Steganalysis Using KL Transform and Radial Basis Neural Network

Safwan Omar Hasoon **Farhad M. Khalifa**
Dr.safwan1971@yahoo.com
College of Computer Sciences and Mathematics
University of Mosul

Received on: 22/03/2011 **Accepted on: 02/11/2011**

ABSTRACT

The essential problem in the security field is how to detect information hiding. This paper proposes a new steganalysis scheme based on artificial neural network as a classifier to detect information hiding in colored and grayscale images. The statistical features extracted from Karhunen-Loève (KL) transform coefficients obtained from co-occurrence matrix of image. Then radial basis neural network (RBNN) trained using these features to discriminate whether the image contains hidden information or not. This system can be used to prevent the suspicious secret communication.

Keywords: Stegnoanalysis, Neural Network, Steganography

Steganalysis والشبكة العصبية ذات الأساس الشعاعي KL باستخدام تحويل

صفوان عمر حسون **فرهاد محمد خليفة**
كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ الاستلام: 2011/03/22 **تاريخ القبول: 2011/11/02**

الملخص

إن المشكلة الجوهرية في حقل الأمنية هي كيفية اكتشاف المعلومات المخفية. في هذا البحث اقتُرح نظام جديد لكشف الإخفاء في الصور الملونة والرمادية بالاعتماد على الشبكات العصبية الاصطناعية كمصنف. حيث تم استخلاص الخصائص الإحصائية من معاملات تحويل KL المكتسب من مصفوفة co-occurrence للصورة. ثم تُدرب شبكة القاعدة الشعاعية العصبية RBNN باستخدام هذه الخصائص لتمييز فيما إذا كانت الصورة تحوي معلومات مخفية أم لا. هذا النظام يستخدم لمنع الاتصالات السرية غير المرغوب بها.

الكلمات المفتاحية: تحليل Stegnoanalysis ، الشبكة العصبية ، إخفاء المعلومات

1. Introduction

Information hiding has become the focus of research now. This is the art of hiding a message signal in a host signal, such as audio, video, still images and text document without any imperceptible distortion of the host signal [7]. Digital watermarking is a technique which allows an individual to add hidden copyright notices or other verification messages to digital audio, video, or image signals and documents [15]. Steganography is the art of invisible communication between trusted parties. Its purpose is to hide from untrusted parties the fact that any secret message is being communicated [14]. Hence, the main difference between steganography and watermarking is the information that has to be secured. In a steganography technique the embedded information is of much importance where as in watermarking the cover image is important [15].

Steganography may provoke negative effects in the outlook of personal privacy, business activity, and national security. The criminals can abuse the technique for planning illegal activities. For example, commercial spies or traitors may thieve confidential trading or technical messages and deliver them to competitors for a great benefit by using hiding techniques [13]. Steganography is considered secure if the stego-images do not contain any detectable artifacts due to message embedding. In other words, the set of stego-images should have the same statistical properties as the set of cover-images. If there exists an algorithm that can guess whether or not a given image

contains a secret message with a success rate better than random guessing, the steganographic system is considered broken [13].

2. Steganalysis

Contrast to the goal of Information hiding, Steganalysis is the art of discovering and rendering useless such covert messages, hence making information hiding failed [7].

Steganalysis has gained prominence in national security and forensic sciences since detection of hidden messages can lead to the prevention of disastrous security incidents. Steganalysis is a very challenging field because of the scarcity of knowledge about the specific characteristics of the cover media (an image, an audio or video file) that can be exploited to hide information and detect the same. The approaches adopted for steganalysis also sometimes depend on the underlying steganography algorithm(s) used [9].

Automating the detection of hidden messages is a requirement, since the shear amount of image data stored on computers or websites makes it impossible for a person to investigate each image separately [5].

Algorithms for image steganalysis are primarily of two types: Specific and Generic. The Specific approach represents a class of image steganalysis techniques that very much depend on the underlying steganographic algorithm used and have a high success rate for detecting the presence of the secret message if the message is hidden with the algorithm for which the techniques are meant for. The Generic approach represents a class of image steganalysis techniques that are independent of the underlying steganography algorithm used to hide the message and produces good results for detecting the presence of a secrete message hidden using new and/or unconventional steganographic algorithms. The image steganalysis techniques under both the specific and generic categories are often designed to detect the presence of a secret message and the decoding of the same is considered complementary not mandatory [9].

3. Related past work

Detection of hidden information has recently taken the attention of many researchers, it became one of the most attractive research fields for last decade. Liu et al [7] proposed a steganalysis method based neural network, they first extracted statistical features of image transformed to frequency domain using DCT, DFT and DWT, then input these features into neural network to get output. They used back propagation (BP) neural network to train and simulate images.

Patricia Lafferty and Farid Ahmed [12] proposed a method for steganalysis utilizing the local binary pattern (LBP) texture operator to examine the pixel texture patterns within neighborhoods across the color planes is presented. Providing the outputs of this simple algorithm to an artificial neural network in order to discriminate clean images from images altered by data embedded by watermarking and steganographic algorithms.

In [16], Yun Q. Shi et al proposed a general blind steganalysis system, in which the statistical moments of characteristic functions of the prediction-error image, the test image, and their wavelet sub-bands are selected as features. Feed forward back propagation neural network is utilized as the classifier.

In [17], Zhen Zhang et al proposed a steganalysis model, which can be treated as a two-class pattern recognition problem, the model built using image quality metrics (IQMs) and moment features which extracted from the multi-size block discrete cosine

transform (MBDCT). Artificial neural network (ANN) is chosen as a classifier to train and test the given images.

Arezoo Yadollahpour and Hossein Miar Naimi [1] proposed a method to detect LSB stego images by using 2-D autocorrelation coefficients of image. Some of autocorrelation coefficients such as distinctive feature are used then these features are applied for classifying the stego image and natural image by using two different classifiers: SVM classifier and Weighted Euclidean distance (WED).

4. Karhunen-Loève Transform

The KL transform can manipulate a sequence of somewhat correlated measurements into an ordered series of principal components and provide a unique means for noise reduction, feature extraction and de-correlation [6]. The Karhunen-Loève (KL) transform is used to decompose measured signals into uncorrelated empirical basis functions. This optimal set of basis functions represents the main modes in the data, based on the variability information. The data is projected onto this optimal set of basis functions, as opposed to predetermined basis functions such as sines and cosines, as in the case of the standard Fourier transform. This property allows for the detection of nonstationary changes in the signals [4].

The KL decomposition of the input data is performed as follows:

i. The covariance matrix is computed first using the M zero-mean input vectors:

$$\hat{S} = \frac{1}{M}\sum_{j=1}^{M} X_j X_j^Y \qquad \cdots 1$$

ii. The eigenvectors Φ_i and eigenvalues λ_i of the covariance matrix are computed using the following matrix relationship:

$$\hat{S}\Phi = \Phi\Lambda \qquad \cdots 2$$

where: Λ is $diag([\lambda_1 \lambda_2 \ldots \lambda_M].)$.

iii. The coefficient vectors Y_i are computed using the matrix transformation:

$$Y = \Phi^T X \qquad \cdots 3$$

The eigenvalues and eigenvectors are listed in descending order and their energies compared to determine each component's significance [4].

5. Artificial Neural Networks

Artificial Neural Networks (ANNs) have been shown over the past two decades to provide solutions to data mining-type problems that are not easily manipulated by classical solutions [5].

Artificial Neural Networks (ANNs) are recognized as powerful data analysis and modeling tools. They have been shown to capture and accurately represent both linear and nonlinear relationships, and are an invaluable tool for approximating functions, clustering data, and recognizing patterns that are otherwise imperceptible; Neural Networks can often be used for steganalysis [11][12].

5.1 Radial Basis Function Network (RBFN)

The radial basis function network (RBFN), shown in figure (1), is a multidimensional nonlinear function mapping that depends on the distance between the

input vector and the center vector [8]. For an input vector (x_1, x_2, … x_n), a neuron i in the hidden layer produces an output, y_i, given by:

$$y_i = f_r(r_i) \quad \cdots 4$$

$$r_i = \sqrt{\sum_{i=1}^{n} (x_i - w_{ij})^2} \quad \cdots 5$$

where w_{ij} are the weights on the inputs to neuron i, and fr is a symmetrical function known as the radial basis function (RBF). The most commonly used RBF is a Gaussian function:

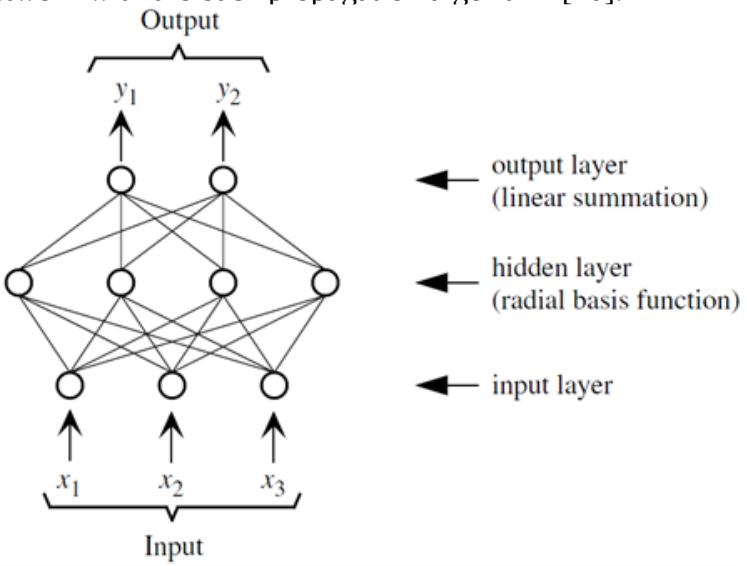
$$f_r(r_i) = exp\left[\frac{-r_i^2}{2\sigma^2}\right] \quad \cdots 6$$

Where: σ_i is the standard deviation. Each neuron, i, in the hidden layer has its own separate value for σ_i [3]. Then calculate the values of the output nodes according to:

$$\varphi_j = \sum_{i=1}^{m} v_{ji} y_i \quad \cdots 7$$

The linear weights associated with the output units of the network tend to evolve on a different "time scale" compared to the nonlinear activation functions of the hidden units. The weight-adaptation process is a linear process compared to the nonlinear parameter adaptation of the hidden-layer neurons. As the different layers of an RBF network are performing different tasks, it is reasonable to separate the optimization of the hidden and output layers by using different techniques. The output layer weights are adjusted according to a simple delta rule [2].

Training in RBFNs is an order of magnitude faster than training of a comparably sized feed forward network with the back propagation algorithm [10].

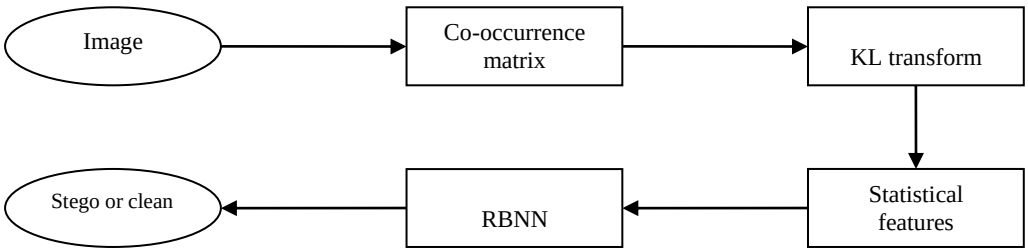


6. Proposed System

The proposed system concerned with the relationship among adjacent pixels, changing any certain pixel affects the relationship of this pixel with its neighbors. So, there are some differences appear when compare the relationship among pixels before and after changing pixels (embedding secret data). But since huge amount of data in

image, and large variety of images, thus the variety of the patterns of data spread in natural images are very huge, so, it is difficult to discern if a certain data spread pattern in image is natural or stego pattern. The challenge is how to find a mechanism that can extract some suitable features that can confer distinction upon natural patterns of data spread and modified data spread patterns.

The proposed system assumes that the message is text embedded using least significant bit (LSB) steganographic algorithm in whole image area, and the image file format is bitmap file format BMP. The proposed system involves the processes as shown in figure (2).



6.1 Data Collection

We have taken two sets of large variety of images (colored and gray images). The two sets contain different images which did not classified in any way, and chosen without any prior knowledge about its statistical features. Each set consists of 726 clean images, and 726 stego images contain hidden information embedded using least significant bit (LSB) algorithm. Each set divided randomly into 70% of images as a training set, and 30% of images as a test set. Figure (3) shows few samples.

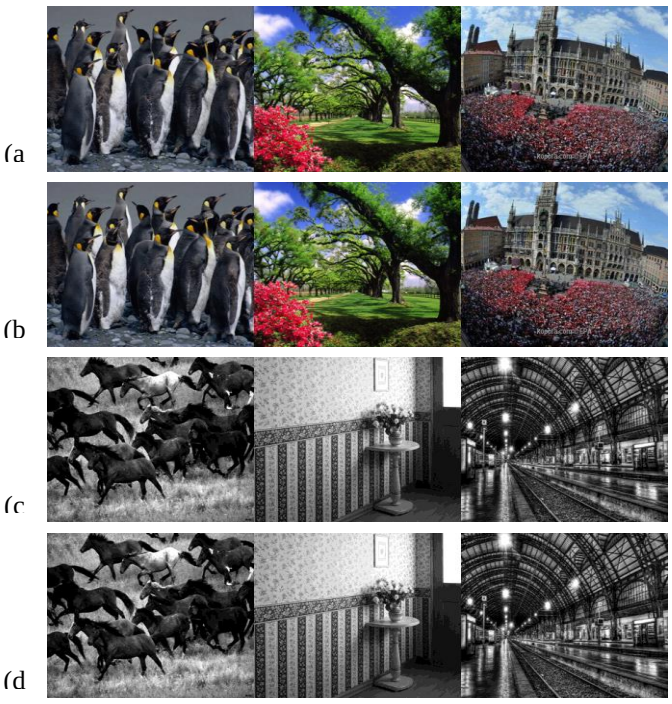
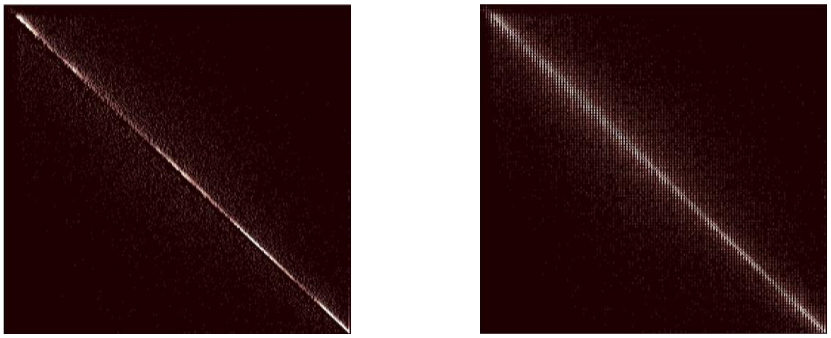


Figure (3). a. Clean Colored Images. b. Stego Colored Images. c. Clean Gray Images. & d. Stego Gray Images.

6.2 Co-occurrence Matrix (cm)

The co-occurrence matrix taken for each image, the co-occurrence matrix of image tests the relationship of adjacent pixels. The test process counts the number of the occurrence of succession of any two specific colors producing 256 X 256 co-occurrence matrix, each element in the matrix represent the succession of the two color values that equal to the row index and the column index of that element. For example if cm(85,83)=62 this means that the number of occurring the succession of the colors 85,83 is 62 times in the whole image. The result is 256 X 256 matrix in which each element is a number represents the co-occurrence of its row index and column index. Figure (4.a) shows the co-occurrence matrix of a clean colored image, while figure (4.b) shows the co-occurrence matrix of a stego colored image using the same image used in figure (4.a), as a cover for the steganography.



(a (b

Figure (4). a. Co-occurrence Matrix of a Clean Colored Image. & b. Co-occurrence Matrix of a Stego Colored Image (Using the Same Image).

6.3 KL-Transform

When the co-occurrence matrix is obtained, the process of kl transform can be implemented by computing the covariance matrix C_i for each column in the co-occurrence matrix, see equation (9)[6], using the mean vector μ that calculated using equation (8)[6],

$$\mu = \frac{\sum_{i=0}^{N-1} X_i}{N} \qquad \cdots 8$$

Where:
X_i is the *i*th column in the co-occurrence matrix,
μ is the mean vector,
N is the number of columns.

$$C_i = (X_i - \mu)(X_i - \mu)^T \qquad \cdots 9$$

Where:
C_i is the *i*th column's covariance matrix.
Then sum these matrices to produce a single matrix (256 X 256).

$$C = \sum_{i=0}^{N-1} C_i \qquad \cdots 10$$

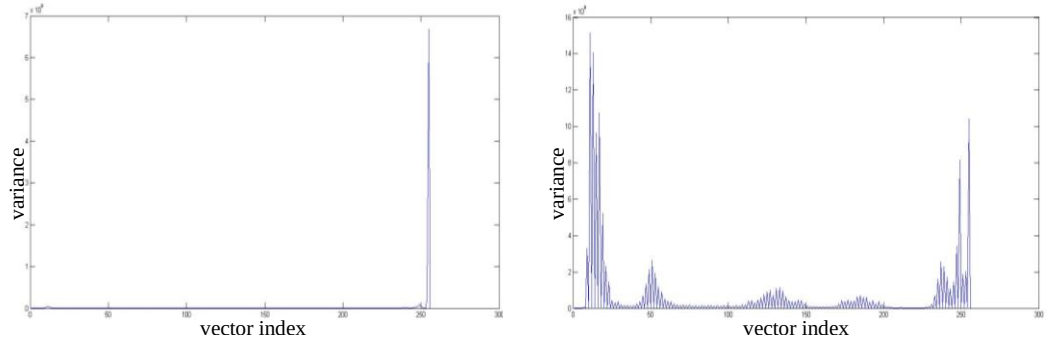
Where:

C is the summation of the covariance matrices.

The eigenvectors Φ and eigenvalues λ obtained from C such as in equation (2). finally the kl transform's coefficient vectors Y_i are computed using the matrix transformation, see equation (3). Kl transform's coefficient is a 256 X 256 matrix in which each column is a single coefficient vector.

6.4 Statistical Features

The statistical features are computed depending on the kl-transform's coefficient vectors. The statistical features in this work involve the calculation of variance, mean and standard deviation. These statistics are row vectors, each element of them corresponds to a coefficient vector in the coefficient matrix. The statistical features vector is a vector of 768 elements. Figure (5.a) shows the variance of kl transform's coefficient of the clean colored image that used in figure (4.a), While figure (5.b) shows the variance of kl transform's coefficient of the stego colored image that used in figure (4.b).



(a) (b)

Figure (5). a. Variance of KL Transform's Coefficient of a Clean Colored Image. & b. Variance of KL Transform's Coefficient of a Stego Colored Image (Using the Same Image).

6.5 Applying RBNN to Detect Information Hiding

The RBNN used to implement the proposed system. The input layer consist of 768 nodes that receive the statistical features, while the hidden layer consist of 600 nodes, finally the output layer consist of a single node. The target of proposed neural network is either +1 indicating to the existence of hidden information in the image, or -1 to represent that the image is clean. Figure (6) shows the architecture of network used to implement the proposed system.

The values of the statistical features varying in the range [0.0017, 0.0039], this is very small variety to be discriminated, therefore, the statistical features preprocessed by normalization so that they fall in the range [-1, 1]. before broad casting to the input layer of the neural network, the normalization process is performed by equation (11).

$$Vn_i = \frac{V_i - V_{min}}{V_{max} - V_{min}} * 2 - 1 \qquad \cdots 11$$

Where:

V_i is the original value,

Vn_i is the normalized value,

V_{min} is the minimum value in the original values,

V_{max} is the maximum value in the original values.

The data normalization applied to each statistical feature vector individually, in order to equalize the effect of each feature vector.

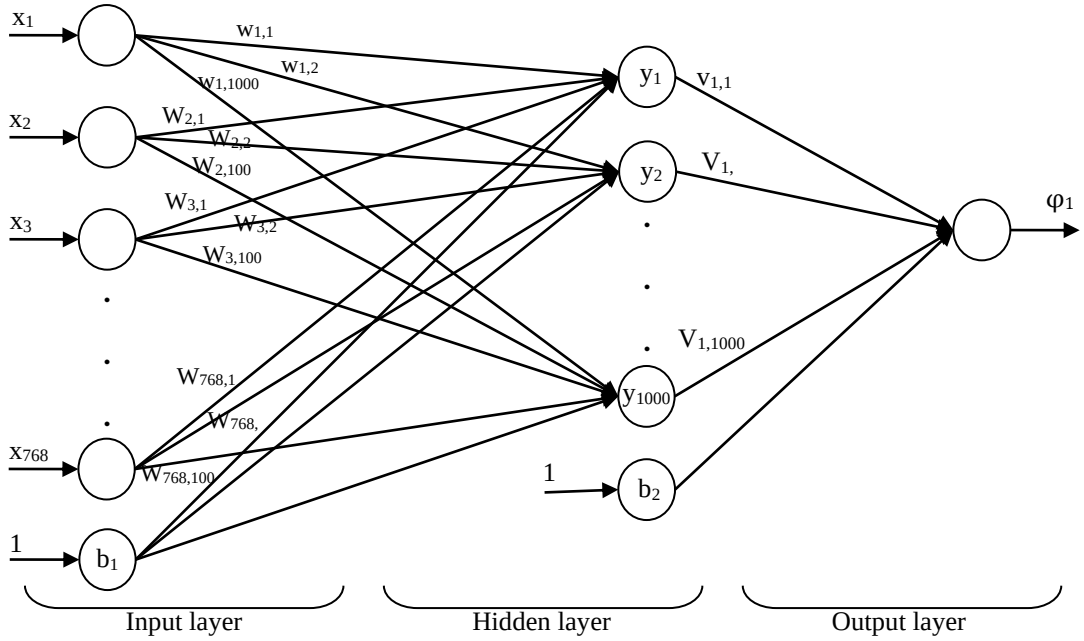


Figure (6). Architecture of RBNN Used to Implement the Proposed System.

7. Results of the Proposed System

The system has detected 96.33 % of the stego colored images correctly (true positive), and detected 91.28 % of the clean colored images (true negative). The average of correct detection using colored images is 93.81 %. Table (1) gives the rates and numbers of classification using colored images. When the proposed system trained using the gray images, it has detected 86.24 % of the stego gray images correctly (true positive), and detected 87.16 % of the clean gray images (true negative). The average of correct detection using gray images is 86.70 %. Table (2) gives the details about the gray images classification rates and numbers.

Table (1). The Result of the Proposed System Using Colored Images

The sets	No. of images in the test set	No. of correctly classified images	Ratio of correctly classified images	No. of incorrectly classified images	Ratio of incorrectly classified images
Clean images	218	199	91.28 %	19	8.72 %
Stego images	218	210	96.33 %	8	3.67 %
Total	436	409	93.81 %	27	6.19 %

Table (2). The Result of the Proposed System Using Gray Images

The sets	No. of images in the test set	No. of correctly classified images	Ratio of correctly classified images	No. of incorrectly classified images	Ratio of incorrectly classified images
Clean images	218	190	87.16 %	28	12.84 %
Stego images	218	188	86.24 %	30	13.76 %
Total	436	378	86.70 %	58	13.30 %

We have trained the texture based steganalysis method in [12] which above mentioned in section 3, with the same data base as in our proposed system. The results were as in table (3) and table (4). Table (3) describes the results of the texture based method trained and tested with colored images, and table (4) describes the results of the same methods trained with gray images.

Table (3). The Result of the Texture Based Method Using Colored Images

The sets	No. of images in the test set	No. of correctly classified images	Ratio of correctly classified images	No. of incorrectly classified images	Ratio of incorrectly classified images
Clean images	218	170	77.98 %	48	22.02 %
Stego images	218	178	81.65 %	40	18.35 %
Total	436	348	79.82 %	88	20.18 %

Table (4). The Result of the Texture Based Method Using Gray Images

The sets	No. of images in the test set	No. of correctly classified images	Ratio of correctly classified images	No. of incorrectly classified images	Ratio of incorrectly classified images
Clean images	218	175	80.28 %	43	19.72 %
Stego images	218	176	80.73 %	42	19.27 %
Total	436	351	80.50 %	85	19.50 %

When the results of the proposed system compared with the results of the texture based steganalysis method under the same conditions of training and testing database, we found that our proposed system achieved the superiority in both situations of colored and gray images. Figure (7) compares the correct (true) detections for both methods with colored images, while figure (8) compares the correct detections for both methods with gray images.

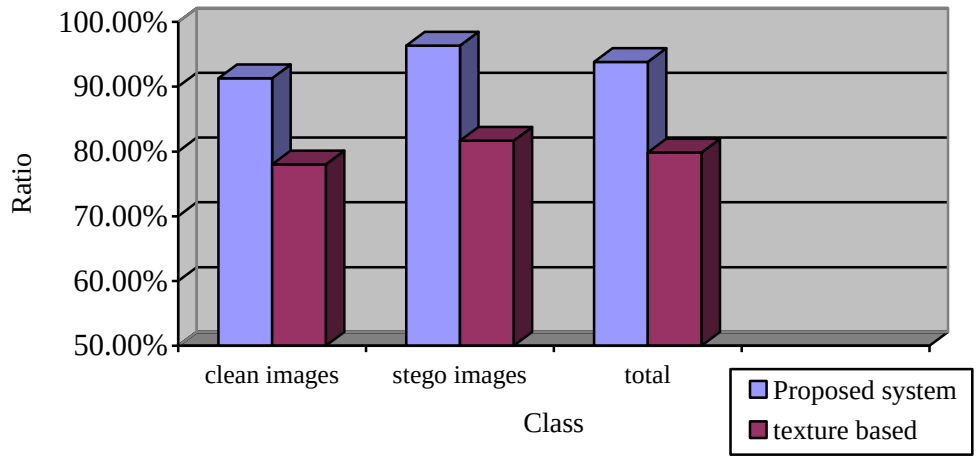


Figure (7). Ratio of Correctly Classified Colored Images

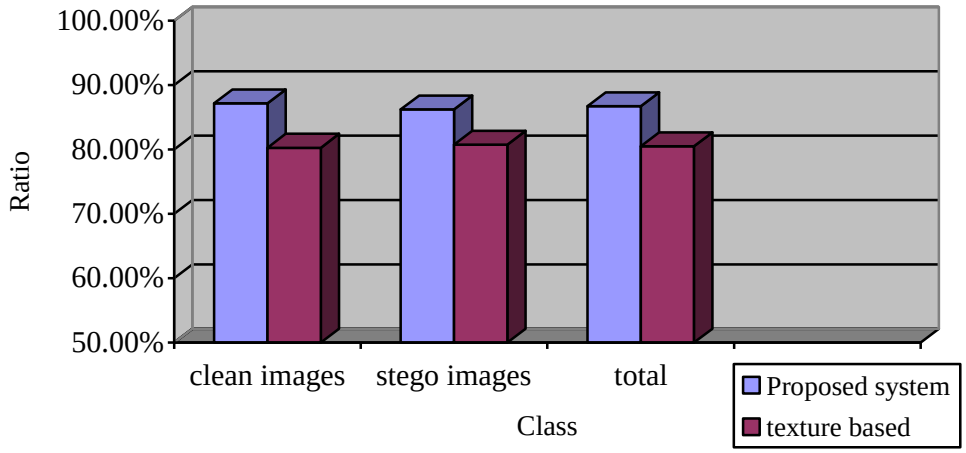


Figure (8). Ratio of Correctly Classified Gray Images

8. Conclusions

This paper provides a new scheme to detect the existence of steganography. The scheme utilize the analyzing power of the kl transform to analyze the co-occurrence matrix of the image to discern if the image contain hidden information or not. The scheme also employ the radial basis neural network as a classifier to classify the statistical features of the kl transform's coefficient.

In comparison with the texture based steganalysis method proposed in [12], the system has proven its activity by giving a superior results using a large variety of images to test its efficiency, see figure (7) and figure (8).

Figure (9) shows samples of misclassified images which incorrectly classified as either stego or clean for both colored and gray images. The most of misclassified images are images adjusted in someway by their producers, like video games and cartoon images. The reason of misclassification may be because the relationship among adjacent pixels in these images differs from that in the other kinds of natural images.

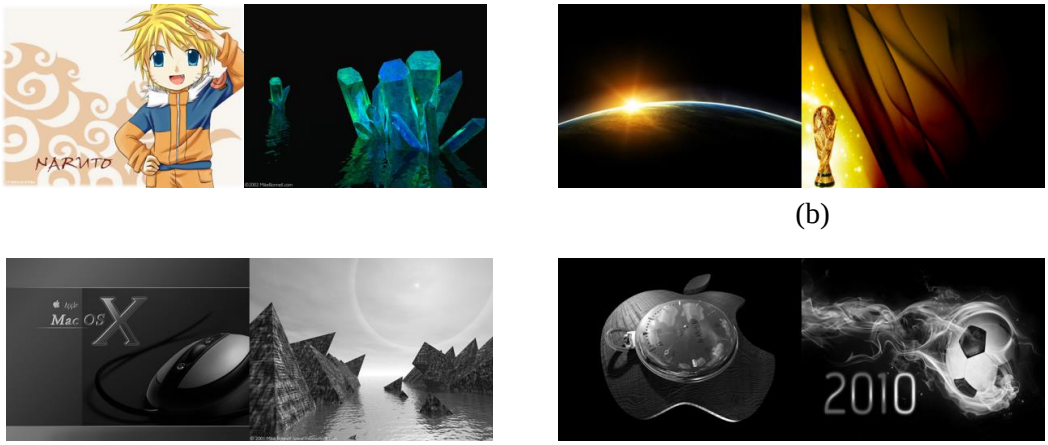


Figure (9). Samples of Misclassified Images: a. Clean Colored Images. b. Stego Colored Images. c. Clean Gray Images. & d. Stego Gray Images.

REFERENCES

[1] Arezoo Yadollahpour and Hossein Miar Naimi, "Attack on LSB Steganography in Color and Grayscale Images Using Autocorrelation Coefficients", European Journal of Scientific Research, vol. 31, no. 2, pp.172-183, 2009.

[2] Bernd Jahne, Horst Haubecker. "Computer Vision and Applications A Guide for Students and Practitioners". Academic Press. 2000.

[3] Hopgood, Adrian A. "Intelligent Systems for Engineers and Scientists", 2nd ed. CRC Press LLC, 2001.

[4] Irem Y. Tumer, Raul G. Longoria, Kristin L. Wood. "Signal Analysis Using Karhunen-Loève Transformation: Application to Hydrodynamic Forces". Transactions of the ASME, vol. 122, August 2000.

[5] Jennifer Davidson, Clifford Bergman, Eric Bartlett. "An Artificial Neural Network for Wavelet Steganalysis". Proceedings of SPIE, vol. 5916, pp. 1-10, 2005.

[6] Jing Wang, Tianfang Li, Hongbing Lu, Zhengrong Liang. "Noise Reduction for Low-Dose Helical CT by Fully 3D Penalized Weighted Least-Squares Sinogram Smoothing", Proc. of SPIE vol. 6142, 2006.

[7] Liu Shaohui, Yao Hongxun, Gao Wen, "Neural Network Based Steganalysis in Still Images", Proceedings of IEEE ICME, vol. 2, pp.509–512, 2003.

[8] Madan M. Gupta, Liang Jin, Noriyasu Homma. "Static and Dynamic Neural Networks from Fundamentals to Advanced Theory", John Wiley & Sons, Inc., Hoboken, New Jersey. 2003.

[9] Natarajan Meghanathan, Lopamudra Nayak. "Steganalysis Algorithms for Detecting the Hidden Information in Image, Audio and Video Cover Media". IJNSA, vol.2, no.1, January 2010.

[10] Nikola K. Kasabov, "Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering". Massachusetts Institute of Technology, Second printing, 1998.

[11] Nouha Kobsi , Hayet Farida Merouani. "Neural Network Based Image Steganalysis: A Comparative Study". JIG'2007 - 3èmes Journées Internationales sur l'Informatique Graphique. pp. 235-240.

[12] Patricia Lafferty, Farid Ahmed, "Texture Based Steganalysis: Results for Color Images", Proc. SPIE, vol. 5561, pp. 145-151,.Aug 2004.

[13] S .Geetha, Dr. N. Kamaraj. "Optimized Image Steganalysis Through Feature Selection Using MBEGA". IJCNC, vol.2, no.4, July 2010.

[14] Tomas Pevny. "Kernel methods in steganalysis", Ph.D. thesis, Computer Science, the Graduate School of Binghamton University, State University of New York, 2003.

[15] Vajiheh Sabeti, Shadrokh Samavi, Mojtaba Mahdavi, Shahram Shirani. "Steganalysis of Embedding in Difference of Image Pixel Pairs by Neural Network". ISC, vol. 1, no. 1, pp. 17-26, January 2009.

[16] Yun Q. Shi, Guorong Xuan, Dekun Zou, Jianjiong Gao, Chengyun Yang, Zhenping Zhang, Peiqi Chai, Wen Chen and Chunhua Chen, "Steganalysis Based on Moments of Characteristic Functions Using Wavelet Decomposition Prediction-Error Image and Neural Network", Multimedia and Expo, Proc. IEEE, 2005.

[17] Zhen Zhang, Yukun Bian and Xijian Ping, "Image Blind Forensics Using Artificial Neural Network", 2008 International Conference on Computer Science and Software Engineering, IEEE, pp. 847-850, 2008.