ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A Technique for Steganalysis of Object Oriented LSB Steganography

Archana O. Vyas, Dr. Sanjay V. Dudul

Ph.D. Student, Department of Applied Electronics, S.G.B. Amravati University, Amravati, India

Professor and Head, Department of Applied Electronics, S.G.B. Amravati University, Amravati, India

ABSTRACT: Steganography and steganalysis are the two faces of a coin. Steganography deals with the information hiding, whereas it also possible to detect the presence of hidden information in cover objects. Steganalysis is composed of “steg” and “analysis”, which aims to analyze the stego carriers to detect or extract the hidden information. In object oriented steganography, the secret information is hidden in some selective object of a cover object and the steganalysis is further a challenging task here. This paper describes the two existing least significant bit steganalysis methods, “Chisquare attack” and “Histogram analysis” applied on object oriented LSB stego images to classify them as stego and nonstego images. The novelty of this approach is that the stego objects, on which steganalysis techniques are applied belongs to two distinct cover images. Difference in histogram is plotted for the stego and nonstego image, the parameters like True positive rate and false positive rate are calculated thereafter.

KEYWORDS: Steganalysis, Histogram, TPR, FPR, Detection rate.

I. INTRODUCTION

Steganalysis is a science of investigating steganography. The objective of steganalysis is to collect sufficient evidence from observed data for discriminating a carrier as ‘stego’ or ‘cover’, based on presence or absence of embedded message. Images are often used as a carrier because of their extensive availability with high resolution of pixels. Data embedding in a multimedia carrier like image may involve varying parameters such as different image formats, different embedding algorithms and various steganographic keys. This has made steganalysis a more difficult and challenging task. Steganography embeds secret messages in images; this causes alterations in the statistics of an image. Statistical steganalysis, as the name implies, analyses this underlying statistics of an image to detect the secret embedded information. Specific statistical steganalysis can be classified based on data hiding techniques i.e. in spatial domain and transform domain [7].

A) Chisquare Attack [6]:

The LSBs in images are not completely random. Rather, they claim, the frequencies of each of the two pixel values in each POV tend to lie far from the mean of the POV. In other words, it is uncommon for the frequency of pixel value 2k to be (nearly) equal to the frequency of pixel value 2k + 1 in a typical image with no embedded information. Furthermore, as information is embedded into an image using an algorithm like Flip Embed which creates POVs, the frequencies of 2k and 2k+1 become equal or nearly so. The Chi-squared attack was designed to detect these near-equal POVs in images and bases the probability of embedding on how close to equal the even pixel values and their corresponding odd pixel values are in the test image. The Chi-squared attack is designed to be adaptable to various embedding algorithms, but the basic concept is the same regardless of the embedding algorithm [6].

a) pov3 Algorithm[6]:

The detection algorithm, pov3, is designed to perform the Chi-squared attack on an image and output the probability of embedding. The pixels are tested sequentially, beginning in the upper left corner, i.e., index (1, 1), and following the same path through the image [6].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

The implementation of pov3 is as follows:
Let X^{128*1} and Y^{128*1} be two vectors such that x_k = frequency (2k) and y_k = frequency (2k + 1), $0 \leq k \leq 127$. Initially, every entry in X and Y is set to 0. Then pov3 counts the gray values in the test image and increments the corresponding entry in X or Y. The theoretically expected frequency of gray values 2k and 2k + 1 is $z_K = \frac{xk+yk}{2}$. Now suppose that there are 'n' categories, that is, 'n' POVs. In the case of 8-bit grayscale images, there are 128 categories (256 gray values /2). Without loss of generality, we concentrate on the even values of the POVs, so that the measured frequency of occurrence in category k is x_k. (Note that the sums of the frequencies of pixel values 2kand 2k + 1 in a cover image and in a stego image are the same, i.e., $x_k + y_k$ in a cover image equals $x_k + y_k$ in its corresponding stego image). A minimum frequency condition, so for if $0 \leq k \leq 127$, if $x_k + y_k \leq 4$, $x_k = y_k = z_K = 0$, and n = n –1 i.e. whenever the combined frequency of 2k and 2k + 1 is less than or equal to 4, the individual frequency counts of 2k and 2k + 1 are set to 0 and the number of categories n is decremented by 1. The Chi-squared statistics, with n − 1 degrees of freedom, is then calculated [6]:

$$x_{n-1}^2 = \sum_{i=0}^{127} \frac{(Xi-Zi)^2}{Zi} \quad \text{, Where} \quad Zi = \frac{Xi+Yi}{2} \qquad (1)$$

The expectation is that for a stego image, X^2_{n-1} is relatively small because x_i should be near z_i, by the hypothesis, and for a non-stego image, X^2_{n-1} is relatively large because x_i should be far from z_i. The final step of the process is calculating p, the probability of embedding, by integrating the density function with as X^2_{n-1} as its upper limit:

$$p = 1 - \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \int_0^{x_{n-1}^2} e^{-\frac{u}{2}} u^{\left(\frac{n-1}{2}-1\right)}\, du \qquad (2)$$

This probability of embedding is the probability of x_{n-1}^2under the condition that $Xi = Zi$ for all i in Equation 1. The density function, 1 − p, converges to 1 as x_{n-1}^2 approaches infinity, so p approaches 0 as x_{n-1}^2 approaches infinity. Therefore, for large $x_{n-1}^2$the probability of embedding is near 0. However, when $x_{n-1}^2$is small relative to n − 1, 1 − p is near zero and hence p is near 1.Thus, for relatively small x_{n-1}^2 the probability of embedding is near 1[6].

B) Histogram analysis [5]:

The histogram of a stego image has less high-frequency power than the corresponding histogram of the cover image. Thus, the center of gravity of |F(h)|, which denotes the Fourier transform of the histogram h, will decrease after LSB embedding. This property can be used as a feature for distinguishing between cover and stego images. The local maxima of the histogram of images will decrease and the local minima will increase after LSB embedding. This property can be used to define a feature that can be used to detect LSB steganography. This feature is the sum of the absolute differences between each local extremum and its neighbors in the intensity histogram of stego images [5].

a) Effects of LSB steganography on histogram[5]:

Let Pc (i, j) denote the pixel value at location (i, j) in the cover image. The intensity histogram is then defined as:

$$h_c(n) = |\{(i, j)|p_c(i, j) = n\}| \qquad (3)$$

Where 'n' denotes a grayscale level in the range 0 to 255. In other words, hc(n) indicates the number of pixels in the cover image with grayscale value n. Considering the effect of LSB, with an embedding rate ρ, on the cover image histogram. There is a 50% chance that the pixel values at selected locations will already have the desired LSB value. Hence, a proportion (1- ρ/2) of the pixels will not be modified. The remaining pixels are incremented or decremented with equal probability. Assuming that the embedding locations are uniformly distributed and independent of the pixel values [5].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

$$h_s(n) = \left(1-\frac{\rho}{2}\right)h_c(n) + \frac{\rho}{4}\,(h_c(n-1) + (h_c(n+1)) \quad (4)$$

A local extremum, n^*, in a histogram, h(), is defined by:

$$(h(n^*) - h(n^* - 1))\,(h(n^*) - h(n^* + 1)) > 0 \quad (5)$$

According to Equation 5, for any local maximum, n*, it turns out that

$$\begin{aligned} h_s(n^*) &= \left(1-\frac{\rho}{2}\right)h_c(n^*) + \frac{\rho}{4}(n^*-1) + h_c(n^*+1) \\ &= h_c(n^*) - \frac{\rho}{4}[\{h_c(n^*) - h_c\ (n^*-1)\} + \{h_c(n^*) - h_c\ (n^*+1)\}] \\ &< h_c(n^*) \end{aligned} \quad (6)$$

Similarly, for any local minimum point, n^*, we have $h_s(n^*)>h_c(n^*)$. Thus, after LSB steganography, the local maxima of an image histogram decrease and the local minima increase. So we can consider the sum of absolute differences between each local extremum and its neighbors in the histogram. These sums are denoted by Dc and Ds for the cover and stego images, respectively. That is,

$$D_c = \sum_{n^*}|2.h_c(n^*) - h_c(n^*-1) - h_c((n^*+1))| \quad (7)$$

$$D_s = \sum_{n^*}|2.h_s(n^*) - h_s(n^*-1) - h_s((n^*+1))| \quad (8)$$

It is expected that Dc > Ds for any image after LSB steganography [5].

C) Steganalysis Performance Parameters:

a) *True Positive Rate (TPR):* It is defined as the ratio of the number of correctly classified images out of the overall test images. TPR should be as high as possible [8].

b) *False Positive Rate (FPR):* It represents the ratio of the wrongly classified the plain images as stego ones. FPR should be as low as possible [8].

c) *Classification Rate:* Classification rate is defined as the average of positive detection (PD) and negative detection (ND), given by (PD + ND) / 2, Where, Positive Detection (PD) is classifying the stego images correctly and Negative Detection (ND) is classifying the non stego images correctly [4].

II. RELATED WORK

In [6] Christy A. Stanley, discussed the theory behind the Chi-squared attack, and implementation of the Chisquared attack when secret data is embedded in a single cover image. He concluded that Chi-squared attack is quite effective against embedding algorithms which flip the LSBs when large amounts of data are embedded. In [5] Jun Zhang and Ingemar J. Cox and Gwena, el Do, err, proposed a targeted steganalysis algorithm that exploits the fact that after LSB embedding, the local maxima of an images gray level or color histogram decreases and the local minima increases. Consequently, the sum of the absolute differences between local extrema and their neighbors in the intensity histogram of stego images will be smaller than for cover images. They demonstrated that their method has superior results as compared with other recently proposed algorithms when the images contain high-frequency noise. In [2] Sorina Dumitrescu, Xiaolin Wuand Zhe Wang, introduced a new, principled approach of detecting LSB steganography in digital signals such as images and audio. It is shown that the length of hidden message embedded in the least significant bits of signal samples can be estimated with relatively high precision. Their new steganalytic approach was based on some statistical measures of sample pairs that are highly sensitive to LSB embedding operations. The resulting

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

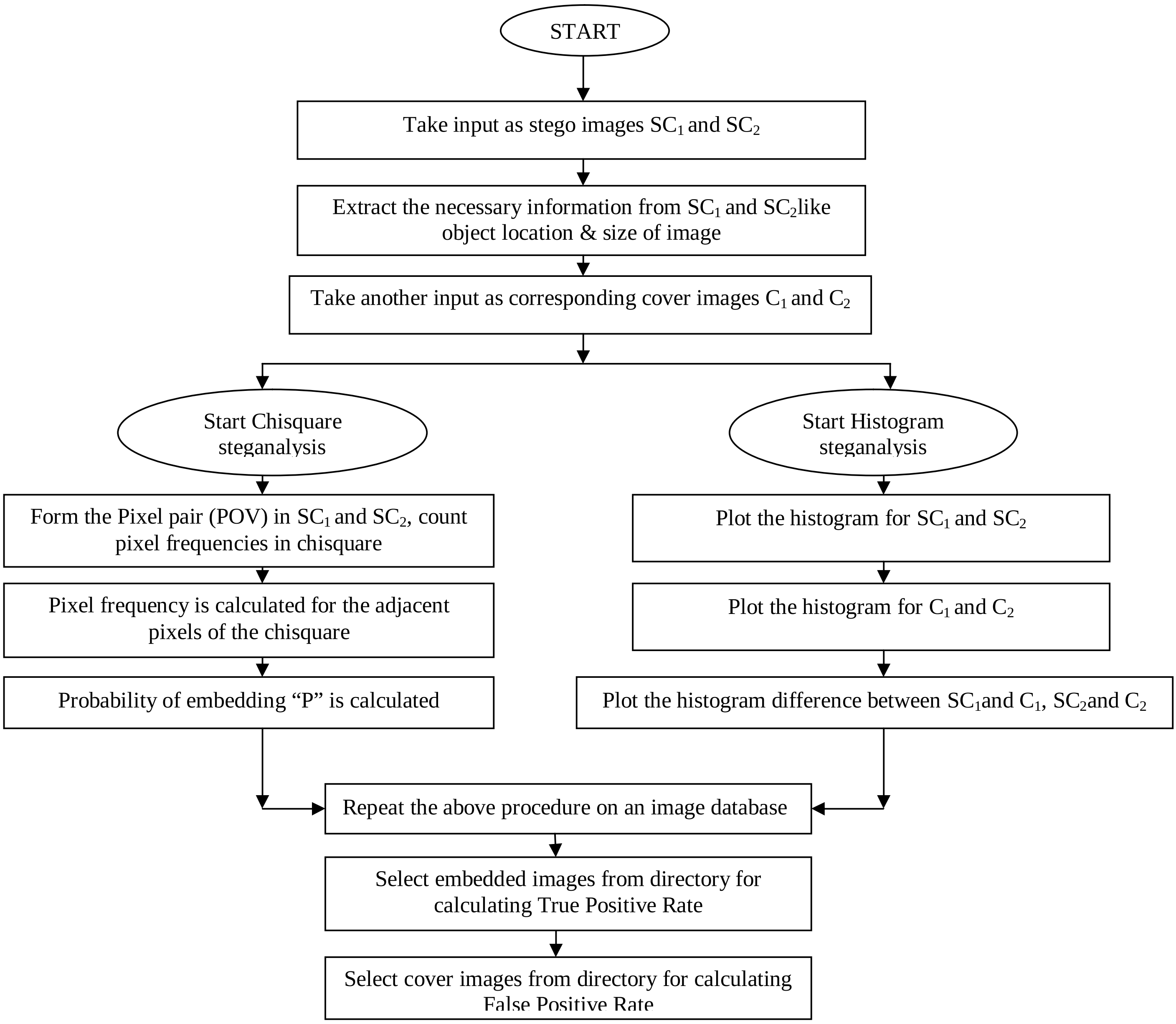
Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

detection algorithm was found to be simple and fast. In [3] Andrew D. Ker has described a general framework for steganalysis of LSB Replacement, which can consider arbitrary tuples of pixels. It involves a new paradigm for detection, in which the effects of embedding a message of known length can be inverted, and a cover image model against which a best fit is found. The frame work can include many of the previously known detectors.

III. PROPOSED METHODOLOGY

The existing two LSB steganalysis techniques, namely, "Chisquare attack" and "Histogram analysis" are applied in a novel paradigm, on the two distinct stego images which are obtained when the LSB embedding is done in some selective objects of the original cover image, not in the whole image. The methodology is explained via the flow chart shown below.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

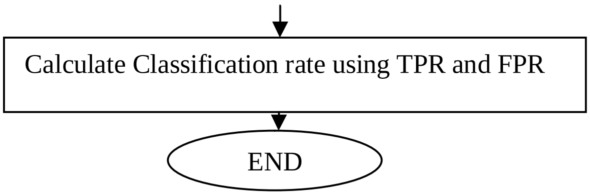


Fig.1. Flowchart of proposed methodology

III. SIMULATION RESULTS

The simulation is accomplished by taking three image data base of different number of cover images and corresponding stego images, algorithm is implemented with MATLAB. For all the three database the simulation procedure followed, involve following steps,

a) Selections of two cover images, plotting their Histogram.
b) Selection of corresponding stego images, plotting their Histogram.
c) Plotting the histogram difference for the stego image and corresponding cover image.
d) Calculation of true positive rate (TPR), false positive rate (FPR), Classification rate.

As illustrated below fig.2, fig. 4, fig.6 gives the histogram difference for the two cover images with their corresponding stego images and fig. 3, fig.5, fig.7 gives the calculated values of TPR, FPR and classification rate. The TPR, FPR and classification rate values obtained after simulation are tabulated in table below for three different data base of images.

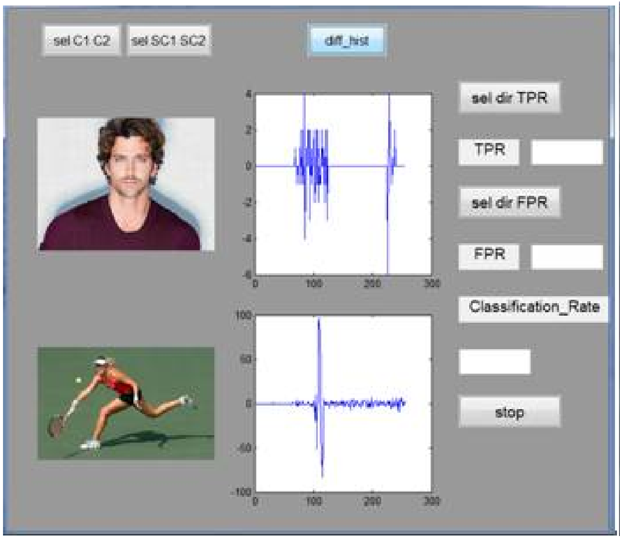


Fig.2. Histogram difference for cover images with stego images

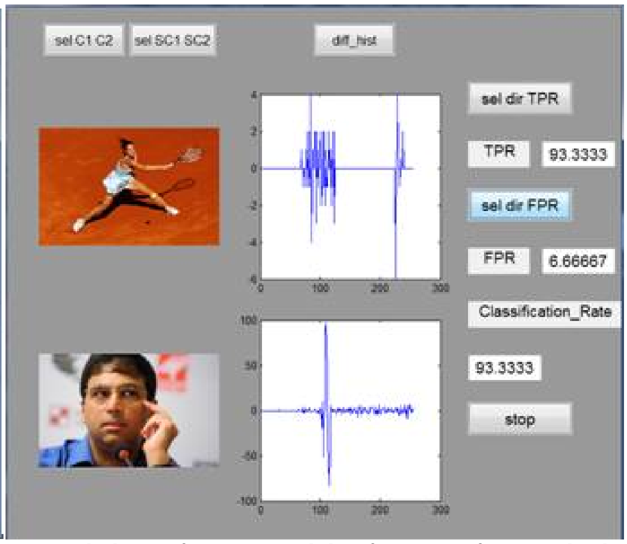


Fig 3. Calculation of TPR, FPR and classification rate for image data base 1

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

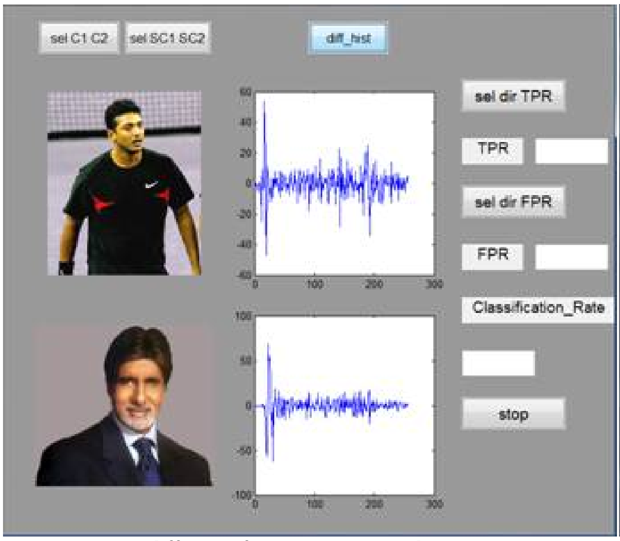


Fig.4. Histogram difference for cover images with stego images

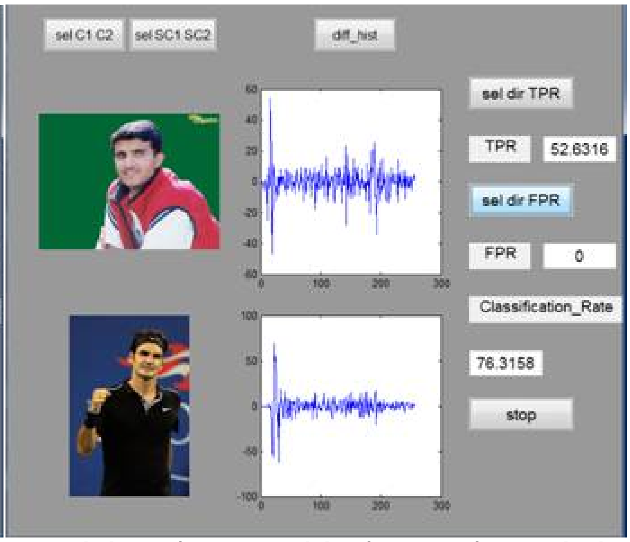


Fig 5. Calculation of TPR, FPR and classification rate for image data base 2

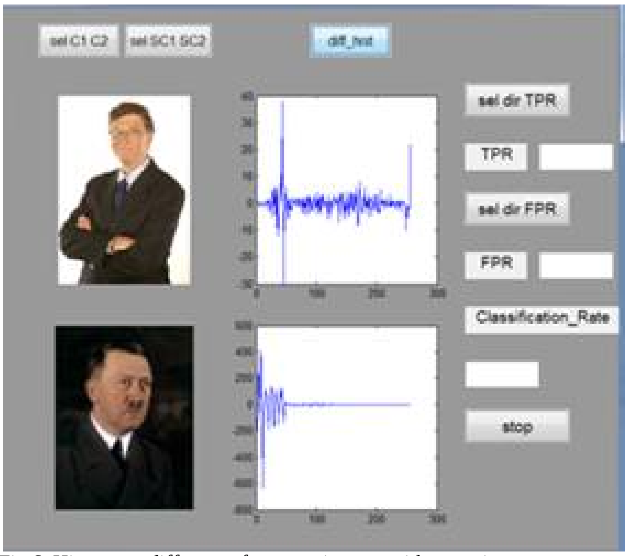


Fig.6. Histogram difference for cover images with stego images

Fig 7. Calculation of TPR, FPR and classification rate for image data base 3

Table 1. Evaluated Parameters

Data Base No.	TPR value	FPR value	Detection rate
I	93.33	52.63	52.63
II	6.66	0	5.263
III	93.33	76.31	73.68

IV. CONCLUSION

When the images are hidden in some selective objects of the cover image, it is a quite difficult task to steganalyze such object oriented stego covers; here the proposed method performs such kind of steganalysis very effectively. The difference in histogram analysis clearly detects the presence of steganography. Obtained value of true positive rate is

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

satisfactorily high, and low value of false positive rate indicates that almost no false detection has been done by the proposed methodology. The classification rate achieved is also satisfactory when tested for a large database of images.

REFERENCES

1. SorinaDumitrescu, Xiaolin Wu, NasirMemon, "On Steganalysis of Random LSB Embedding in Continuous-tone Images", IEEE, proceeding for international conference on image processing, vol. 3, 2002.
2. SorinaDumitrescu, Xiaolin Wuand Zhe Wang, "Detection of LSB Steganography via Sample Pair Analysis", IEEE, Transaction on signal processing, 2003, vol. 51, issue 7.
3. Andrew D. Ker, "A General Framework for Structural Steganalysis of LSB Replacement", IEEE, Transaction on information forensics and security, 2005, vol. 2, issue 1.
4. Wen-Nung Lie and Guo-Shiang Lin, "A Feature-Based Classification Technique for Blind Image Steganalysis", IEEE Transactions On Multimedia, vol. 7, no. 6, December 2005.
5. Jun Zhang, Ingemar J. Cox and Gwena el Do err, "Steganalysis for LSB Matching in Images withHigh-frequency Noise", IEEE, Workshop on multimedia signal processing, 2007.
6. Christy A. Stanley, "Pairs of Values and the Chi-squared Attack", IEEE conference publication, 2008.
7. Rita Chhikara, Latika Singh, "A Review on Digital Image Steganalysis Techniques Categorised by Features Extracted", International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 4, October 2013.
8. Archana O. Vyas Dr. Sanjay V. Dudul, "Study of Image Steganalysis Techniques", International Journal of Advanced Research in Computer Science, Volume 6, No. 8, Nov-Dec 2015.

BIOGRAPHY

Ms.Archana O. Vyas was born in Indore, Madhyapradesh in 1980. She received the B.E. Degree in Electronics and Tele-communication from S.G.B. Amravati University, Amravati in 2009 and completed her M.Tech. in Electronic Systems and Communication from Government college of Engineering Amravati in 2011. Currently she is working as Assistant Professor in Electronics and Tele communication Engineering, Department at G. H. Raisoni College of Engineering & Management, Amravati. She is pursuing PhD. degree in Electronics Engineering from Sant Gadge Baba Amravati University, Amravati, India. Her interest of research is image steganography and steganalysis using computational intelligence approach.

Dr. Sanjay V. Dudul was born on 28th August 1964 at Amravati, Maharashtra. He was conferred on Ph.D. in Electronics Engineering on topic of Computational Intelligence by Sant Gadge Baba Amravati University, Amravati in 2003. He obtained B.E. degree in Electronics Engineering in 1986 from Nagpur University. He received M.E. degree in Electronics Engineering with specialization in Computer Applications in 1989. Currently, he is a Professor and Head of the Department of Applied Electronics at Sant Gadge Baba Amravati University, Amravati, India. He has published 81 research articles in peer reviewed and refereed International journals and 12 research articles in refereed National Journals. His numerous publications have been listed in SCOPUS (author's h index is 9.0). Hitherto, 13 students have been awarded Ph.D. degree in Electronics Engineering under his guidance. He has filed 6 Indian patents. He is a member of editorial boards of many International Journals. His research area includes Image Processing, AI in Pattern Recognition, Prediction and Regression, Machine Intelligence: A Soft-Computing Approach, Advance digital signal processing, Pattern Recognition, Analysis and Machine Intelligence, Computational Intelligence, Bio-medical Signal Processing, Intelligent/Smart Sensors and Learning Systems.