the DCT features and calibrated Markov transition probabilities to improve the detection performance. Recently, based on the method in Shi et al. [2007], Chen and Shi [2008] and Liu et al. [2008c] individually expanded the original intra-block Markov approach to inter-block approach. The experimental results demonstrate that the approach integrating intra-block and inter-block Markov transition features outperforms other popular methods [Chen and Shi 2008].

Unfortunately, Chen and Shi [2008] and Liu et al. [2008c] have not elaborated on the reason why the approach is successful. To study this issue, we explored the statistical property of DCT coefficients and found that, most JPEG steganographic systems

[1] http://www.msnbc.msn.com/id/38028696/ns/technology_and_science-science/.

[2] http://www.justice.gov/opa/documents/062810complaint1.pdf.

[3] http://www.justice.gov/opa/documents/062810complaint2.pdf.

modify the DCT coefficients and hence change the correlation of neighboring DCT coefficients. Specifically, embedding of the secret information changes the neighboring joint distribution and, therefore, Markov transition probabilities are affected. Our study also shows that, that Markov approach does not completely represent the neighboring joint relation, that is, it falls short of fully exploring the modification caused by information-hiding [Liu et al. 2009c].

In this article, we present an approach based on the neighboring joint probabilities of the DCT coefficients on intra-block and inter-block. A support vector machine is employed for pattern analysis. Additionally, an evolving neural-fuzzy inference system is used to predict the hiding amount in JPEG steganograms. Experimental results show that, in detecting several JPEG-based steganographic systems, our method prominently outperforms the well-known Markov-process based approach. Meanwhile, our method also obtains good prediction of the hiding amount in the steganograms.

The remainder of the paper is organized as follows: Section 2 studies the neighboring joint statistical model on the DCT coefficients and the modification caused by information-hiding. Section 3 describes our method of feature extraction and an introduction of feature selection method of SVMRFE. Section 4 presents our experiments and analysis, followed by conclusions in Section 5.

2. RELATED STATISTICAL MODELS AND INFORMATION HIDING

Generalized Gaussian distribution (GGD) is widely used in modeling probability density function (PDF) of a multimedia signal. It is very often applied to transform coefficients such as discrete cosine transform (DCT) or wavelet ones [Ohm 2004]. Experiments show that adaptively varying two parameters of the generalized Gaussian distribution (GGD) [Sharifi and Leon-Garcia 1995] can achieve a good probability distribution function (PDF) approximation, for the marginal density of transform coefficients. The GGD model is described as

$$p(x;\alpha,\beta) = \frac{\beta}{2\alpha\cdot\Gamma(1/\beta)}\exp\{-(|x|/\alpha)^{\beta}\}, \tag{1}$$

where $\Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}dt, z > 0$ is the Gamma function, scale parameter α models the width of the PDF peak, and shape parameter β models the shape of the distribution.

Although there does not appear to exist a generally agreed upon multivariate extension of the univariate generalized Gaussian distribution, some researchers define a parametric multivariate generalized Gaussian distribution (MGGD) model that closely fits the actual distribution of wavelet coefficients in clean natural images, exploit the dependency between the estimated wavelet coefficients and their neighbors or other coefficients in different subbands based on the extended GGD model, and achieve good image denoising [Cho and Bui 2005]. The MDDG model is given by

$$p(\mathrm{x}) = \gamma\exp\left\{-\left(\frac{(\mathrm{x}-\mu)^t\sum_{\mathrm{x}}^{-1}(\mathrm{x}-\mu)}{\alpha}\right)^{\beta}\right\}, \tag{2}$$

where γ indicates a normalized constant defined by α and β, $\sum_{\mathrm{x}}$ is the covariance matrix and μ is the expectation vector.

To exploit the dependency between the compressed DCT coefficients and their neighbors, we study the neighboring joint density of the DCT coefficients, and postulate that information hiding will modify the neighboring joint density. Let the left (or upper) adjacent DCT coefficient be denoted by random vector $\mathbf{X}_1$ and the right (or lower) adjacent DCT coefficient be denoted by random vector $\mathbf{X}_2$; let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$. When hidden data are embedded in the compressed DCT domain in JPEG images by using one of the several JPEG steganographic algorithms/tools, the DCT neighboring joint probability

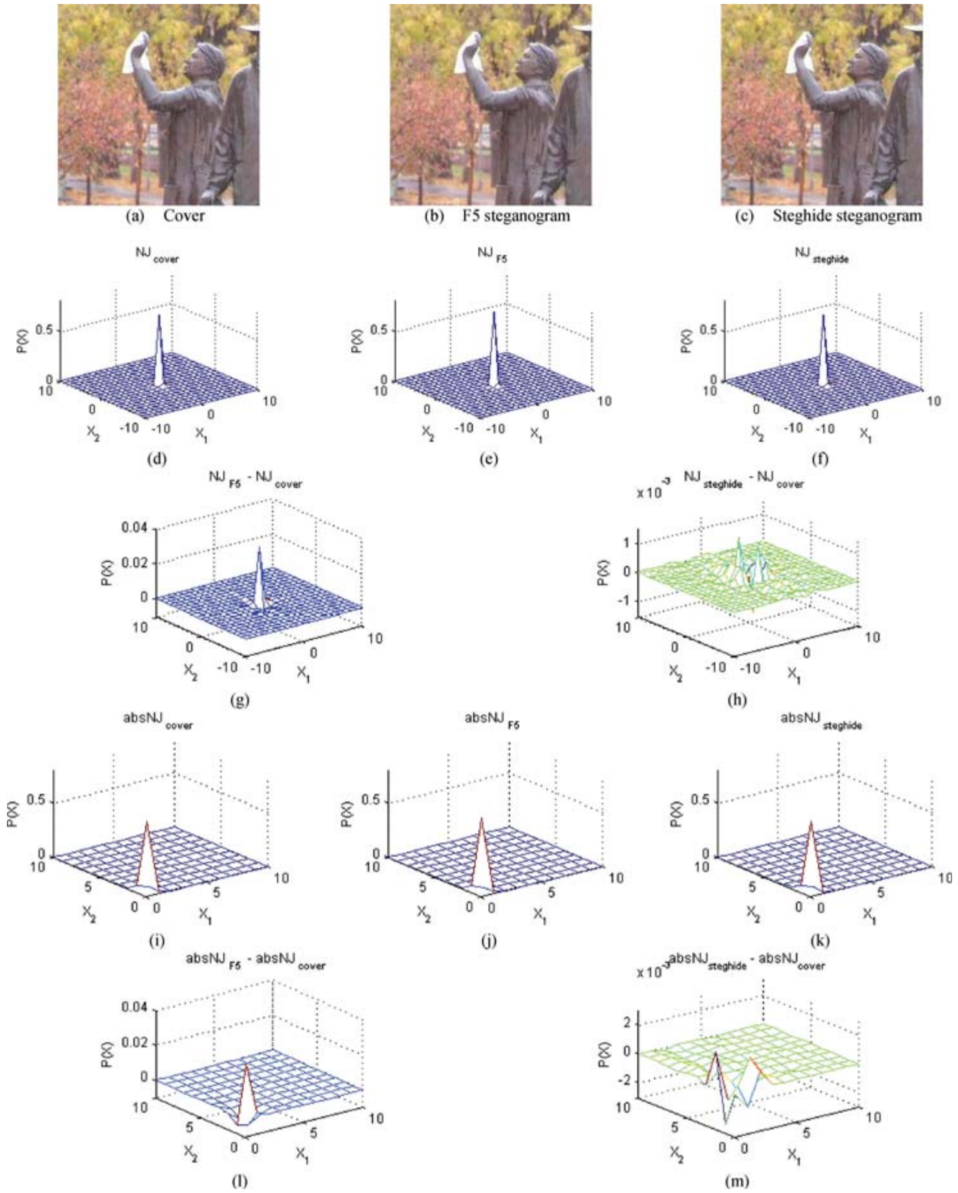
(a) Cover (b) F5 steganogram (c) Steghide steganogram

Fig. 1. Neighboring joint densities of the DCT arrays and the absolute arrays and the density differences.

density is affected, and the change hence leaves a track for steganalysis. An example of the modification of the joint density is illustrated by Figure 1. Figures 1(a) and (d) show a JPEG cover and the compressed DCT neighboring joint density probability, respectively. Figures 1(b) and (e) give the F5 steganogram carrying some hidden data and the neighboring joint density distribution. Figures 1(c) and (f) are the steghide steganogram and the joint density, respectively. Although the three images look identical, the neighboring joint densities are different. The differences are given in Figures 1(g) and (h), indicating that information hiding modifies the neighboring joint density. Figures 1(c), (d), and (e) also demonstrate that the neighboring joint density is approximately symmetric about the origin. Considering this property, we anticipate that the neighboring joint density of the absolute values of the DCT coefficient array will be changed, by the information hiding, which is validated by Figures 1(l) and (m). Figures 1(i), (j), and (k) exhibit the neighboring joint densities of the absolute values of the DCT coefficient arrays.

3. FEATURE EXTRACTION

In our detection algorithm, the neighboring joint features are extracted on intra-block and inter-block from the DCT coefficient array and the absolute array, respectively, described as follows.

3.1. DCT Coefficient Array Based Feature Extraction

3.1.1. Neighboring Joint Density on Intra-Block. Let F denote the compressed DCT coefficient array of a JPEG image, consisting of M × N blocks F_{ij} ($i = 1, 2, \ldots, \mathrm{M}$; $j = 1, 2, \ldots, \mathrm{N}$). Each block has a size of 8 × 8. The intra-block neighboring joint density matrix on horizontal direction NJ_{1h} and the matrix on vertical direction NJ_{1v} are constructed as follows:

$$NJ_{1h}(x, y) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{m=1}^{8} \sum_{n=1}^{7} \delta(c_{ijmn} = x, c_{ijm(n+1)} = y)}{56MN}, \tag{3}$$

$$NJ_{1v}(x, y) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{m=1}^{7} \sum_{n=1}^{8} \delta(c_{ijmn} = x, c_{ij(m+1)n} = y)}{56MN}. \tag{4}$$

where c_{ijmn} stands for the compressed DCT coefficient located at the mth row and the nth column in the block F_{ij}; $\delta = 1$ if its arguments are satisfied, otherwise, $\delta = 0$; x and y are integers. For computational efficiency, we define NJ_1 as the neighboring joint density features on intra-block, calculated as follows:

$$NJ_1(x, y) = \frac{NJ_{1h}(x, y) + NJ_{1v}(x, y)}{2}. \tag{5}$$

Based on the modification of the neighboring joint density caused by information hiding, shown in Figure 1, in our experiment, the values of x and y are in the range of [−6, +6], so NJ_1 has 169 features.

It is worth noting that if the ranges of x and y increase, the number of features will increase greatly. As shown by Figure 1, the modification caused by information hiding is negligible while the values of x and y are away from the origin; in such case, the feature extraction and classification take more time but the detection may not be better. On the contrary, if the ranges of x and y are too small, say, in the range of [−1, 1], the features will be unable to completely reflect the modification caused by information hiding.

3.1.2. Neighboring Joint Density on Inter-Block. The inter-block neighboring joint density matrix on horizontal direction NJ_{2h} and the matrix on vertical direction NJ_{2v} are constructed as follows:

$$NJ_{2h}(x, y) = \frac{\sum_{m=1}^{8} \sum_{n=1}^{8} \sum_{i=1}^{M} \sum_{j=1}^{N-1} \delta(c_{ijmn} = x, c_{i(j+1)mn} = y)}{64M(N-1)}, \tag{6}$$

$$NJ_{2v}(x, y) = \frac{\sum_{m=1}^{8} \sum_{n=1}^{8} \sum_{i=1}^{M-1} \sum_{j=1}^{N} \delta(c_{ijmn} = x, c_{(i+1)jmn} = y)}{64(M-1)N}. \tag{7}$$

We define NJ_2 as the neighboring joint density features on inter-block, calculated as follows:

$$NJ_2(x, y) = \frac{NJ_{2h}(x, y) + NJ_{2v}(x, y)}{2}. \tag{8}$$

Similarly, the values of x and y are in the range of $[-6, +6]$ and NJ_1 has 169 features.

3.2. Absolute DCT Coefficient Array-Based Feature Extraction

3.2.1. Neighboring Joint Density on Intra-Block. Let F denote the compressed DCT coefficient array as before. The intra-block neighboring joint density matrix on horizontal direction $absNJ_{1h}$ and the matrix on vertical direction $absNJ_{1v}$ are given by:

$$absNJ_{1h}(x, y) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{m=1}^{8}\sum_{n=1}^{7}\delta(|c_{ijmn}| = x, |c_{ijm(n+1)}| = y)}{56MN}, \tag{9}$$

$$absNJ_{1v}(x, y) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{m=1}^{7}\sum_{n=1}^{8}\delta(|c_{ijmn}| = x, |c_{ij(m+1)n}| = y)}{56MN}, \tag{10}$$

where c_{ijmn} is the DCT coefficient located at the mth row and the nth column in the block F_{ij}; $\delta = 1$ if its arguments are satisfied, otherwise $\delta = 0$; x and y are integers. For computational efficiency, we define $absNJ_1$ as the neighboring joint density features on intra-block, calculated as follows:

$$absNJ_1(x, y) = \frac{absNJ_{1h}(x, y) + absNJ_{1v}(x, y)}{2}. \tag{11}$$

In our algorithm, the values of x and y are in the range of $[0, 5]$, so $absNJ_1$ consists of 36 features.

3.2.2. Neighboring Joint Density on Inter-Block. The inter-block neighboring joint density matrix on horizontal direction $absNJ_{2h}$ and the matrix on vertical direction $absNJ_{2v}$ are constructed as follows:

$$absNJ_{2h}(x, y) = \frac{\sum_{m=1}^{8}\sum_{n=1}^{8}\sum_{i=1}^{M}\sum_{j=1}^{N-1}\delta(|c_{ijmn}| = x, |c_{i(j+1)mn}| = y)}{64M(N-1)}, \tag{12}$$

$$absNJ_{2v}(x, y) = \frac{\sum_{m=1}^{8}\sum_{n=1}^{8}\sum_{i=1}^{M-1}\sum_{j=1}^{N}\delta(|c_{ijmn}| = x, |c_{(i+1)jmn}| = y)}{64(M-1)N}. \tag{13}$$

We define $absNJ_2$ as the neighboring joint density features on inter-block, calculated as follows:

$$absNJ_2(x, y) = \frac{absNJ_{2h}(x, y) + absNJ_{2v}(x, y)}{2}. \tag{14}$$

Similarly, the values of x and y are in $[0, 5]$ and $absNJ_2$ has 36 features.

4. EXPERIMENTS

4.1. Experimental setup

The 5000 original TIFF raw format digital images used in the experiments are 24-bit, 640 × 480 pixels, lossless true color and never compressed. In accordance with the method of Lyu and Farid [2005], Liu and Sung [2007], and Liu et al. [2008a, 2008b], we cropped these original images into 256 × 256 pixels in order to eliminate the low-complexity parts and converted the cropped images into JPEG format with the default quality. The following six types of steganograms are generated by hiding different data into the 5000 JPEG images with different hiding ratios.

(1) *CryptoBola*. CryptoBola is a commercial information-hiding software that determines the parts (bits) of the JPEG-encoded data that play the least significant role in the reproduction of the image, and it replaces those bits with the bits of the secret message. CryptoBola JPEG is available at http://www.cryptobola.com/.

(2) *JPHS* (*JPHIDE and JPSEEK*). The design objective of JPHS was not simply to hide a file but rather to do this in such a way that it is impossible to prove that the host file contains a hidden file. JPHS for Windows (JPWIN) is available at: http://digitalforensics.champlain.edu/download/jphs_05.zip/.

(3) *Steghide*. Hetzl and Mutzel [2005] designed a graph-theoretic approach for information-hiding based on the idea of exchanging rather than overwriting pixels. Their approach preserves first-order statistics, and the detection on the first order doesn't work.

(4) *F5*. Westfeld [2001] proposed the algorithm F5 that withstands visual and statistical attacks, yet it still offers a large steganographic capacity.

(5) and (6) *Model Based Steganography without Deblocking* (*MB1*) *and with Deblocking* (*MB2*). Sallee [2004] presented an information-theoretic method for performing steganography. Using the model-based methodology, an example steganography method is proposed for JPEG images that achieves a higher embedding efficiency and message capacity than previous methods while remaining secure against first-order statistical attacks.

Some steganogram created by using these steganographic systems and cover examples are shown in Figure 2.

4.2. Comparison of Detection Performance

Chen and Shi [2008] designed an expanded Markov transition probability based feature set, and their study shows that the feature set outperforms other popular feature sets. Specifically, the method applies Markov approach to the differential neighboring coefficients on intra-block and inter-block. In this article, the Markov approach based feature set, DCT coefficient array based neighboring joint density feature sets, defined by Eqs. (5) and (8), and DCT coefficient absolute array-based neighboring density feature sets, defined by Eqs. (11) and (14), are compared with respect to their performance in detecting JPEG steganography. Our approaches are abbreviated as NJ and absNJ, respectively. Table I lists the feature sets in our experiments.

SVM^{light} is an implementation of Vapnik's Support Vector Machine [Vapnik 1998] for the problem of pattern recognition, for the problem of regression, and for the problem of learning a ranking function. The optimization algorithms used in SVM^{light} are described in Joachims [1999, 2000, 2002]. The algorithm has scalable memory requirements and can handle problems with many thousands of support vectors efficiently. In our experiments, we apply SVM^{light} with Radial Basis Function (RBF) kernel as the learning classifier to the features for classification, the kernel parameter is 0.01. In detection of each type of stego-images, three types of feature sets are compared: intra-block, inter-block and both. 100 experiments are performed on each type of feature set. In each experiment, 30% samples are randomly chosen for training and other 70% samples are used for testing. The testing results consist of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The detection performance is evaluated by the classification accuracy, $w \times \text{TP}/(\text{TP} + \text{FN}) + (1 - w) \times \text{TN}/(\text{FP} + \text{TN})$, where w is a weighting factor in the range of [0, 1]. Without losing generality, w has been set to 0.5 in our experiments. The mean values over the 100 testing are shown in Table II.

In this article, the ratio of the number of modified DCT coefficients to the total number of non-zero DCT coefficients is used to measure the embedding strength; the ratio of the number of modified DCT coefficients to the total number of DCT coefficients is used

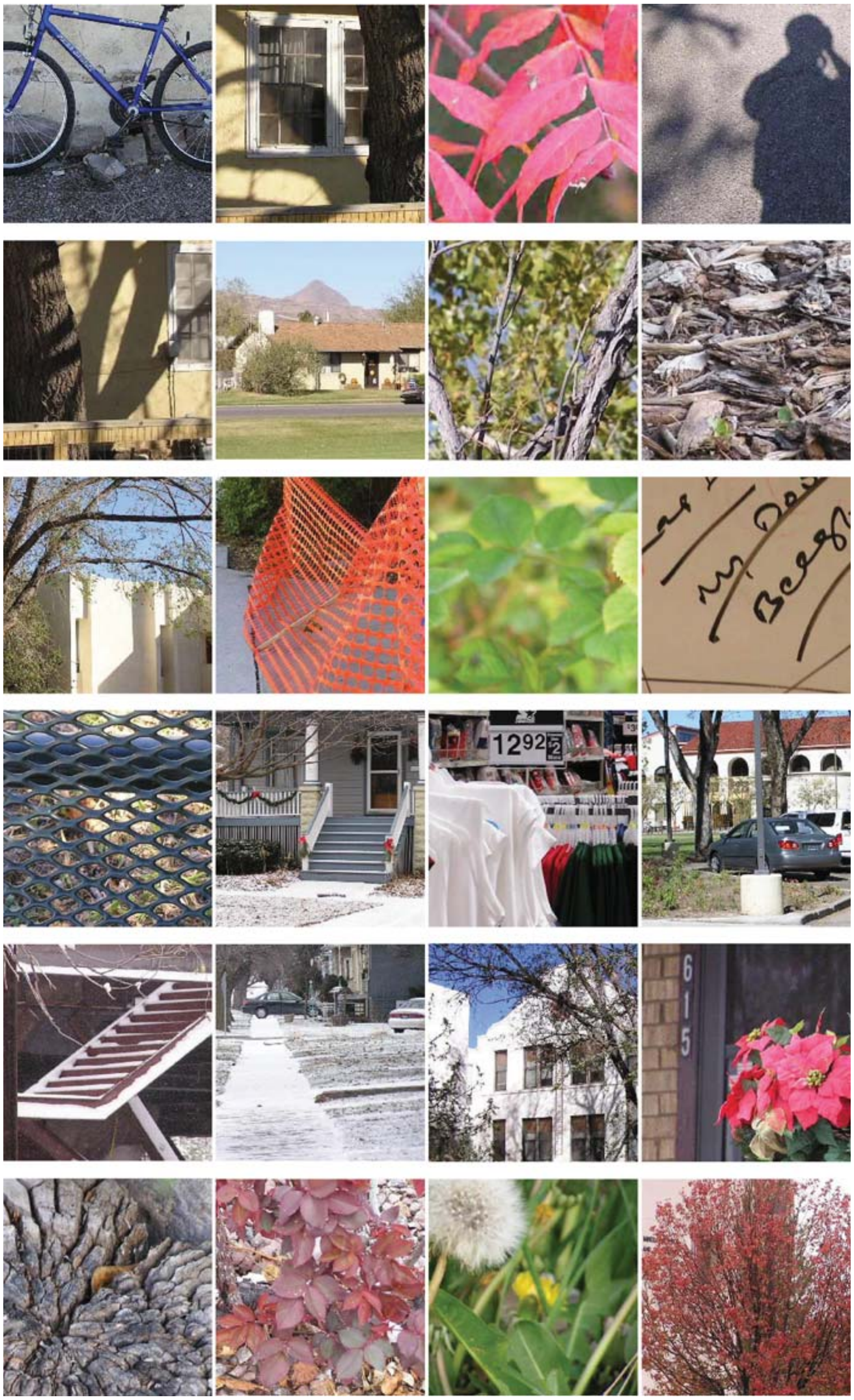

Fig. 2. Steganogram examples produced by using CryptoBola (row 1), F5 (row 2), JPHS (row 3), steghide (row 4), MB1(left two, row5) and MB2 (right two, row 5) and cover examples (row 6).

Table I. Feature Sets Tested in Our Detection

	The number of features		
Detection method	Intra-block feature set	Inter-block feature set	Intra-&inter-block feature set
Markov	324	162	486
NJ	169	169	338
absNJ	36	36	72

Table II. Detection Performance by Using Intra-Block Features, Inter-Block Features, and Total Features (M: Markov Transition Feature Set; NJ: Neighboring Joint Density Feature Set)

		Classification accuracy, 0.5* TP/(TP+FN) + 0.5*TN/(TN+FP)								
		Intra-block features			Inter-block features			Intra-&inter-block features		
Steganographic systems	Average embedding strength	M	NJ	absNJ	M	NJ	absNJ	M	NJ	absNJ
CryptoBola	0.11	**99.2%**	99.1	98.9	96.9	**99.4**	98.9	96.3	**99.4**	99.1
	0.18	99.5	**99.9**	99.8	97.9	**99.9**	99.8	95.7	**99.9**	**99.9**
JPHS	0.10	67.9	**76.8**	76.1	53.2	**75.6**	74.1	57.6	77.1	**78.2**
	0.12	76.5	82.8	**92.8**	62.2	81.2	**90.7**	66.3	83.3	**93.7**
Steghide	0.04	73.8	80.1	**82.8**	66.8	**75.1**	62.3	65.3	81.9	**89.5**
	0.06	83.9	92.7	**95.4**	75.3	**88.7**	76.7	76.0	93.4	**97.1**
F5	0.12	67.2	**82.3**	80.8	56.4	62.2	**74.8**	63.1	**87.6**	85.2
	0.22	83.9	**95.3**	93.2	67.3	78.0	**89.4**	82.4	**97.2**	95.3
MB1	0.09	82.6	89.2	**89.6**	58.2	**76.2**	75.3	78.0	88.6	**89.1**
	0.18	94.6	96.9	**97.1**	74.0	**93.2**	92.7	90.6	**97.4**	97.1
MB2	0.12	86.5	93.3	**94.0**	59.5	**78.7**	76.1	80.9	92.6	**93.5**
	0.24	95.5	98.1	**98.3**	74.5	**93.3**	93.1	92.5	98.1	**98.2**

as modification ratio. With the same amount of hidden data embedded into different JPEG images, the embedding ratio will likely be different since different JPEG images have different number of non-zero DCT coefficients. The average embedding strengths shown in Table II were used for experiments to demonstrate the detection performance under different embedding strength. In comparison of the three types of features, the best result is highlighted in bold. The results show that neighboring joint density based approaches outperform the Markov process based approach; some improvements are as high as over 20%.

Figure 3 demonstrates ROC curves in detecting the six types of steganograms by using absNJ, NJ, and Markov intra- and inter-block features. These results show that the advantage of the neighboring joint density based approaches over the Markov approach is significant and quite noticeable.

4.3. Detection Performance, Image Complexity, and Hiding Amount

As shown by our previous work on steganalysis of LSB matching steganography, even when the same amount of hidden data was embedded in images with the same spatial size, the detection performances under different image complexities will be different. To study the detection performance under different image complexities for image steganalysis, we previously adopted the shape parameter β of the GGD to measure the image complexity and found that image complexity is a significant factor in the evaluation of detection performance. To decrease the computational cost in measuring the image complexity, here we simply take the ratio of the number of non-zero DCT coefficients to the number of total DCT coefficients, including non-zero and zero DCT coefficients, to measure the image complexity. Additionally, we roughly represent the amount of covert message by the ratio of the number of modified DCT coefficients to the total number of DCT coefficients. Due to different hiding methods and different hiding capacity associated with different image complexity, in addition to investigating the

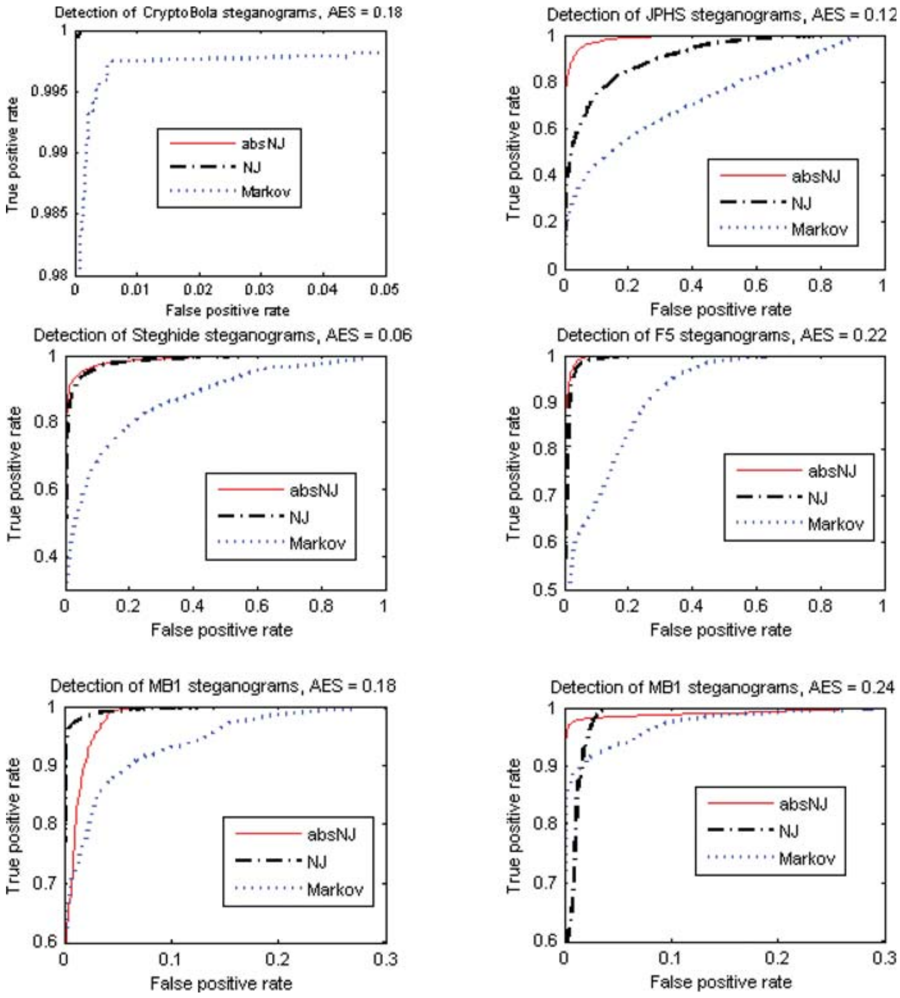

Fig. 3. ROC curves by using total features in absNJ, NJ and Markov approaches.

relation between detection performance and image complexity, it would be more comprehensive and rigorous to explore the relation among detection performance, image complexity, and amount of hidden message.

In machine learning, the Matthews Correlation Coefficient (MCC), which can be calculated by (TP × TN − FP × FN) / sqrt((TP + FP) × (TP + FN) × (TN + FP) × (TN + FN)), is generally used as a balanced measure even if the classes are of very different sizes regarding the quality of binary classification. In Figure 4, the figures on the left demonstrate the MCC values in our detection by using Support Vector Machine; the figures in the middle column show the hiding statistics of the average modification ratio of the DCT coefficients in the steganograms and the covers under different image complexities; the figures on the right exhibit the prediction error of the modification ratio with the use of Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) that was presented in the reference [Kasabov and Song 2002].

It is clear that neighboring joint density based approaches are generally superior to the Markov approach. The results in Figure 4 also demonstrate the relation among

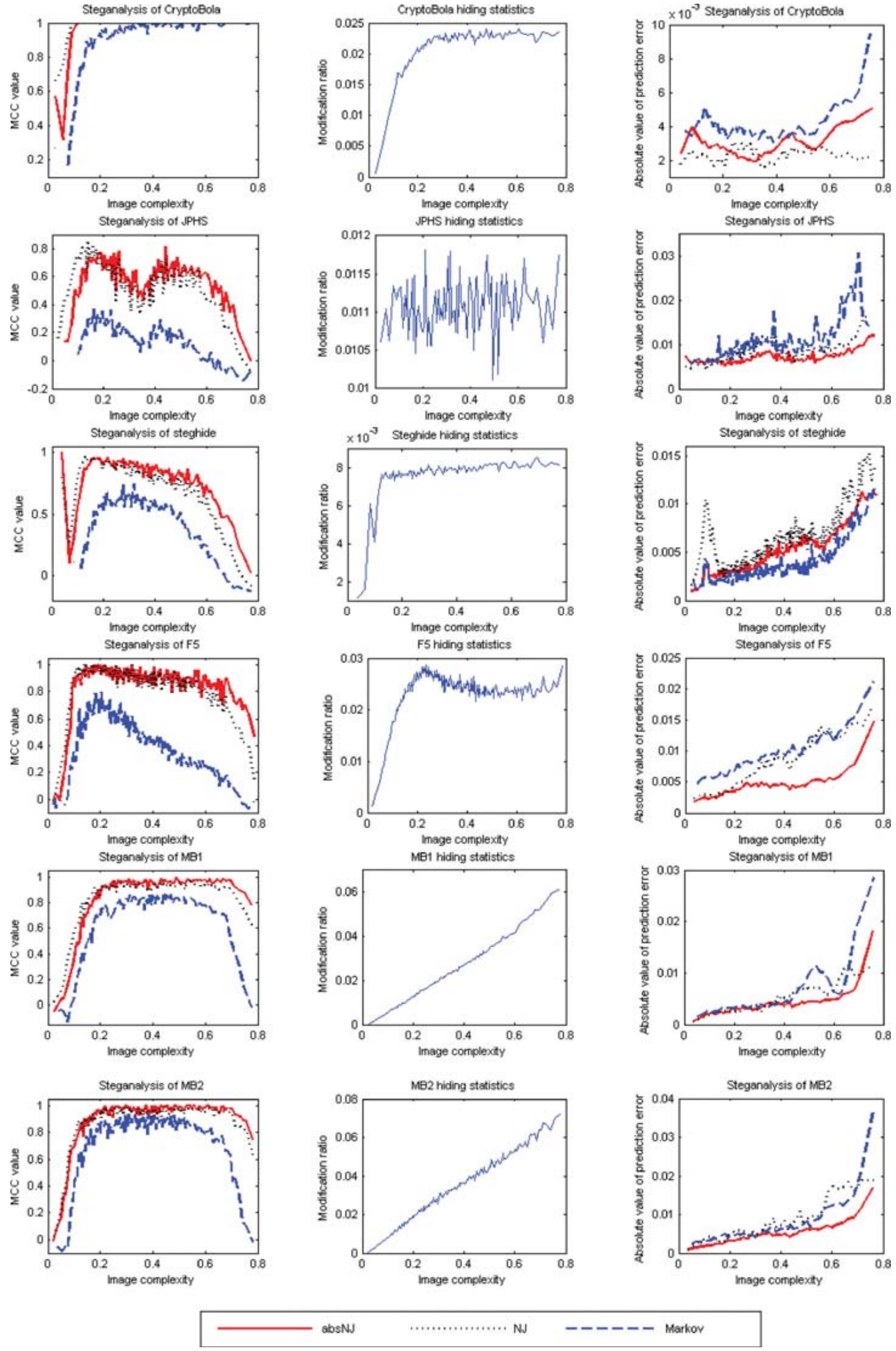

Fig. 4. Detection performance under different image complexities and hiding ratios by using total features in absNJ, NJ, and Markov approaches.

Table III. Mean Value of MCCs

Steganographic system	MCC values under the three approaches		
	Markov	NJ	absNJ
CryptoBola	0.92	**0.99**	**0.99**
JPHS	0.14	0.54	**0.57**
Steghide	0.44	0.75	**0.80**
F5	0.42	0.85	**0.87**
MB1	0.67	**0.86**	**0.86**
MB2	0.75	0.91	**0.92**

Table IV. Mean Values of Absolute Values of Prediction Errors

Steganographic systems	Absolute value of prediction errors		
	Markov	NJ	absNJ
CryptoBola	0.016	**0.009**	0.011
JPHS	0.037	0.032	**0.028**
Steghide	**0.011**	0.020	0.015
F5	0.032	0.023	**0.015**
MB1	0.017	0.014	**0.012**
MB2	0.021	0.022	**0.014**

detection performance, image complexity, and the modification ratio. At low-image complexity, as the modification ratio apparently increases with the increasing of image complexity, the detection performance measured by MCC values increases, which shows that information hiding strength is the significant factor for the evaluation of detection performance. At middle image complexity, as the modification ratio increases (steghide, MB1, and MB2) with the increasing of image complexity, there is no noticeable increasing of the MCC values, some detection performance even drops, which shows that image complexity is an significant factor for the detection evaluation. At high-image complexity, when the modification ratio increases as the image complexity increases, the detection performance deteriorates, and image complexity is a crucial factor for the detection evaluation.

Regardless of the specifics of information hiding ratio and image complexity, Table III lists the means of the MCC values and Table IV gives the absolute values of the prediction errors on the modification ratio, wherein the mean values corresponding to the best detection are highlighted in bold. The results show that absNJ approach performs the best, followed by the NJ approach. Both neighboring joint-density-based approaches are superior to the Markov approach in most cases.

4.4. Feature Selection by SVMRFE

For applications requiring speedy or real-time steganalysis, the task of processing large amounts of JPEG images can be extremely demanding. In such case, the number of features is an important issue since the real-time performance will be negatively affected if the number of features is too large. In our experiments, the performing time by using NJ and Markov approaches is much higher than the computational cost of absNJ approach due to the large number of features of these two approaches. Therefore, we analyze the relation between the detection performance and the number of features that are selected by using the feature selection method, SVMRFE, proposed by Guyon et al. [2002], an application of recursive feature elimination using the support vector weight magnitude as ranking criterion. Figure 5 demonstrates the average detection performance over fifty experiments under different image complexities (here the image complexity is measured by the GGD shape parameter β of the DCT coefficients). Compared to the results in Table II, the comparable detection accuracy is obtained by using even much fewer features.

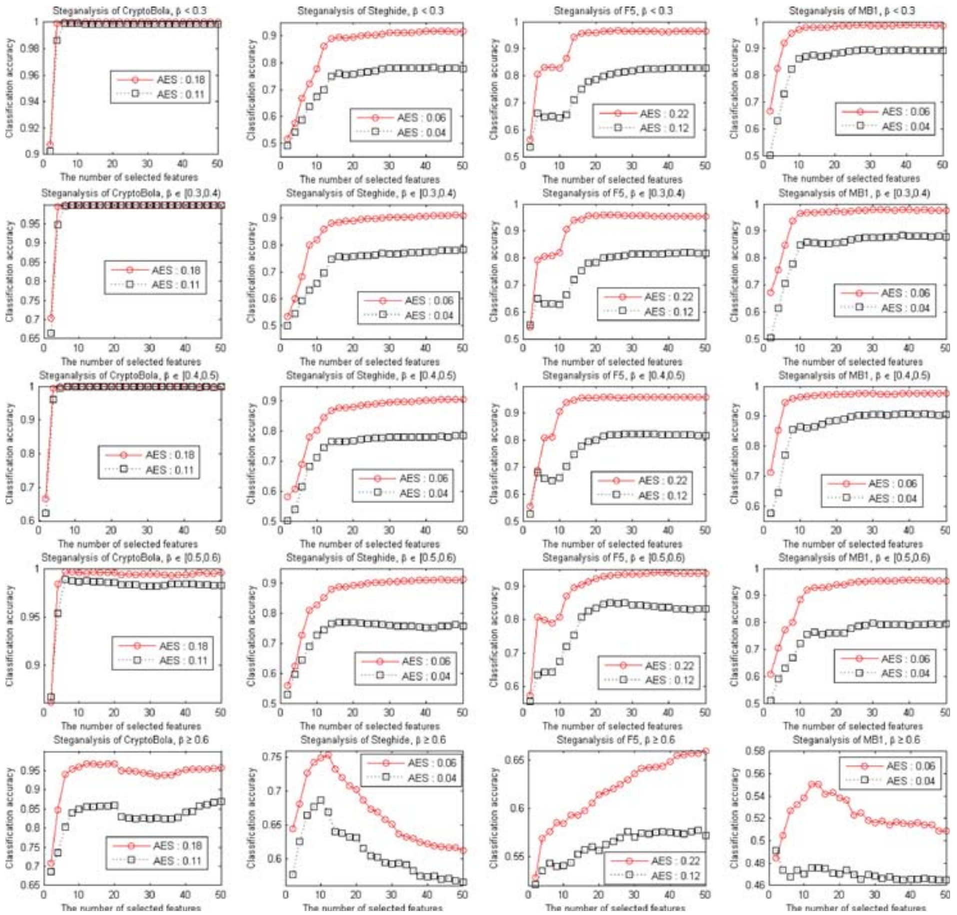

Fig. 5. Detection performance of NJ approach with SVMRFE.

4.5. Discussions

In steganalysis, the Markov process based approach originally proposed by Shi et al. [2007] is highly referenced due to its promising detection accuracy. It has been widely applied, adopted or modified with additional features by several researchers in the field. However, the authors did not elaborate on the reason for the success of the Markov transition probability features. Following their idea, we further explore the modification of the statistical property of the DCT coefficients caused by the information hiding. We found that the neighboring joint density of the DCT coefficients has been modified in several JPEG-based steganographic systems. It is the modification of the neighboring joint density that results in the modification of Markov transition probability; however, the change of the Markov transition probability cannot completely reflect the modification of the neighboring joint density. Therefore, the neighboring joint density features should be generally superior to Markov transition probability features with regard to detection accuracy, which has been verified by our experiments.

For practical applications, the detection time—in addition to detection accuracy—is another major performance issue due to the high-volume of JPEG images; in such case, it is preferable to have a smaller number of features. The two types of neighboring joint density features have the advantage of less processing time compared to the Markov approach−especially for the feature set extracted from the absolute DCT array, where the construction time of the classification model and the detection time are much shorter. Therefore, the two neighboring joint density features possess a great advantage in terms of both detection accuracy and processing time, and they provide an attractive solution to the problem of having to steganalyze large numbers of images with fast processing time and satisfactory detection accuracy.

Besides information-hiding ratio, or embedding strength, image complexity is a critical parameter in the evaluation of detection performance. The detection at high-image complexity is not as reliable as the detection at low image complexity. It should be noted that the neighboring joint density features are also good to detect JPEG double compression, resized JPEG images and spliced images [Liu and Sung 2009].

5. CONCLUSIONS

In this article, we present neighboring joint density-based JPEG steganalysis from the DCT coefficient array and the absolute DCT coefficient array, respectively. Compared to the well-known Markov process based approach, our proposed method remarkably improves the steganalysis performance.

In terms of the two types of neighboring joint density-based approaches, the features extracted from the absolute DCT coefficient array has the advantage of achieving comparable or better detection accuracy with much fewer number of features and less detection time.

For most real-world steganalysis applications, the ability to identify a small or minimal feature set that achieves satisfactory detection performance is a most important evaluation criteria for steganalysis algorithms, this is because of the extremely large number of media files that must be routinely steganalyzed and the fact that feature extraction usually constitutes the most time-consuming step in running the steganalysis algorithm. The results of this article demonstrate that, with judicious and sophisticated feature mining, it is possible to simultaneously achieve smaller feature set, faster detection time, and higher detection performance for JPEG image steganography.

REFERENCES

Avcibas, I., Memon, N., and Sankur, B. 2003. Steganalysis using image quality metrics. *IEEE Trans. Image Proc. 12*, 2, 221–229.

Chang, C., Lin, C., Tseng, C., and Tai, W. 2007. Reversible hiding in DCT-based compressed images. *Inf. Sci. 177*, 13, 2768–2786.

Chen, C. and Shi, Y. 2008. JPEG image steganalysis utilizing both intrablock and interblock correlations. In *Proceedings of the IEEE International Symposium on Circuits and Systems*. 3029–3032.

Cho, D. and Bui, T. 2005. Multivariate statistical modeling for image denoising using wavelet transforms. *Signal Process. Image Commun. 20*, 1, 77–89.

Fridrich, J., Goljan, M., and Hogeam, D. 2002. Steganalysis of JPEG images: Breaking the F5 algorithm. In *Proceedings of the 5th Information Hiding Workshop*. 310–323.

Fridrich, J. 2004. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. Lecture Notes in Computer Science, vol. 3200, 67–81.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. N. 2002. Gene selection for cancer classification using support vector machines. *Mach. Learn., 46*, 1-3, 389–422.

Harmsen, J. and Pearlman, W. 2003. Steganalysis of additive noise modelable information hiding. In *Proceedings of the SPIE,* vol. 5020, 131–142.

Harmsen, J. and Pearlman, W. 2004. Kernel fisher discriminant for steganalysis of JPEG hiding methods. In *Proceedings of the SPIE*, vol. 5306, 13–22.

Hetzl, S. and Mutzel, P. 2005. A graph-theoretic approach to steganography. Lecture Notes in Computer Science, vol. 3677, 119–128.

Joachims, T. 1999. Making large-scale SVM learning practical. In *Advances in Kernel Methods—Support Vector Learning*, B. Schölkopf, C. Burges and A. Smola, Eds., MIT Press.

JOACHIMS, T. 2000. Estimating the generalization performance of a SVM efficiently. In *Proceedings of the 7th International Conference on Machine Learning,* Morgan Kaufman.

JOACHIMS, T. 2002. *Learning to Classify Text Using Support Vector Machines*. Kluwer.

KASABOV, N. K. AND SONG, Q. 2002. DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series. *IEEE Trans. Fuzzy Syst. 10*, 2, 144–154.

KATZENBEISSER, S. AND PETITCOLAS, F. 2000. *Information Hiding Techniques for Steganography and Digital Watermarking.* Artech House Books.

KER, A. 2005. Improved detection of LSB steganography in grayscale images. Lecture Notes in Computer Science, vol. 3200, 97–115.

KIM, Y., DURIC, Z., AND RICHARDS, D. 2006. Modified matrix encoding technique for minimal distortion stegangoraphy. In *Proceedings of the 8th International Workshop on Information Hiding.* Lecture Notes in Computer Science, vol. 4437, 314–327.

KODOVSKY, J., PEVNY, T., AND FRIDRICH, J. 2010. Modern steganalysis can detect YASS. In *Proceedings of the SPIE, Electronic Imaging, Media Forensics and Security XII*, 02-01–02-11.

LI, B., SHI, Y., AND HUANG, J. 2009. Steganalysis of YASS. *IEEE Trans. Inf. Foren. Sec. 4*, 3, 369–382.

LIU, Q. AND SUNG, A. H. 2007. Feature mining and nuero-fuzzy inference system for steganalysis of LSB matching steganography in grayscale images. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence.* 2808–2813.

LIU, Q., SUNG, A. H., CHEN, H., AND XU, J. 2008a. Feature mining and pattern classification for steganalysis of LSB matching steganography in grayscale images. *Patt. Recog. 41*, 1, 56–66.

LIU, Q., SUNG, A. H., RIBEIRO, B. M., WEI, M., CHEN, Z., AND XU, J. 2008b. Image complexity and feature mining for steganalysis of least significant bit matching steganography. *Inf. Sci. 178*, 1, 21–36.

LIU, Q., SUNG, A. H., RIBEIRO, B., AND FERRIERA, R. 2008c. Steganalysis of multi-class JPEG images based on expanded Markov features and polynomial fitting. In *Proceedings of the 21st International Joint Conference on Neural Networks*, 3351–3356.

LIU, Q., SUNG, A. H., AND QIAO, M. 2009a. Temporal derivative-based spectrum and mel-cepstrum audio steganalysis. *IEEE Trans. Inf. Foren. Sec. 4*, 3, 359–368.

LIU, Q., SUNG, A. H., AND QIAO, M. 2009b. Novel stream mining for audio steganalysis. In *Proceedings of the 17th ACM Conference on Multimedia*. 95–104.

LIU, Q., SUNG, A. H., AND QIAO, M. 2009c. Improved detection and evaluation for JPEG steganalysis. In *Proceedings of the 17th ACM Conference on Multimedia*. 873–876.

LIU, Q. AND SUNG, A. H. 2009d. A new approach for JPEG resize and image splicing detection. In *Proceedings of the ACM Multimedia Workshop on Multimedia Forensics*. 43–47.

LIU, Q., SUNG, A. H., QIAO, M., CHEN, Z., AND RIBEIRO, B. 2010a. An improved approach to steganaysis of JPEG images. *Inf. Sci. 180*, 9, 1643–1655.

LIU, Q., SUNG, A. H., AND QIAO, M. 2010b. Derivative based audio steganalysis. *ACM Trans. Multimed. Comput. Comm. Appl.* To appear.

LYU, S. AND FARID, H. 2005. How realistic is photorealistic. *IEEE Trans. Signal Proc. 53*, 2, 845–850.

MARVEL, L., BONCELET, C., AND RETTER, C. 1999. Spread spectrum image steganography. *IEEE Trans. Image Proc. 8*, 8, 1075–1083.

MIELIKAINEN, J. 2006. LSB matching revisited. *IEEE Sig. Proc. Lett. 13*, 5, 285–287.

OHM, J. R. 2004. *Multimedia Communication Technology, Representation, Transmission and Identification of Multimedia Signals*. Springer, Berlin.

PEVNY, T. AND FRIDRICH, J. 2007. Merging Markov and DCT features for multi-class JPEG steganalysis. In *Proceedings of the SPIE*, vol. 6505, DOI: 10.1117/12.696774.

PEVNY, T., FILLER, T., AND BAS, P. 2010. Using high-dimensional image models to perform highly undetectable steganography. In *Proceedings of the 12th International Workshop on Information Hiding*. Lecture Notes in Computer Science, vol. 6387, 161–177.

PROVOS, N. 2001. Defending against statistical steganalysis. In *Proceedings of the 10th USENIX Security Symposium*. 323–335.

SACHNEV, V., KIM, H. J., AND ZHANG, R. 2009. Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome code. In *Proceedings of the 11th ACM Multimedia & Security Workshop*. 131–140.

SALLEE, P. 2004. Model based steganography. Lecture Notes in Computer Science, vol. 2939, 154–167.

SARKAR, A., NATARAJ, L., MANJUNATH, B. S., AND MADHOW, U. 2008a. Estimation of optimum coding redundancy and frequency domain analysis of attacks for YASS—A randomized block based hiding scheme. In *Proceedings of the 15th IEEE International Conference on Image Processing*. 1292–1295.

SARKAR, A., SOLANKI, K., AND MANJUNATH, B. S. 2008b. Further study on YASS: Steganography based on randomized embedding to resist blind steganalysis. In *Proceedings of the SPIE—Security, Steganography, and Watermarking of Multimedia Contents (X)*, vol. 6819. DOI: 10.1117/12.767893.

SHARIFI, K. AND LEON-GARCIA, A. 1995. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Trans. Circuits Syst. Video Technol. 5*, 52–56.

SHARP, T. 2001. An implementation of key-based digital signal steganography. Lecture Notes in Computer Science, vol. 2137, 13–26.

SHI, Y., CHEN, C., AND CHEN, W. 2007. A Markov process based approach to effective attacking JPEG steganography. Lecture Notes in Computer Science, vol. 4437, 249–264.

SOLANKI, K., SARKAR, A., AND MANJUNATH, B. 2007. YASS: Yet another steganogrpahic scheme that resists blind steganalysis. Lecture Notes in Computer Science, vol. 4567, 16–31.

VAPNIK, V. 1998. *Statistical Learning Theory*, Wiley.

WESTFELD, A. 2001. High capacity despite better steganalysis (F5—a steganographic algorithm). Lecture Notes in Computer Science, vol. 2137, 289–302.

YU, X. AND BABAGUCHI, N. 2008. Breaking the YASS algorithm via pixel and DCT coefficients analysis. In *Proceedings of the 19th International Conference on Pattern Recognition.* 1–4.

Received February 2010; revised June 2010 and August 2010; accepted August 2010