International Journal of Computer Applications Technology and Research
Volume 3– Issue 7, 439 - 445, 2014, ISSN: 2319–8656

A Review on Classification Based Approaches for STEGanalysis Detection

Anjani Kumar Verma
SPM College, Department Of
Computer Science,
University Of Delhi
New Delhi, India

Abstract: This paper presents two scenarios of image steganalysis, in first scenario, an alternative feature set for steganalysis based on rate-distortion characteristics of images. Here features are based on two key observations: i) Data embedding typically increases the image entropy in order to encode the hidden messages; ii) Data embedding methods are limited to the set of small, imperceptible distortions. The proposed feature set is used as the basis of a steganalysis algorithm and its performance is investigated using different data hiding methods. In second scenario, a new blind approach of image Steganalysis based on contourlet transform and nonlinear support vector machine. Properties of Contourlet transform are used to extract features of images, the important aspect of this paper is that, it uses the minimum number of features in the transform domain and gives a better accuracy than many of the existing steganan-lysis methods. The efficiency of the proposed method is demonstrated through experimental results. Also its performance is compared with the Contourlet based steganalyzer (WBS). Finally, the results show that the proposed method is very efficient in terms of its detection accuracy and computational cost.

Keywords: Steganography, Steganalysis, Contourlet transform, Structural similarity measure, Non linear support vector Machine, MAE, MSE, wMSE, Bayesian

1. INTRODUCTION

Steganography refers to techniques that establish a covert (subliminal) communications channel within regular, innocuous message traffic [1-3]. Steganography is the art and science of hiding secret messages by embedding them into digital media while steganalysis is the art and science of detecting the hidden messages. The goal of a high quality steganography is hiding information imperceptibly not only to human eyes but also to computer analysis.

The obvious purpose of steganalysis is to collect sufficient evidence about the presence of embedded message and to break the security of the carrier. Steganalysis can be seen as a pattern recognition problem also since based on whether an image contains hidden data or not, images can be classified into Stego or Cover image classes.

Steganalysis is broadly classified into two categories. One is meant for breaking a specific steganography. The other one is universal steganalysis, which can detect the existence of hidden message without knowing the details of steganography algorithms used. Universal steganalysis is also known as blind steganalysis and it is more applicable and practicable [4, 5] than the specific steganalysis. Based on the methods used, steganalysis techniques are broadly classified into two classes; signature based steganalysis and statistical based steganalysis. Specific signature based steganalysis are simple, give promising results when message is embedded sequentially, but hard to automatize and their reliability is highly questionable [6, 7]. The first blind steganalysis algorithm to detect embedded messages in images through a proper selection of image quality metrics and multivariate regression analysis was proposed by Avcibas et al. [8, 9]. In universal steganalysis, using statistical methods and identifying the difference of some statistical characteristic between the cover and stego image becomes a challenge. Due to the tremendous increase in steganography, there is a need for powerful blind steganalyzers which are capable of identifying stego images.

In this paper, the focuses are on image steganalysis problem and develop new algorithms based on rate-distortion concepts. In particular, it has been observe the effect of different steganographic methods on image rate-distortion characteristics and construct detectors to separate innocuous cover images from message bearing stego images. This paper proposes a new approach to blind steganalysis does not need any knowledge of the embedding mechanism. This approach utilizes contourlet transform to represent the images. A Gaussian distribution is used to model the contourlet subband coefficients and since skewness and kurtosis of a distribution could be analyzed using the first four moments, the first four normalized statistical moments are considered as the features along with the similarity measure among the medium frequency bands. The experimental results show the efficiency of our approach when analyzed with various steganography methods.

The rest of the paper is organized as below. Section 2 discusses the proposed methods. Section 3 gives Experimental evaluation of the proposed Methods or Steganalyzer with the actual results. At the end, section 5 concludes this paper.

2. PROPOSED METHODS

2.1 METHOD 1

It has been propose novel steganalysis algorithms based on the effect of data hididng process on image rate-distortion characteristics. In particular, we make the following assumptions/observations about the data embedding process:

1. *Data embedding typically increases the image entropy*: In order to encode the hidden messages, steganography methods modify parts of the image data. These modifications typically do not conform with the existing image statistics and therefore result in a net increase in image entropy.

2. *Data embedding methods are limited to the set of small, imperceptible distortions:* Typical steganography methods make only small modifications to ensure perceptual transparency. Perceptually significant parts of the image remain intact.

Stochastic embedding does not result in artifacts with known structures; therefore it is used to develop a generalized method to detect the changes in the rate- distortion characteristics. Since real rate-distortion points for signals with unknown probability distributions –such as natural images – cannot reliably calculated, the data rates achieved by lossy compression scheme. The flowchart of the detection process is seen in Figure. 1.

An image feature extraction phase is followed by a classifier that is trained on relevant data sets. As image features we use the distortion values at different rate points. Mean square error (MSE), mean absolute error (MAE) and weighted mean square error (wMSE) are used as distortion metrics. Here, Bayesian classifier preceded by a KL transform, which reduces the dimensionality of the feature vector.

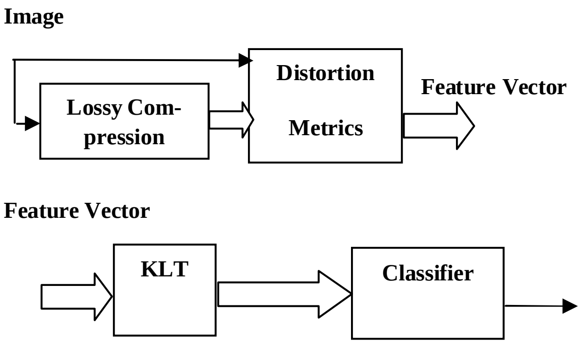

Figure 1. Flow chart describing detection of stochastic embedding

2.1.1 Classifier

Let us denoting w_i as different classes, where each corresponds to a different stego method. This is assuming that $1 \leq i \leq M$ that M such classes exist. This denotes the L dimensional feature vector by x.

$$p(x|w_i) = 1/(2\pi)^{1/2} \, |\textstyle\sum_i|^{1/2} \exp(-1/2(x- \mu_i)^T \textstyle\sum_i^{-1} (x- \mu_i)) \qquad (1)$$

where bf μ_i = E[x]is the mean value of the w_i class and $\sum_i$ is the covariance matrix defined as

$$\textstyle\sum_i = E\,[(x-\mu_i)\,(x-\mu_i)^T] \qquad (2)$$

and $|\sum_i|$ denotes the determinant of $\sum_i$ and E[.] denotes the expected value.
It is also define the discriminant function in the logarithmic form as

$$g_i(x) = \ln(p(x|w_i)P(w_i)) \qquad (3)$$

$$= -1/2\,(x- \mu_i)^T \textstyle\sum_i^{-1} (x- \mu_i)) + \ln(P(w_i)) - 1/2\ln(2\pi) - 1/2\ln(|\textstyle\sum_i|) \qquad (4)$$

Assuming equiprobable classes and eliminating constant terms, Eqn.3 can be reduced to

$$g_i(x) = (x-\mu_i)^T \textstyle\sum_i^{-1}(x-\mu_i)+\ln(|\textstyle\sum_i|) \qquad (5)$$

μ_i and $\sum_i$ are estimated from the training samples for each class during the training phase.

When the classifier has to operate on a limited number of training samples with relatively small number of classes, the high dimensionality of the problem adversely affects the classifier performance. In particular, the covariance matrix becomes nearly singular and classifications results become sensitive to acquisition noise. A method of reducing the dimensionality of the classification problem while keeping the discriminatory power of the feature vector is to project the feature vector onto a proper subspace.

Let us define the within class and between class scatter matrices, S_w and S_b as,

$$S_w = \textstyle\sum^M_{i=1} P_i E[(x-\mu_i)(x-\mu_i)^T] \qquad (6)$$

$$S_b = \textstyle\sum^M_{i=1} P_i(x-\mu_0)(x-\mu_0)^T \qquad (7)$$

where μ_0 is the global mean vector

$$\mu_0 = \textstyle\sum^M_{i=1} p_i \mu_i \qquad (8)$$

This is further define the scattering matrix criterion J_3 as

$$J_3 = \text{trace}\{ S\text{-}^{1}_{w}\, S_b \} \qquad (9)$$

It can now define a linear projection from the L dimensional feature space to N dimensional sub-space.

$$\hat{A} = C^T x \qquad (10)$$

The optimal projection matix w.r.t. the scattering matrix criterion J_3 is the eigenvectors corresponding to the largest eigenvalues of the system $S\text{-}^{1}_{w}\, S_b$. As the individual scatter matrices S_i, the within class scatter matrix may also be ill conditioned. Therefore, in practice it has been used the pseudo-inverse of S_w in the calculations.

2.2 METHOD 2

The objective of the proposed scheme is to select the most relevant features using statistical characteristics of the subband coefficients, thus reduce the dimensionality of feature set and increase the accuracy of detection. In this paper, the first four normalized moments of high frequency, low frequency subband coefficients and structural similarity measure of medium frequency sub band coefficients are taken as the feature set. With these five features, a Non linear Support Vector Machine is trained for further classification. The block diagram of the proposed model is given in Figure 2. The following sub sections briefly explain contourlet transformation and how the feature set is extracted from images.

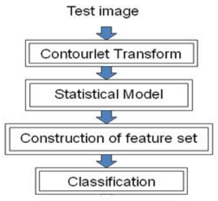

Figure 2. Block diagram for proposed scheme

2.2.1 Contourlet Transform

The Contourlet transform is a two-dimensional extension of the wavelet transform proposed by Do and Vetterli [11, 12] using multiscale and directional filter banks. The contourlet expansion is composed of basis images oriented at various directions in multiple scales with flexible aspect ratio that could effectively capture smooth contours of all images. The contourlet employs an efficient tree structured implementation, which is an iterated combination of Laplacian Pyramid (LP) [13] for capturing the point discontinuities, and the Directional Filter Bank (DFB) [14] to gather nearby basis functions and link point discontinuities into linear structures. Contourlet transform is more powerful than the wavelet transform in characterizing images rich of directional details and smooth contours [15, 16].

Let the image be a real-valued function I(t) defined on the integer valued Cartesian grid [2^l, 2^l]. The Discrete Contourlet Transform with scale j, direction k and level n of I(t) is defined as follows [17,19]:

$$\lambda_{j,k,n}(t) = \sum_{i=0}^{3} \sum_{m \in Z^2} d_k(m) \psi_{j,n}{}^{(i)}(t)$$

where $d_k(m)$ is the directional coefficient and

$$\psi^{(i)}{}_{j,n}(t) = \sum_{m \in Z^2} f_i(m) \phi_{j,n+m}(t)$$

where ϕ (t) is the scaling function and f_i is the spatial domain function.

Furthermore, the current existing steganalysis algorithms are limited to the domain of wavelet and DCT transforms. Therefore, identifying stego (constructed by embedding data into their contourlet coefficients) and cover image from the image data set is not easy by these steganalysis algorithms. This fact motivates us to develop efficient steganalysis algorithm in contourlet domain. In this paper, contourlet subband based features are used for steganalysis. *Sub-band Coefficient Modelling* The coefficients in the produced sub bands of contourlet transformed image are very appropriate to obtain the texture feature due to coarse to fine directional details of the image in these sub-bands. Besides, the distribution of the sub-bands coefficients is symmetric and unimodal with mean skewness approximately near to zero, though they have not exactly Gaussian distribution [18]. These special characteristics of subband coefficients make them suitable for modelling by Gaussian distribution with density function.

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \prec x \prec \infty$$

where μ and σ are the mean and standard deviation of all the coefficients of sub-bands.

2.2.2 Feature Extraction

There are various methods in the literature to extract the relevant features of digital images based on different transforms or filtering techniques. Even though the accuracy of classifiers is based on the number of suitable features, higher the number of features slower will be the classification. So identifying a minimum number of features which can produce efficient classification is a challenge. In this paper, only 5 features have been used which is very less compared to the number of features used in the existing steganalysis. Contourlet transform is more sparser than wavelet as the majority of the coefficients have amplitudes close to zero. Also the moments of contourlet coefficients are more sensitive to the process of information hiding. The first four normalized moments of the high frequency and low frequency subband coefficients are more sensitive to the process of steganography. Since these moments could be a good measure for skewness and kurtosis due to information hiding, the first four normalized moments are extracted as features. Moments are computed as below:

$$m_k = \frac{E(X-\mu)^k}{\sigma^{2k}} \qquad k=1,2,3 \text{ and } 4.$$

where X represents the coefficients of contourlet sub bands. Since these moments alone are not sufficient to detect the changes in the medium frequency sub-bands, another feature namely structural similarity measure (SSIM) is also included. For estimating SSIM, medium frequency band is split into two equal number of subband groups X and Y respectively. SSIM includes three parts: Luminance Comparison (LC), Contrast Comparison (CC) and Structural Comparison (SC) and they are defined as below [20, 21, 22]:

$$CC(x,y) = \frac{2\sigma_x\sigma_y}{\sigma^2{}_x + \sigma^2{}_y}$$

$$LC(x,y) = \frac{2\mu_x\mu_y}{\mu^2{}_x + \mu^2{}_y}$$

$$SC(x,y) = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

$$SSIM(X,Y) = [LC(X,Y)][CC(X,Y)][SC(X,Y)]$$

The similarity of the whole image (I) is

$$SSIM\ (I) = \frac{\sum_{j=1}^{n} SSIM_{\ j}}{n}$$

where n is the number of middle frequency sub bands in the image. Feature set consists of the first four normalized moments m_k (k=1,2,3,4) and the similarity measure SSIM(I).

2.2.3 Classification

A three back propagation Neural Network (NN) is used as a classifier for identifying stego images as well as images [10]. The power of back propagation is that it enables us to compute an effective error for each hidden unit, and thus derive a learning rule for the input to hidden weights. Non linear Support Vector Machine (NSVM) classifier is used for effective classification of stego images and cover images in this work.

3. EXPERIMENTAL RESULTS

3.1 Method 1

The database collection is done through 108 images in Kodak PhotoCD format and spans a large variety of image subjects. In fact, the collection even includes some digitally manipulated images. It is assume that these images have not been modified by steganography software and hence represent the set of cover images.

All images are converted from their original photoCD format to RGB TIFF images at the base resolution of 768X512 pixels. As proposed methods operate on mono-chrome images, only green channel is used. Furthermore, the image is crop to remove black boundary regions (30 and 20 pixels).

3.1.1 Detection of stochastic embedding

The process can be modeled by noise addition, without loss of generality. Although, the method allows for alternative noise statistics, this uses a Gaussian noise in this experiment. It uses two different embedding strengths at σ^2 =3 and σ^2 =9, corresponding to PSNR of 41dB and 38dB, and embedding rates of 0.84bpp and 0.91bpp, respectively.

For each image, mean square error, mean absolute error, and weighted mean square error between the image and compressed version are computed. Compression is performed with JPEG2000 at 95, 90, 85, 80, 70, 60, and 50% of the lossless rates.

During training, feature vectors are processed to obtain an optimal projection onto a two dimensional feature space. Then a Bayesian classifier is trained on the reduced features using three classes (namely no embedding, low embedding, and high embedding). In the test phase, the projection matrix obtained in the training phase is used to reduce the feature vector dimensions. Afterwards, classification is performed using the previously learned parameters.

In the whole scenario, 9 cover images out of 54 are mislabeled as stego-image, while 13 stego-images are mis-labeled as a cover image. Corresponding false alarm and miss rates are 16.7% and 12% respectively.

3.2 Method 2

The proposed steganalysis is implemented using MATLAB 7.6.0 with MATLAB scripts. The experiments are conducted on a personal computer with a 1 GB RAM and P-IV processor. For training we have used 12,200 images from Computer Vision image dataset and INRIA image dataset. It contains 5,500 cover images and 6,700 stego images which are generated by different embedding algorithms like LSB, F5, Contsteg, and YASS. Washington image dataset [22] is used for testing the proposed steganalytic method. 100 images are used to test the proposed scheme, with 60 cover images and 40 stego images.

In order to analyze the proposed method, four typical steganography methods are used. Table 1 gives a comparison of the average detection accuracy between NN classification and non linear support vector classification with same feature set. From this table, one can see that Non linear Support Vector Machine classifies stego images and cover images more accurately. Figure 3. Depicts the performance comparison of NN classifier and NSVM classifier in classifying stego images.

Table 1. Average correct detection rates for natural images and stego images

Steganography Methods	Classifier	Average correct detection rates						
		Embedding rates			Different image size			
		100%	50%	25%	512X512	256X256	128X128	64X64
LSB	NSVM	.945	.922	.904	.952	.976	.937	.902
	NN	.922	.912	.894	.975	.973	.895	.870
F5	NSVM	.950	.971	.977	.957	.951	.931	.894
	NN	.910	.969	.898	.985	.942	.973	.763
ConSteg	NSVM	.928	.917	.908	.905	.957	.903	.823
	NN	.921	.897	.878	.870	.856	.831	.723
YASS	NSVM	.966	.936	.914	.941	.987	.912	.853
	NN	.956	.906	.844	.901	.892	.879	.733

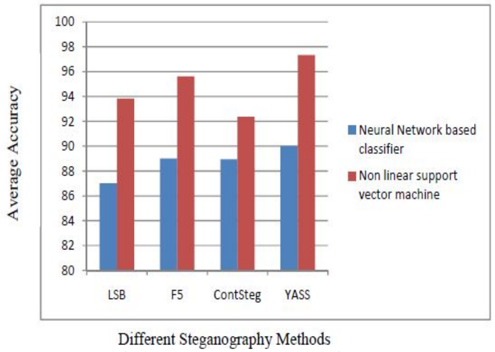

Figure 3. Performance comparison of Neural Network and Non linear Support vector machine based classifiers

The relevancy of the extracted features used in this steganalysis is evaluated using error estimation. Table 2 and Figure 4 display the sample Median Absolute Error (MAE) which exhibits a higher error than bias for all the embedding algorithms. So it is clear that, with this minimum dimensional feature set, proposed method can able to detect the stego image.

Table 2. Median absolute error and bias for the proposed method

Algorithm	MAE	Bias
LSB	5.91X10^{-3}	-1.70X10^{-4}
F5	6.63 X10^{-3}	-3.78X10^{-4}
YASS	4.19X10^{-3}	1.87X10^{-4}
ContSteg	3.25X10^{-3}	0.58X10^{-4}

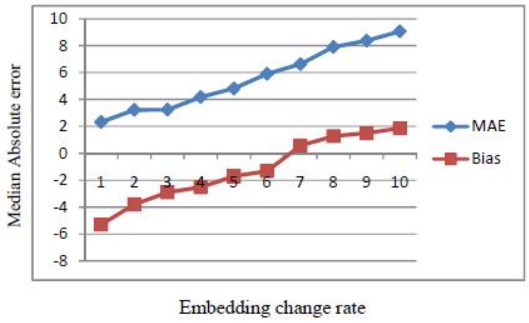

Figure 4. Median Absolute Error (MAE) and Bias of proposed steganalyzer, with respect to embedding rates

The proposed work is compared with the Contourlet-Based Steganalysis (CBS) [23] methods and the results show significant improvement and they are tabulated in Table 3. The Data set used in the proposed scheme for comparison is the Washington dataset which is used in ContSteg [24] and CBS [23].

Table 3. Accuracy of CBS and proposed steganalysis methods on detection of stego-image produced by ContSteg

Secret Data Size (bits)	Steganalysis Method	Average Detection Accuracy (%)
5000	CBS	59
	Proposed Method	77
10,000	CBS	63
	Proposed Method	89
15,000	CBS	68
	Proposed Method	93

The correct detection rate is improved in the proposed method compared to existing steganalysis schemes [23]. Especially proposed scheme is independent of file formats and image types. The new method based on statistical steganalysis utilizes fewer features than rest of the methods. Hence, it is fast and the computational cost of the new method in extracting the features and detecting the stego image are much less than that of the methods based on feature extraction.

4. CONCLUSIONS

In this paper, it has been proposed new steganalysis techniques based on rate-distortion arguments. These techniques are base on the observation that the steganographic algorithms invariably disturb the underlying statistics therefore change in rate-distortion characteristics of the signals. This is demonstrated the effectiveness of the proposed approach against the stochastic embedding algorithms with varying degrees of success.

On the other hand another approach has been proposed a steganalysis blind detection method based on contourlet transform and non linear support vector machine. This method extracts the statistical moments and structural similarity of the contourlet coefficients as the feature set. The performance of the proposed scheme is illustrated using various testing metrics. The average correct detection rate is improved, at the same time the dimension of the feature set and the average run time is reduced in this proposed scheme. Furthermore, the method proposed here is an universal blind scheme, which is independent of image type and file format.

5. ACKNOWLEDGMENTS

I would like to thanks the experts who are involved in this area for so long time, as well as their valuable resources and also like to indebtness towards my colleagues of their valuable help to write this paper.

6. REFERENCES

[1] R.J. A nderson and F.A. Petitcolas, “On the limits of steganography,” IEEE Journal of selected Areas in Communications 16, pp. 474-481, May 1998. Special issue on copyright & privacy protection.

[2] I.J. Cox, M.L. Miller, and J.A. Bloom, Digital Watermarking, Morgan Kaufmann Publishers, San Francisco, CA, USA, 2002.
S.Katzenbeisser and F.A.P. Petitcolas, eds., Information Hiding: techniques for steganography and digital watermarking, Artech House, Boston, MA, 2000.

[3] J. Fridrich and M. Goljan, “Digital image steganography using stochastic modulation,” in Proc. SPIE: Security and Watermarking of Multimedia Contents V, E.J. Delp and P.W. Wong , eds., E123, pp. 191-202, Jan. 2003.

[4] Fridrich J, Goljan M.(2002) “Practical: Steganalysis of digital images- state of the art. In:” Proceedings of SPIE, Security and Watermarking Multimedia Content IV.Vol. 4675. New York: SPIE, pp 1-13.

[5] McBride B T, Peterson G L, Gustafson S C.(2005) “A new blind method for detecting novel steganography”. Digit Invest, 2: 50-70.

[6] Johnson.N.F, Jajodia.S.:(1998) “Steganalysis: the investigation of hidden information”, In: Proc. IEEE Information Technology Conference, Syracuse, NY.

[7] Fridrich.J, Goljan.M.: Practical steganalysis of digital images state of the art, in: Proc. SPICE Photonics West, Electronic Imaging (2002), Security and watermarking of multimedia contents, San Jose, CA, vol. 4675, January 2002, pp 1-13.

[8] Avcibas I, Memon N D, Sankur B.:(2001) “Steganalysis of watermarking techniques using image quality metrics” In: Proceedings of SPIE, Security and Watermarking of Multimedia Content III, vol. 4314. New York: SPIE, 2001. 523-531.

[9] Avcibas I, Memon N, Sankur B. (2003):” Steganalysis using image quality metrics”. IEEE Trans Image Process, 12: 221- 229.

[10] V. Natarajan and R. Anitha,(2012) “Universal Steganalysis Using Contourlet Transform”, Advances in Intelligent and Soft Computing, Springer – Verlag, Volume 167/2012, 727-735.

[11] Do, M.N., Vetterli, M.: Contourlets (2002) “A directional multiresolution image representation”, Proc. of IEEE Int. Conf. on Image Process., Piscataway, NJ, pp. 357-360.

[12] Minh N.Do, Martin Vertterli. (2006) “The Contourlet Transform: An Efficient Directional Multiresolution Image Representation”, IEEE Transaction on Image Processing, vol.14 no.12,pp.2091-2106.

[13] Burt P.J and Adelson E.H.(1983) “The Laplacian pyramid as a compact image code”, IEEE Trans.Commun,vol 31,no.4, ppl 532-540.

[14] Bamberger R.H and Smith M.J.T.(1992) “A filter bank for the directional decomposition of images: theory and design”, IEEE Trans.Signal Process. Vol.40,n0.4,pp.882-893.

[15] `Yazdi, M., Mahyari, A.G. (2010) “ A new 2D fractal dimension estimation based on contourlet transform for texture segmentation”, The Arabian Journal for Science and Engineering, vol. 35, No. 13, pp.293-317.

[16] Ali Mosleh, Farzad Zargari, Reza Azizi. (2009) “Texture Image Retrieval Using Contourlet Transfrom”, International Symposium on Signal, Circuits and Systems.

[17] M.N.Do, and Vetterli.M. (2006) “Directional multiscale modeling of images using contourlet transform”, IEEE Transactions on Image Processing, Vol.15, no.6,pp.1610-1620.

[18] Chun Ling Yang, Fan Wang, Dongqin Xiao.(2009) “Contourlet Transform based Structural Similarity for image quality assessment”, Intelligent computing and intelligent systems.

[19] http://www.cs.washington.edu/research/imagedatabase.

[20] “Kodak PhotoCD images.” ftp://ftp.kodak.com/www/images/pcd.

[21] “HP Labs LOCO-I/JPEG-LS Home Page.” http://www.hpl.hp.com/loco/.

[22] Mathworks (MATLAB) http://www.mathworks.com

[23] Hedieh Sajedi, Mansour Jamzad.(2008)" A Steganalysis method based on contourlet transform coefficients", International Conference of Intelligent Information Hiding and Multimedia Signal Processing.

[24] Sajedi H., and Jamzad M. "ContSteg: Contourlet-Based Steganography Method", Wireless Sensor Network, Scientific Research Publishing (SRP) in California (US),1(3),163-170.