NEURAL NETWORK BASED STEGANALYSIS IN STILL IMAGES

Liu Shaohui[1] *Yao Hongxun*[1] *Gao Wen*[1, 2]
[1]Department of Computer Science, Harbin Institute of Technology
150001, Harbin, the People's Republic of China
[2]Institute of Computing Technology, Chinese Academy of Science
Email: {shaohl,yhx}@vilab.hit.edu.cn , wgao@ict.ac.cn

ABSTRACT

Seganalysis has recently attracted researchers' interests with the development of information hiding techniques. In this paper we propose a new method based neural network to get statistics features of images to identify the underlying hidden data. We first extract features of image embedded information, then input them into neural network to get output. And experiment results indicate this method is valid in steganalysis. This method will be used for internet/network security, watermarking and so on.

1. INTRODUCTION

Information hiding (maybe called data 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. Digital watermark is one of the most important data hiding applications. The major driving force for this is the need for effective copyright protection for digital media. This hidden information (mark, logo and so on) can be used to verify by owner with software if breaking of copyright law, and maintaining owner's legal rights. This is the positive application of information hiding technology. But there must to exist the negative aspect for everything. Some who have ulterior motives utilize this technology to transmit illegal information for avoiding law enforcement. It is said after 911 terrorist attacks, people think terrorist may be communicated with each other by data hiding technology. US today printed an article "Terror groups hide behind web encryption" by Tuck Kelley[1]. In his article, he writes "US officials and experts say steganography is the latest method of communication being used by Osaw bin Laden and his associates to out for law enforcement". And this may be true or will be true. In this case, the data hiding in digital media on Internet has proven to be a boon for terrorists. So how to determine whether digital media has hidden information become a emergency problem for social security and stability. This new research area is called as steganalysis similar with cryptanalysis. Contrast to the goal of Information hiding, Steganalysis is the art of discovering and rendering useless such covert messages, hence making information hiding failed. Though steganography is often confused with the relatively well-know cryptography, the two are but loosely related. Cryptography is about hiding the contents of a message, steganography, on the other hand, is about concealing the very fact that a message is hidden. Detection of steganography, estimation of message length, and its extraction belong to the field of steganalysis. Some definitions and several methods of steganalysis were proposed in the literature [2~,9,16]. In [2], authors give an overview of some characteristics to detect the existence of hidden information. And author in [3] gives a good description of popular free software's steganalysis. H. Farid in [4] tells a steganalysis method based fisher linear classifier. I.Aveibas et al in [5] take the regress analysis to analysis image based image metrics. And for LSB

embedding methods, the most successful resarchers may be J. Fridrich[6,7,8,9]. Some good review papers [10~12] are recommend. And with the help of steganalysis, ones can find more robustness methods to resist attack and analysis [12].

2. INFORMATION DETECTION BASED ON NEURAL NETWORK

The information detection is based on result that host image must be difference with hidden information host image. Maybe because human eye's mask features, we can't find any difference between host images and hidden information host images. But in essentially, data hiding process have to alter host image for embedding data. In other words, data hiding algorithms reveals statistical evidence or traces which can be used to detect the existence of hidden information in still images.

Because now popular data hiding methods can be divided into two major classes: spatial domain and transform domain. Spatial domain is simple and easy to implementation, but their robustness is weaker than other methods' based other domain. Transform domain includes discrete Fourier transform, discrete cosine transform, discrete wavelet transform mainly. And in spatial domain, mainly data hiding method is least significant bit (for short writing LSB) including all kinds of improved LSB methods. In transform domain, methods can be divided into several classes, quantization based, LSB based. Because we hope our method is feasible to all kinds of hiding methods no regardless embedding information in spatial or transform domain. So we try our best to consider all kinds methods' features. But due to methods is too much, so we only consider transforms used in common data hiding methods.

In this paper, we only consider following transforms, DFT, DCT and DWT. Firstly we analysis object digital image according these three different kinds transforms in this method. The object image is transformed into transform domain data according these three transforms. Then calculate these transforms data's statistical features which can be exploited to detect hided information. The reason for selecting DFT, DCT and DWT is that most data hiding method operate in these domains. So we select these domains to design algorithm to be able to detect methods as many as possible. These selected features should be significantly impacted by the data hiding processing. But it is difficult to find those features, so we select neural network to process this problem, neural network has the super capability to approximation any nonlinear functions. For these features which have more effected by data hiding process, neural network will assign larger weight coefficients and for these features which have less effected by data hiding process, neural network will assign less weight coefficients. Next section we explain our method and features selecting processing.

3. IMAGE FEATURES EXTRACTING

First we analysis image's discrete cosine transform domain's statistics features. We divide each image into 8×8 sub-block and then take DCT of each sub-block. Based on analysis of I. Aveibas et.al work [5], data hiding process possesses statistical difference in image quality metric scores obtained from blurred-and-hidden images as compared to blurred-but-non-hidden host images. We select spectral measures based on DFT and DCT. In DFT data hiding process, one quantize the magnitude to hide information and can't change the phase information (this is very important), so selecting metrics based on magnitude (two statistics, image and its sub block) In DWT, we select these same metrics with in DFT. Finally in DFT and DCT, we select 4 statistics. Next we take three levels DWT of each training images, and we calculate the mean value, variance, skewness and kurtosis of each part of every level. Then according pyramid algorithm's characteristic to forecast the

original data then calculate the error's statistics features [4]. Because in calculating process, variance is very larger than other statistics, we give up variance statistics. So every image only has 36 statistics. Added with DCT's 4 statistics based image metrics, each image has 40 statistics. So we set the number of neural network input as 40, the output is one. We use –1 and 1 to denote without hidden information and with hidden information respectively.

4. PERFORMANCE ANALYSIS AND EXPERIMENT RESULTS

From the measured statistics of training sets of images with and without hidden information, our destination is to determine whether an image has been hidden information or not. Because data hiding process is a nonlinear process, if we only use linear classifier to classify images, it is not a good simulation. And neural network has an excellent capability to simulate any nonlinear relation, so we make use of neural network to classify images.

In I.Aveibas paper [5], he used following regression model, each decision label *y* in a sample of *n* observations as a linear function of the image quality measure scores *x's* plus a random error:

$$\begin{aligned} y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_q x_{1q} + \varepsilon_1 \\ y_2 &= \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_q x_{2q} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_q x_{nq} + \varepsilon_n \end{aligned}$$

And the complete model is

$$y = \mathbf{X}_{nxq}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad such\ that \quad \begin{cases} rank(\mathbf{X}) = q \\ E[\boldsymbol{\varepsilon}] = \mathbf{0} \\ Cov[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I} \end{cases}$$

The corresponding optimal MMSE linear predictor β can be obtained by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\mathbf{y}\right).$$

but we found that neural network is more feasible to process nonlinear problem, so In this paper we take use of BP neural network to train and simulate images. This BP neural network uses three levels: input level, hidden level and output level. In neural network, the important issue is the slow of convergence. In practice, this is the main limitation of neural network applications. And many new algorithms claimed fast convergence were developed. In this paper a single parameter dynamic search algorithm is used to accelerate network train. Each time only one parameter to be searched to achieve best performance, so this learning algorithm has a better improvement than other old algorithms ([12 13]). We set the number of this network's input as features, and node number of hidden level is set to be 40, and output is either yes or no.

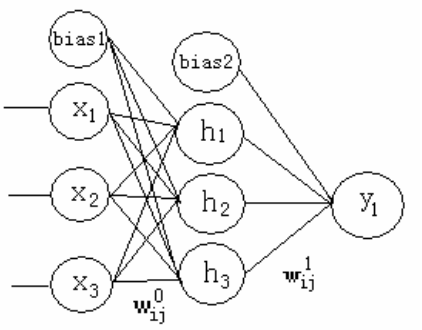

Fig.1. Three level BP neural network

We select two training set of images. In this example, two trainings include 23(no hidden) and 21 (hidden) images respectively. The data hiding method is Brain Chen's quantization index modulation methods. The training process time is 21.763 second. Total iteration number is 400 steps. Ant the test image base includes 40 (no hidden) and 41(hidden) images. The result is follow as table 1. (Note: This hiding method based on B. Chen's work [15])

Table1. Test results

type	Total number	Right detection	rate
Hidden images	41	35	85.4%
No hidden images	40	30	75.0%

5. CONCLUSION

This paper provides a new method to detect the existence of hiding information. This method find statistically evidences after host images has been hidden information, then using the capability of approximation of neural network for determining whether an image has been hidden information or not. We analysis the difference and some traces in detail. And results indicate that this method is promising. There is a lot of work that still needs to be done. Many other watermarking schemes and algorithm will to be included this research and extensive tests need to be done with a larger number of images

Detection technique development in this area of data hiding will continue. We find image's statistic features are important clues to determine whether hiding information or not from the detection process. So this suggests us to develop more robustness method with statistic features altered as little as possible. Steganalysis not only prevent will but also advance information hiding to provide another method to hide important information in digital media for transmitting on Internet rather than cryptography.

6.Reference

[1] Jack Kelley. Terror groups hide behind Web encryption. USA Today, Feburary 2001. http://www.usatoday.com/life/cyber/tech/2001-02-05-binladen.htm.

[2] Neil F.Johnson and Sushhil Jajodia. Steganalysis: The Investigation of Hidden Information. Proceedings of the IEEE Information Technology Conference, Syracuse, New York,USA.1998

[3] N. F. Johnson and S. Jajodia, "Steganalysis of Images Created Using Current Steganography Software," Lecture Notes in Computer Science, vol.1525, Springer-Verlag, Berlin, 273-289, 1998

[4] H. Farid, "Detecting Steganographic Message in Digital Images", Technical Report, TR2001-412, Dartmouth College, Computer Science, 2001.

[5] I.Aveibas, N.Memon, B.Sankur. Steganalysis based on image quality metrics. Multimedia Signal Processing, 2001 IEEE Fourth Workshop on , 2001 Page(s): 517 -522

[6] R. Chandramouli and N. Memon, "Analysis of LSB based Image Steganography Techniques", Proceedings of ICIP 2001, Thessaloniki, Greece, 2001.

[7] J. Fridrich, R. Du, and L. Meng, "Steganalysis of LSB Encoding in Color Images," Proceedings IEEE International Conference on Multimedia and Expo, New York ,2000

[8] J. Fridrich, M. Goljan, and R. Du, "Steganalysis based on JPEG compatibility," SPIE Multimedia Systems and Applications IV, August 20–24, 2001.

[9] N. F. Johnson and S. Jajodia, "Steganography: Seeing the Unseen," *IEEE Computer*, February, 26–34,1998

[10] Jessica Fridrich, Miroslav Goljan. Practical Steganalysis of Digital Images – State of the Art. Preprint 2001.

[11] N. Provos and Peter Honeyman, "Detecting Steganographic Content on the Internet", CITI Technical Report 01-11, 2001,

[12]N. Provos, "Defending Against Statistical Steganalysis", 10th USENIX Security Symposium, Washington, DC, August 2001.

[13] Xuefeng Wang and Yingjun Feng. A New Learning Algorithm for Neural Networks. Journal of Harbin institute of technology. 29(2):23-25

[14] Xuefeng Wang and Yingjun Feng. A Fast Learning Algorithm of Multi-Layer Neural Network. OR Transactions. 2(3):25-29 1998

[15] B. Chen and G. W. Wornell. Implementations of quantization index modulation methods for digital watermarking and information embedding of multimedia. Special Issue on Multimedia Signal Processing, vol. 27:7-33, 2001

[16] Voloshynovskiy, S.,Herrigel and Rytsar Y. StegoWall: Blind statistical detection of hidden data. Proceedings of SPIE. Vol. 4675:57-68.2002