International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:6, 2007

Steganalysis of Data Hiding via Halftoning and Coordinate Projection

Woong Hee Kim, and Ilhwan Park

***Abstract*—**Steganography is the art of hiding and transmitting data through apparently innocuous carriers in an effort to conceal the existence of the data. A lot of steganography algorithms have been proposed recently. Many of them use the digital image data as a carrier. In data hiding scheme of halftoning and coordinate projection, still image data is used as a carrier, and the data of carrier image are modified for data embedding. In this paper, we present three features for analysis of data hiding via halftoning and coordinate projection. Also, we present a classifier using the proposed three features.

***Keywords*—**Steganography, steganalysis, digital halftoning, data hiding.

I. Introduction

In recent years, there has been a lot of interest in steganography and steganalysis. Steganography is the art of hiding and transmitting data through apparently innocuous carriers in an effort to conceal the existence of the data [1]. The digital image data such as BMP, JPEG, and GIF are usually used as a carrier for steganography. Data hiding via halftoning and coordinate projection is proposed by Wu [2]. In this algorithm, still image data is used for data hiding, and the carrier image data are modified during data hiding process, halftoning and coordinate projection.

We present three features for analysis of data hiding via halftoning and coordinate projection. We extract three features using histogram data. After embedding data using that algorithm, the histogram of the carrier image is changed significantly. To extract the features, we transformed the histogram data into frequency domain. The first feature is the ratio of between the sum of higher frequencies power and the sum of lower frequencies power in the frequency domain of the histogram of the carrier image data. The second feature is the position of second peak in the frequency domain of the histogram of the carrier image data. The final proposed feature is the difference of the first moment. Also we design a classifier using the proposed three features, which uses EBP (Error Back Propagation) algorithm [3].

This paper is organized as follows. In section II, data hiding scheme via halftoning and coordinate projection is explained. Section III presents the proposed features for steganalysis and architecture of classifier. Section IV shows experimental results using the proposed method. Finally, section V concludes the paper.

Woong Hee Kim is with the NSRI, 161, Gajeong-dong, Yuseong-gu Daejeon, 305-350, South Korea (e-mail: whkim@ etri.re.kr).

Ilhwan Park is with the NSRI, 161, Gajeong-dong, Yuseong-gu Daejeon, 305-350, South Korea (e-mail: ihpark@ etri.re.kr).

II. Data Hiding Scheme

Data hiding via halftoning and coordinate projection algorithm is proposed by Wu [2]. This algorithm uses a digital still image as a carrier for data hiding. This data hiding scheme is composed of two algorithms, data embedding algorithm and data extraction algorithm. Given a source image M_0 and additional images $M_1, M_2, \cdots, M_k$, the data embedding algorithm embeds $M_1, M_2, \cdots, M_k$ into the source image M_0 , which results in a modified source image with embedded data M_0' [2]. The data extraction algorithm extracts the embedded images $M_1, M_2, \cdots, M_k$ from the modified source image M_0' [2].

A simple description of this algorithm is as follows [2]. First of all, this data embedding scheme generates a set C of extended colors. This algorithm assumes that both the sender and the receiver share the algorithm to generate a set C or share a set C . For example, a set C could be $C = (a, reverse(a))$ where $reverse(a)$ is a with the bits reversed. An element of a set C is selected using distance measure such as Euclidean distance. Suppose that two images, M_0 and M_1 , are given, and M_0 is the source image. Also, let $C = \{(0,3), (1,10), (3,20)\}$. If the data or pixel value of the M_0 is 1, and the data of M_1 is 3, we select $(0,3)$ from C . After selecting one element from a set C , the second coordinate value is discarded, which is called coordinate projection. Since the receiver has a set C , the receiver could recover the element, $(0,3)$ from only single coordinate value, (0) . Fig. 1 shows the process of data embedding [2].

III. Features for Steganalysis and Design of Classifier

In this section, we present three features for analysis of data hiding via halftoning and coordinate projection algorithm, and they are used for a classifier. We extract these features in the frequency domain of the histogram. The data embedding algorithm changes the image data during halftoning and coordinate projection processes. Fig. 2 is an innocuous

256x256 gray image, and Fig. 3 shows the histogram of it. Fig. 4 is the histogram of the image after embedding Fig. 5 into Fig. 2. As shown in the Fig. 4, after embedding data using halftoning and coordinate projection algorithm, the histogram of the image is changed significantly. We extract three features to detect the images which have secret messages in the frequency domain of the histogram of an image. The proposed features are as follows.

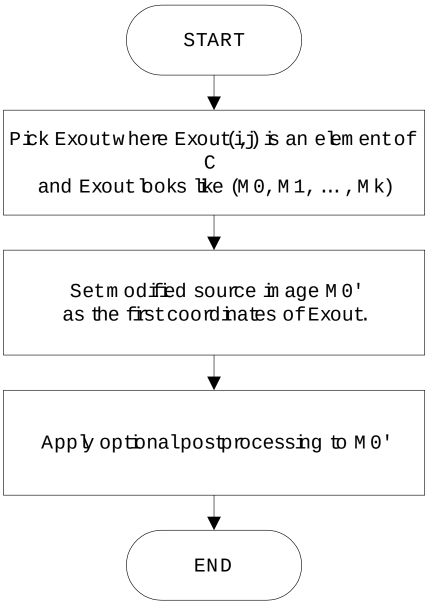

Fig. 1 Data Embedding Algorithm of Halftoning and Coordinate Projection

Fig. 2 Carrier Image

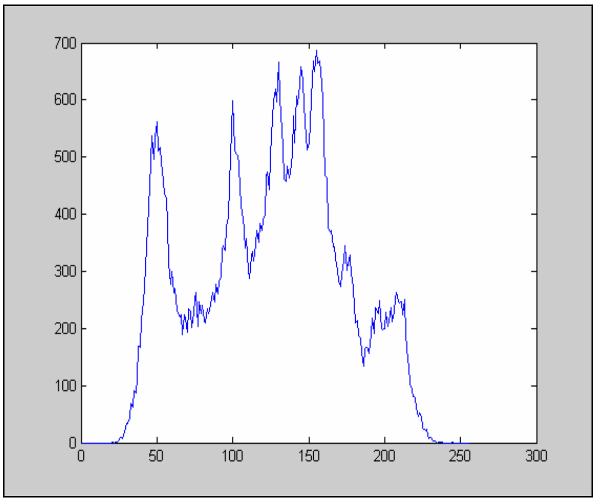

Fig. 3 Histogram of Carrier Image

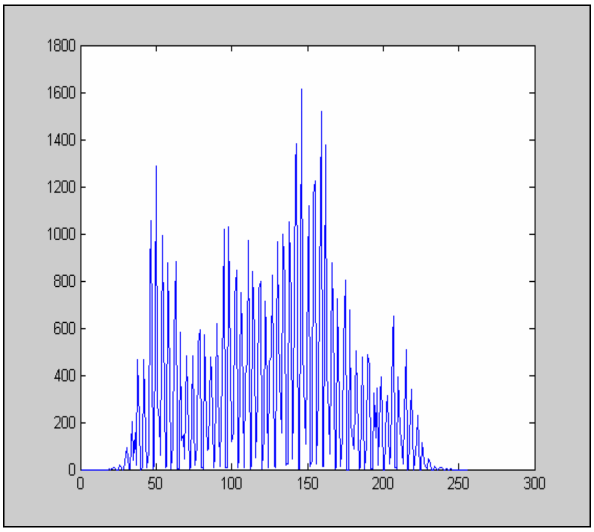

Fig. 4 Histogram of the Modified Image

A. *Feature1 (FT1)*

The first feature (FT1) is the ratio of between the sum of higher frequencies and the sum of lower frequencies. As shown in the Fig. 6 and Fig. 7, the frequency spectrum of the histogram of the carrier image is significantly different from the frequency spectrum of the histogram of the modified image. The power of the higher frequency components is much greater than that of lower frequency components in the modified image. Other images show the similar characteristic as Fig. 6 and Fig. 7. Thus, we select the first feature (FT1) as the ratio of between the sum of higher frequencies and the sum of lower frequencies.

B. *Feature2 (FT2)*

The second feature (FT2) is the position of the second peak in the frequency domain of the histogram. As shown in the Fig. 6 and Fig. 7, the first peak is DC component in the frequency spectrum of the histogram of the image. However, the position of the second peak in the frequency domain is much different.

In most case, the second peak of the modified image is located in the much higher frequency. Many other images have the similar characteristic as this image. Therefore, we choose the second feature (FT2) as the position of the second peak in the frequency domain of the histogram.

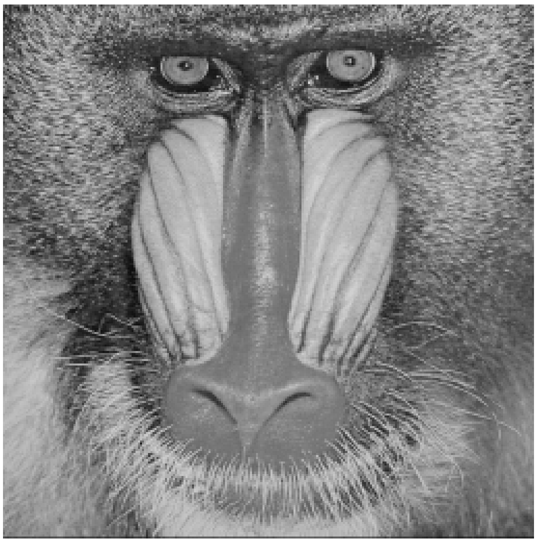

Fig. 5 Embedded Image

C. Feature3 (FT3)

The final feature (FT3) is the difference of between the first moment of image I and modified image I' in the frequency domain where I is the image to be tested, and I' is a modified test image using data hiding via halftoning and coordinate projection algorithm. In other words, when we test an image, we don't know whether the image has secret messages or data. Assuming that the image (I) to be test does not have secret messages or data, we embed some data into I , which results in another image I'. If I does have secret message or data, the first moment of the histogram in the frequency domain does not change significantly. Otherwise, there would be greater change in the first moment of the histogram in the frequency domain.

$$FT3 = |E_1 - E_2|$$

where

$$E_1 = \frac{1}{128}\sum_{i=0}^{127} X_i ,$$

$$E_2 = \frac{1}{128}\sum_{i=0}^{127} Y_i$$

X_i is the magnitude of the frequency spectrum of the image to be tested, and Y_i is the magnitude of the frequency spectrum of the modified image using embedding algorithm.

D. Design of Classifier

We design a classifier using the presented features to detect images with hidden messages or data. The classifier uses EBP (Error Back Propagation) algorithm [3]. The classifier has two hidden layers. Fig. 8 shows presented features and the architecture of detector.

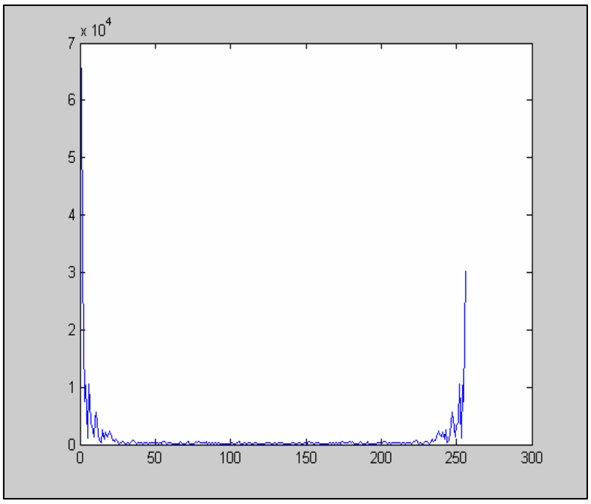

Fig. 6 Frequency Spectrum of the Histogram of the Carrier Image before Data Embedding

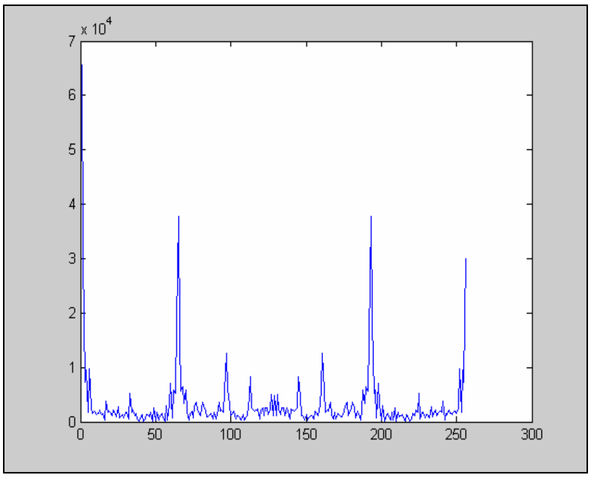

Fig. 7 Frequency Spectrum of the Histogram of the Carrier Image after Data Embedding

IV. Experimental Results

This section describes the experimental results on the proposed features and classifier. We have implemented a classifier using the presented features, which uses EBP algorithm. In the experiment, we train the classifier using 250 train images, which have different embedded images. Fig. 8 is the overall architecture of the designed detector for the suspicious images. Fig. 9 shows the learning curves. As we expected, the error of the detector is decreased as the epochs increases. We test the detector using 750 images, and the result is 6 wrong reports. Thus, the detection performance of the detector using the presented features is 93.33%. If we have much more training images, the performance of the detector would be better.

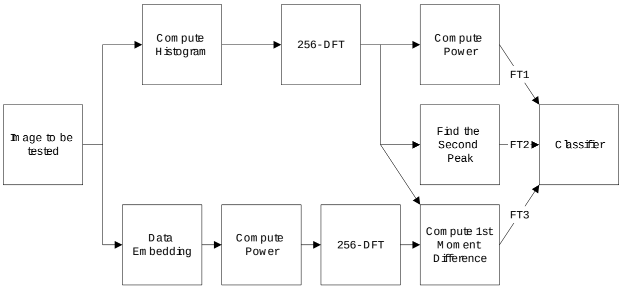

Fig. 8 Architecture of Detector

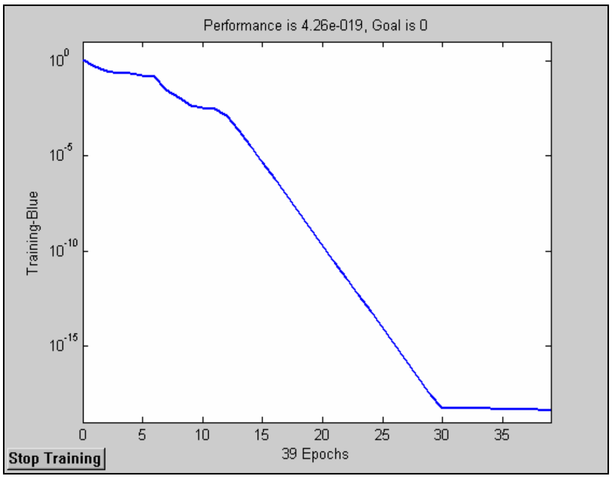

Fig. 9 Learning Curves of the Classifier Using the Presented Features

V. Conclusion

We present three features for analysis of data hiding via halftoning and coordinate projection in this paper. We extract these three features from the frequency domain of the carrier image histogram. Data hiding via halftoning and coordinate projection algorithm modifies still image data during data embedding, and this changes the histogram significantly. The proposed features are the ratio of between sum of higher frequencies and the sum of lower frequencies, the position of the second peak in the frequency domain of the histogram, and the difference of the first moment. We design a classifier using theses three features, which uses EBP algorithm. Experimental results show that the classifier could be used for detecting the suspicious images.

References

[1] Neil F. Johnson, Zoran Duric, and Sushil Jajodia, *Information Hiding : Steganography and Watermarking – Attacks and Countermeasures*, MA, Kluwer Academic Publishers, 2001.

[2] Chai Wah Wu, "Multimedia Data Hiding and Authentication via Halftoning and Coordinate Projection," *EURASIP Journal on Applied Signal Processing*, 2002:2, pp. 143-151, 2002

[3] Simon Haykin, *Neural Networks*, NJ, Prentice-Hall, 1999.