I.J. Image, Graphics and Signal Processing, **2014, 8, 19-24**
Published Online July 2014 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2014.08.03

Statistical Image Classification for Image Steganographic Techniques

Seyyed Amin Seyyedi
Belarusian State University Informatics and Radioelectronics/Department of Electronic Computing Machines, Minsk, 220013, Belarus I.A.U Maku Branch/ Department of computer, Maku, Iran
Email: amseyyedi@gmail.com

Nick Ivanov
Belarusian State University Informatics and Radioelectronics/ Department of Electronic Computing Machines, Minsk, 220013, Belarus
Email: ivanovnn@gmail.com

Abstract—Steganography is the method of information hiding. Free selection of cover image is a particular preponderance of steganography to other information hiding techniques. The performance of steganographic system can be improved by selecting the reasonable cover image. This article presents two level unsupervised image classification algorithm based on statistical characteristics of the image which helps Sender to make reasonable selection of cover image to enhance performance of steganographic method based on his specific purpose. Experiments demonstrate the effect of classification in satisfying steganography requirements.

Index Terms—Data Hiding**,** Steganography, Watermark, Cover Image Selection, Stego-image, Steganalysis.

I. Introduction

Nowadays the digital communication channels and Internet play important role in data transmission and sharing, hence there is a great need for security of information. Data hiding is a science that its aim is hiding the information in a media such as image without any remarkable trace on that media [1, 2]. Depending on the relationship between the embedded message and the cover image, data hiding techniques are divided into two categories, digital watermarking and steganographic applications.

Digital watermarking has a close relationship to the cover image such as adding the cover image caption, author signature, and authentication code in it. Applications of steganographic have no relationship to the cover image used for communication. The cover image means nothing to the sender except masking the secret message [3, 4].

Steganography is the art and science of hiding confidential information in an innocuous container that can be a text, image, audio, video, etc. The container is called a file intended to conceal it the confidential information. At present most steganographic methods hide information inside images because of their popularity in internet [2, 4]. The main objective of image steganographic techniques has been to maximize embedding payload while minimizing the distortion rate and detectability of stego-image. But development of steganalysis methods which detect existence of secret message is facing challenges for satisfying the objectives of steganographic techniques [5, 6]. Applying a steganographic technique on two images is not guaranteeing the same results. Selection of suitable cover image is very vital in steganographic methods. It significantly influences the result obtained from the proposed algorithms. Hitherto the cover image selection is not sufficiently investigated in the proposed embedding techniques.

This article presents a new unsupervised image classification algorithm based on edge and texture features of image. This classification helps Sender, choose an appropriate cover image based on his specific purpose in order to enhance steganographic objectives.

The remainder of this paper is organized as follows. Section 2 describes related works. Section 3 introduce proposed image classification algorithm. Experimental results are given in Section 4 and finally, Section 5 concludes the article.

II. Related work

A brief description of some related works is presented in this section.

A. Cover Selection Methods

Z. Kermani [7] was the first introduced the image selection technique for hiding a secret message in it. His method operates on image texture similarity and replaces some blocks of a cover image with similar secret image blocks; then, block location indices of secret image are stored in the cover image. In this method, the blocks of the secret image are compared with the blocks of a set of cover images. The image with the most similar blocks to those of the secret image is introduced as the best candidate to carry the secret image.

H. Sajadi [8] used statistical features of image blocks and their neighborhoods in order to improve Kermani

methods. Using block neighborhood information prevents appearance of virtual edges in the sides and corners of the replaced blocks.

S. Sadkhan [9] proposed an agent system based on some statistical feature of images that helps Sender to choose the best cover image from the image database. After picking up an image from the images database, agent system computes some specific image parameters as histogram, mean, standard deviation and entropy and make a decision on choice. Steganographic agent system tries to find or detect an image with highest variance, maximum contrast, and high entropy.

M. Kharrazi [10] experimentally investigated the problem of cover image selection only for detectability of stego-image by three scenarios in which the Sender has either no knowledge, partial knowledge, or complete knowledge of steganalysis methods.

Y. Sun [11] proposed cover selection method based on correlation coefficient for spatial domain image steganography. The cover data are modeled as Gauss-Markov process, where the correlation coefficient of two arbitrary data elements is the exponent of correlation parameter. The KL divergence and Bhattacharyya distance of Spread Spectrum steganographic system is extended with increasing the correlation parameter. Thus the cover with smaller correlation parameter is selected to improve security.

The above mentioned cover selection methods lack any theoretical background or cannot be generalized to all steganographic requirements. This article presents a new unsupervised image classification algorithm based on edge and texture features of image. This classification helps Sender, choose an appropriate cover image based on his specific purpose in order to enhance steganographic objectives.

B. Steganographic System and Requirements

The concepts of steganography have been agreed at the Information Hiding Workshop in Cambridge [12]. The scheme of steganographic system is shown in fig 1.

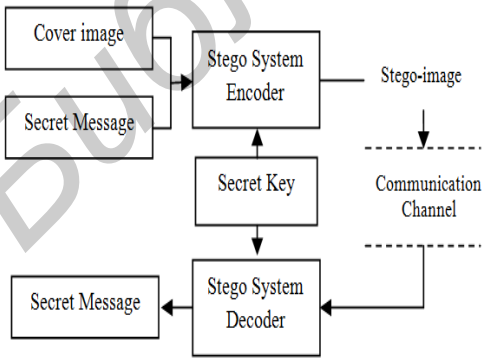

Fig 1 Scheme of steganographic system

Generally, steganographic technique comprises of two steps. At the first step, the selection of input objects of steganographic method, secret message, container and secret key. Any information: text, audio file, image, etc can be used as secret message. This information is called secret message or just message. Choosing the container has a significant impact on the reliability steganographic method and the ability to detect the fact of transfer concealed messages.

Second step determines proper embedding region and then modifies selected regions with secret message bits. Therefore selection of embedding regions is one of the main factors in image steganographic techniques. The choice of embedding regions within cover image mainly depends on the cover image contents. The Human Vision System (HVS) is not very sensitive to changing on the edge and texture regions because these regions are very noisy and a variation in these regions for hiding secret message is difficult to detect. Thus the image with high level of textures and edges satisfies steganographic requirements.

Steganographic methods can be classified into two categories namely spatial-domain techniques and frequency-domain techniques. In spatial domain techniques, the secret messages are embedded directly into a cover image. The simplest spatial domain method is the LSB (Least Significant Bit) approach. In frequency domain methods, the cover image is converted into frequency domain before embedding the secret message in it. A frequency domain method, especially wavelet methods is more secure than other ones [2, 3].

A steganographic method must satisfy three requirements, Payload, Fidelity and Security [13, 14].

1. Payload refers to the amount of information that can be hidden in the cover image. The embedding rate is usually given in absolute measurement such as the size of the secret message or in bits per pixel, etc. It depends on the embedding function, and may also depend on properties of the container.
2. Fidelity (imperceptibility) refers to inability of human eyes to distinguish between cover image and stego-image. The fidelity of stego-image measures by various image similarity metrics such as Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR) and Cross Correlation (CC).

Mean Square Error (MSE) is a simple non-perceptual error metric that is obtained from the cover image C and stego-image S where lower MSE value are assumed to be indicative of lesser detectability. The MSE is calculated using following formula:

$$MSE = \frac{1}{(M \times N)^2} \sum_{i=1}^{M} \sum_{j=1}^{N} (C_{i,j} - S_{i,j})^2. \quad (1)$$

The peak signal-to-noise ratio (PSNR) is the ratio between a signal‘s maximum power and the power of the signal‘s noise. Engineers commonly use the PSNR to measure the quality of reconstructed signals that have been compressed. Signals can have a wide dynamic range, so PSNR is usually expressed in decibels, which is a logarithmic scale. The PSNR calculated using following formula:

$$PSNR = 10\log_{10}\frac{Max^2}{MSE}\,dB, \qquad (2)$$

where Max denotes the maximum pixel value of the image. A higher PSNR value indicates the better quality of stego algorithm. HVS is unable to distinguish the images with PSNR more than 36 dB [14].

Cross-Correlation (CC) is a measure of similarity of cover image C and stego-image S that is computed as:

$$CC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(C_{i,j}-\mu_1)(S_{i,j}-\mu_2)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(C_{i,j}-\mu_1)^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(S_{i,j}-\mu_2)}} \qquad (3)$$

Values μ_1 μ_2 are the mean pixel values of the cover image C and stego-image S.

3. Security of steganographic system is defined in term of undetectability. There are many approaches in defining the security of a steganographic method [4, 15]. J. Zollner [12] theoretically proved that a steganographic system is secure, if secret message has a random nature and is independent from the cover image and stego-image. C. Cachin [16] defined a steganographic method (by Kullback-Leibler KL divergence) to be ε-secure ($\varepsilon \geq 0$), if the relative entropy between probability distribution of cover image (P_C) and stego-image (P_S) are at most ε. Then the detectability $D(P_C \| P_S)$ is defined by:

$$D(P_C \| P_S) = \int P_C \log\frac{P_C}{P_S} \qquad (4)$$

Thus, for a completely secure stego system, $D = 0$ and if $D \leq \varepsilon$, then stego system is named ε-secure.

Increasing payload rate is in conflict with fidelity and security. Sender must make the best tradeoff between requirements.

III. Proposed Image Classification Algorithm

The Sender according to his requirements can improve the steganographic algorithm with selection of suitable cover image. The image containing many specific details is a proper candidate for selection as a container. Image classification algorithm is proposed in two levels. The scheme of proposed image classification is shown in fig 2.

Small patches of image are investigated to detect local features of the image. In the first level image C is divided into 3×3 non overlapping blocks. The Maximum Deviation (MD) of intensity is calculated for each block s:

$$\overline{X} = \frac{1}{9}\sum_{i=1}^{3}\sum_{j=1}^{3} I(i,j), \qquad (5)$$

$$MD(s) = \max_{i,j=1,2,3}\{|I(i,j) - \overline{X}|\}, \qquad (6)$$

where $I(i,j)$ is intensity value of pixel (i,j) within each block of 3×3 dimension and s is the block number.

Number of blocks $NMD(C)$ in the image C with MD greater than given threshold T_1 is counted. These blocks are suspected as non-smooth regions. The threshold T_1 is defined as:

$$T_1 = \alpha \times mean(MD), \qquad (7)$$

where $mean(MD)$ is the mean value of $MD(s)$ for all blocks of image C and α is an accuracy parameter, that $0.5 \leq \alpha \leq 1$.

The first level of proposed classification scheme comprises the following steps:

Input: Image database (DB) containing N grayscale images of the size 512×512.

Output: Four classes K_1, K_2, K_3, K_4 of images.

Step1. Create an array VNC with the size of $2 \times N$ as, elements of $VNC(1,j)$ denotes image number and elements of $VNC(2,j)$ will be assigned with numerical parameter of image C.

Step2. Set the loop control variable $k = 1$. Loop builds numerical characteristic for each image from the database DB and write it to array VNC.

Step3. Get next image C_k from database DB and divide C_k into 3×3 non overlapping blocks.

Step4. Calculate value of $MD(s)$ for each block s by formula (6).

Step5. Calculate $NMD(C_k)$ and assign $VNC(2,j) = NMD(C_k)$.

Step6. Set $k = k+1$. If $k \leq N$, then go to Step 3.

Step7. Sort array VNC in non-descending order based on $VNC(2,k)$ for all $k = 1,2,\ldots,N$.

Step8. Find $M_0 = \min_{k=1,\ldots,N} VNC(2,k)$, $M_4 = \max_{k=1,\ldots,N} VNC(2,k)$,

$$M_i = \left\lfloor i\frac{M4 - M0}{4} \right\rfloor, i = 1,2,3,$$

Step9. Define the elements of classes K_i as:

$$K_i = \{C \in DB \,|\, M_{i-1} < VNC(C) \le M_i\}, i = 1,2,3,4.$$

For constructing classes, $N = 2000$ gray scale images of the size 512×512 are selected randomly from image database BOSS Base (v0.92) [17].

Second level takes into account details of textured regions in images of each class in order to improve the accuracy of classification. Selecting the last elements of each class helps Sender more accurately satisfy steganographic requirements in comparison with the other elements.

The image is divided into 8×8 non overlapping blocks and Entropy $En(s)$ of each block s is calculated.

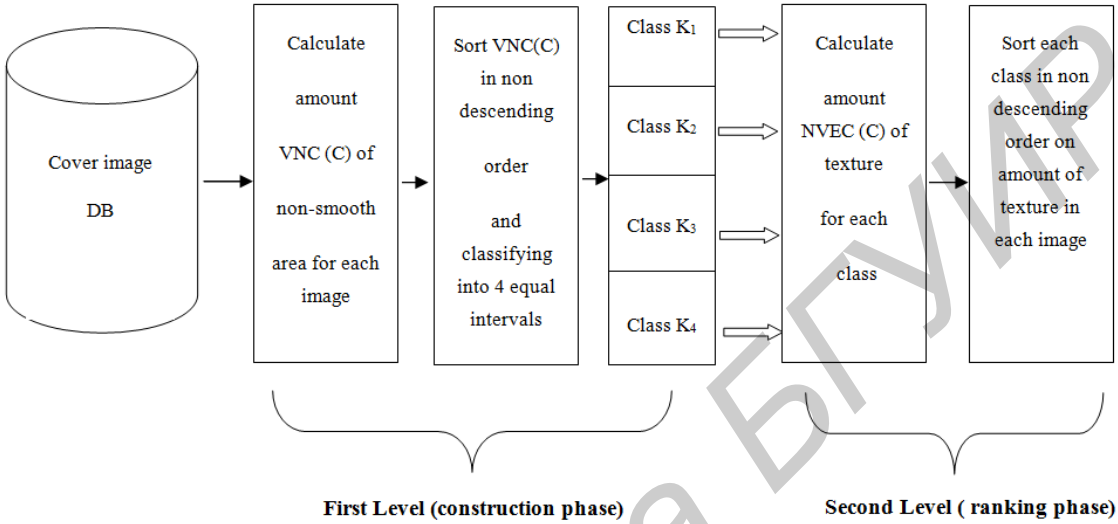

Fig 2 the scheme of classification process

The entropy a block *P* is defined as:

$$En(s) = \sum_{i=1}^{8}\sum_{j=1}^{8} P(i,j)\log_2 P(i,j), \quad (8)$$

where s is block number and $P(i,j)$ probability of pixel intensity.

Number of blocks $NEn(C)$ in the image C with En greater than given threshold $T_2(C)$ is counted. These blocks are suspected as texture regions. The threshold $T_2(C)$ is defined as:

$$T_2(C) = \frac{1}{K}\sum_{s \in C} En(s), \quad (9)$$

$$NEn(C) = \left|s \in C \,|\, En(s) \ge T_2(C)\right|, \quad (10)$$

The second level scheme comprises the following steps:

Input: Images of class.

Output: The same class with enhanced stress on texture.

Step1. Get image C from the class and divide C into 8×8 non-overlapping blocks.

Step2. Calculate $NEn(C)$ and assign it to $VNEC(k)$. Where $VNEC$ and k denotes an array containing NEn and image number respectively.

Step3. Repeat first and second steps for all of images in class.

Step4. Sort the values of $VNEC$ in non-descending order.

IV. Experimental Results

Some experiments were carried out to assess the efficiency of the proposed scheme. The classification has been simulated using the MATLAB 8.1 (R2013a) tools on Windows 7 version 6.1 platform. The proposed classification algorithm was conducted on image database of Granada University [18] and Wisconsin-Madison University [19]. The value of accuracy parameter α =0.7 was selected for classification. The cover images were classified into four classes. The stego-images for all classes were created using Wu methods in spatial domain [20] and Lai method [21] and Seyyedi method [22] in frequency domain. Table 1 compares the maximum payload of each class based on Wu, Lai and Seyyedi methods. As table 1 shows increasing the class number enlarges the payload volume.

PSNR, MSE and Kullback-Leibler (KL) divergence are used as an objective metrics to assess the impact of each class into fidelity and security with fixed message lengths as shown in table 2. These results are approved by three steganographic methods. The correlation of steganography requirements between Lai, Wu and Seyyedi method are approximately equal to 0.988.

According to the experimental results, one can set rules to determine which one of the classes is better for each specific purpose.

- If the main goal of the Sender is high volume of secret message communication, the choices of images from class K_4 are more suitable.
- For security point of view data transmission, the images from classes K_1, K_2 are proper candidate.
- For compromise between the requirements of steganographic requirements (fidelity, data payload and security) the selection cover image from classes K_2, K_3 are more suitable.

V. Conclusion

Visual selection or classification of the container does not satisfy the steganographic requirements. The visually similar images may result in different amounts of payload, fidelity and security. A new image classification algorithm has been proposed for selecting the suitable container. Assigning an image to a certain class means that user is able to predict degree of requirement satisfaction. If the Sender likes to select the image with high payload, he can choose appropriate class in order to provide requested volume. The last elements of each class provide steganographic requirements in the best way. Increasing the number of classes increases the accuracy of algorithm because of decreasing the overlap between classes. High correlation coefficient between Wu, Lai and Seyyedi methods suggests that idea of the image classification may be generalized to other steganographic methods.

Table 1. Comparison maximum payload of each class

Method	Max Payload (Bit)	Class One	Class Two	Class Three	Class Four
Wu	Mean	414948	427067	443922	449386
Lai (K=1)		320668	383377	446683	491946
Seyyedi		390991	453073	511886	535721

Table 2. Comparison fidelity and security metrics of each class

Method	Metrics		Class One	Class Two	Class Three	Class Four
Wu	PSNR	Mean	42.11	41.72	41.11	40.78
	MSE		4.096	4.788	5.44	6.863
	CC		0.9990	0.9990	0.9988	0.9987
	KL		1.79E-04	2.16E-04	2.52E-04	3.71E-04
Lai (K=1)	PSNR	Mean	41.90	40.94	40.32	39.97
	MSE		4.359	5.362	6.284	6.611
	CC		0.9990	0.9987	0.9986	0.9982
	KL		4.32E-04	1.9E-03	5.04E-03	6.07E-03
Seyyedi	PSNR	Mean	43.43	42.64	41.77	41.61
	MSE		3.069	3.685	4.538	4.565
	CC		0.9993	0.9991	0.9990	0.9987
	KL		3.37E-04	1.47E-03	4.07E-03	4.89E-03

References

[1] L. Huang, L. Tseng, and M. Hwang, "A Reversible Data Hiding Method by Histogram Shifting in High Quality Medical Images", *Journal of Systems and Software*, Vol. 86, no. 3, pp. 716-727, 2013.

[2] A. Cheddad, J. Condell, K. Curran, and P.M. Kevitt, "Digital Image Steganography: Survey and Analysis of Current Methods", *Digital Signal Processing*, Vol. 90, No.3, PP. 727-752, 2010.

[3] M. Mohananthini, and G.Yamuna, "A Robustness Image Watermarking Scheme Based Multiresolution Analysis", *International journal of Image, Graphics and Signal Processing*, Vol.11, PP. 9-15, 2012.

[4] B. Li, J. He, J. Huang, and Y.Q. Shi, "Survey on Image Steganography and Steganalysis", *Journal of Information Hiding and Multimedia Signal Processing*, Vol.2, No.2, PP. 142-172, 2011.

[5] A. Nissar, and A.H. Mir, "Classification of Steganalysis techniques", *Digital Signal Processing*, Vol. 90, No. 6, PP. 1758-1770, 2010.

[6] D.C. Lou, and J.L. Liu, "Steganography method for secure communications", *Computers& Security*, Vol. 21, No. 5, PP. 449-460, 2000.

[7] Z. Z. Kermani, and M. Jamzad, "A Robust Steganography Algorithm Based on Texture Similarity UsingGaborFilter", *In IEEE Symposium on Signal processing and Information Technology*, PP. 578-582, 2005.

[8] H. Sajedi, and M. Jamzad, "Cover Selection Steganography Method Based on Similarity of Image Blocks", *In IEEE CIT*, PP. 8-11, 2008.

[9] S.B. Sadkhan, A.M. Al-Barky, and N.N. Muhammad, "An Agent based Image Steganography using Information Theoretic Parameters", *MASAUM Journal of Computing*, Vol. 1, No. 2, PP. 258-264, 2009.

[10] M. Kharrazi, Y.N. Sencar, and N. Memon, "Cover Selection for Steganographic Embedding", *In IEEE image processing*, PP. 117:120, 2006.

[11] Y. Sun, and F. Lui, "Selecting Cover for Image Steganography by Correlation Coefficient", *In IEEE ETCS*, PP. 159-162, 2010.

[12] J. Zollner, at el, "Modeling the Security of Steganographic Systems", *Information Hiding Workshop*, PP.345-355, 1998.

[13] R. Roy, S. Changder, A. Sarkar and N.C Debnath, "Evaluating Image Steganography Techniques: Future Research Challenges", *In ComManTel*, PP.21-24, 2013.
[14] S. Goel, A. Rana, and M. Kaur, "ADCT-based Robust Methodology for Image Steganography", *International journal of Image, Graphics and Signal Processing*, Vol.11, PP. 23-34, 2013.
[15] W. Zhang, and Sh. Li, "Security Measurments of Steganographic System", *Applied Cryptography and Network Security Springer*, Vol.3089, PP. 194-204, 2004.
[16] C. Cachin, "An Information theoretic Model for Steganography", Information and Computation, Vol. 192, No. 1, PP. 41-56, 2004.
[17] http://exile.felk.cvut.cz/boss/BOSSFinal/index.php?mode=VIEW&tmpl=materials.
[18] http://decsai.ugr.es/cvg/dbimagenes/g512.php.
[19] http://homepages.cae.wisc.edu/~ece533/images/
[20] D. Ch. Wu and W.H. Tsi. A Steganographic Method for Images by Pixel-Value Differencing, *Pattern Recognition Letters*, Vol. 24, No. 9, PP. 1613-1626, 2003.
[21] B.L. Lai and L.W. Chang, "Adaptive Data Hiding for Images Based on Harr Discrete Wavelet Transform", *Springer Advances in Image and Video Technology*, PP. 1085-1093, 2006.
[22] S.A Seyyedi, and N. Ivanov, "High Payload and Secure Steganography Method Based on Block Partitioning and Integer Wavelet Transform", *International Journal of Security and Its Applications*, under review.

Authors' Profiles

Seyyed Amin Seyyedi received the M.E in software engineering from Islamic Azad University, Iran 2008. He is a member of computer department in Islamic Azad University. Now he is studying for PhD in Belarusian State University Informatics and Radioelectronics. His research interests include image steganography and watermark.

Nick Ivanov took his PhD degree in applied mathematics from National Academy of Belarus in 1978. Now he is Associate Professor of Belarusian State University Informatics and Radioelectronics. He was supervisor for several Graduate students. His research interests include discrete mathematics, image analysis, and image steganography.