STEGANALYSIS OF HALFTONE IMAGES

[1]*Ming Jiang,* [1]*Edward K. Wong,* [1]*Nasir Memon,* [2]*Xiaolin Wu*

[1]Dept. of Computer and Inf. Science
Polytechnic University
Brooklyn, New York 11201

[2]Dept. of Electrical & Comp. E.
McMaster University
Hamilton, Ontario, L8G 4K1

ABSTRACT

In this paper, we present a novel steganalysis technique for halftone images without knowledge of the original cover image. We first convert halftone images into grayscale-like images by low-pass filtering. The low-pass-filtered image is then decomposed using quadrature mirror filters, and a set of subband coefficients are generated at different scales and orientations. Next, a set of statistical features are computed from the subband coefficients and their predicated errors. Using Fisher linear discriminant analysis, a statistical classifier is designed to detect marked images. Experimental results demonstrate the effectiveness and accuracy of the proposed technique.

1. INTRODUCTION

Steganography is the science of inconspicuously hiding data within data. Although steganography is an old subject, its modern version was formulated by Simmons as the *prisoners' problem* [1] where Alice and Bob, two prison inmates covertly communicate by embedding a *secret message* M into a *cover-object* C, to obtain the *stego-object* S. The stego-object S is then sent through the public channel. Wendy, the warden, who examines the stego-object is unaware of the embedded message M within S and hence permits the communication to take place. For a good survey of steganographic techniques, the reader is referred to [2, 3].

Steganalysis, in this context, is the art of detecting and sometimes even decoding hidden data within a given medium. The basic idea behind most steganalysis techniques is that they compute image features that are not typically "normal" in a given candidate image. Based on these features, a steganalysis technique classifies an image as a marked image or a clean image.

In the past few years, we have seen the development of a number of steganalysis techniques for image data. Perhaps the most successful ones are techniques that can detect the presence of LSB embedding. Representative techniques can be found in [4, 5, 6]. The basic idea behind these techniques is to examine the relationship between neighboring pixels. In [7, 8, 9, 10] general purpose or universal steganalysis techniques that are effective for a wide variety of embedding techniques are presented.

THIS WORK WAS SUPPORTED BY AFOSR GRANT F30602-03-C-0091

Several embedding techniques have been developed for half-tone images that can be found routinely in printed matters such as books, magazines, newspapers, printer outputs, etc. These methods can only be used for half-tone images, and are not suitable for other types of binary images. The methods described in [11, 12, 13] embed data during the half-toning process. This requires the original grayscale image. The methods described in [14, 15, 16, 17] embed data directly into the half-tone images after they have been generated. The original grayscale image is therefore not required. For a good survey of embedding techniques for halftone images, the reader is referred to [18, 19].

In this paper, we propose a steganalysis technique for halftone images. To the best of our knowledge, this is the first steganalysis technique developed for halftone images. In Section 2, we present our proposed method. In Section 3, an experiment is conducted to evaluate the performance of our proposed technique, and in Section 4, we make our conclusions.

2. PROPOSED APPROACH

Embedding messages in halftone images introduces "noise," and the resulting stego-images can be detected by statistical classification, regardless of whether the embedding was made during or after the halftoning process. In our method, low-pass filtering is first performed on a candidate halftone image. The filtered image is then recursively decomposed by using quadrature mirror filters. A set of statistical features are extracted at each scale and orientation after decomposition. Using Fisher linear discriminant analysis and a set of training images, we design a statistical classifier that will classify a candidate image as a clean image or a marked image. The classifier used here is similar to the one used in [20].

(a) $\frac{1}{25}$

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

(b) $\frac{1}{273}$

1	4	7	4	1
4	16	26	16	4
7	26	41	26	7
4	16	26	16	4
1	4	7	4	1

Fig. 1. Two types of low-pass filters. (a) Averaging filter; (b) Gaussian filter.

2.1. Low-pass Filtering

The purpose of low-pass filtering is to recover a grayscale-like image from the candidate halftone image. This process is not the same as inverse halftoning. Inverse halftoning techniques do not give much benefit since they often smooth away the "noise" caused by the embedded message. Good low-pass filters should still retain the trace of "noise", namely embedded information, for later steganalysis. Two types of low-pass filters are considered here: averaging filter and Gaussian filter. An averaging filter can be defined using an $m \times n$ template with weights equal to

$$A(x, y) = \frac{1}{m \times n} \tag{1}$$

An isotropic (i.e. circularly symmetric) Gaussian filter has the form:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

We use a template size of 5 by 5 as shown in Figure 1 for these two types of filters.

2.2. Statistical Features

The statistical features used in [20] were proven to be effective for the steganalysis of grayscale images. We choose the same set of features for our proposed technique. These statistical featues can be obtained by decomposing the candidate image with separable quadrature mirror filters (QMFs). The frequency space of the candidate image is split into multiple scales and orientations. This is accomplished by applying separable lowpass and highpass filters along the image axes, thus generating a vertical, a horizontal, a diagonal, and a low frequency subband. By filtering the low frequency subband recursively, we can create the subbands for all subsequent scales. We denote the vertical, horizontal, and diagonal subband coefficients at scale i as $V_i(x, y)$, $H_i(x, y)$, and $D_i(x, y)$, respectively, for $i = 1, ..., P$. The predicted errors for the vertical, horizontal, and diagonal subband coefficients for scales $i = 1, ..., P$ are denoted as $EV_i(x, y)$, $EH_i(x, y)$, and $ED_i(x, y)$, respectively. Given this image decomposition, the statistical features we use consist of the mean, variance, skewness and kurtosis of each of the subband coefficients $V_i(x, y)$, $H_i(x, y)$, $D_i(x, y)$, and each of the predicted errors $EV_i(x, y)$, $EH_i(x, y)$, and $ED_i(x, y)$ over the scales $i = 1, ..., P$. The mean, variance, skewness and kurtosis for $V_i(x, y)$ are defined as follows:

$$\mu = \frac{1}{|D|} \sum_{(x,y)\in D} V_i(x, y) \tag{3}$$

$$\sigma^2 = \frac{1}{|D|} \sum_{(x,y)\in D} (V_i(x, y) - \mu)^2 \tag{4}$$

$$\gamma_1 = \frac{1}{|D|} \sum_{(x,y)\in D} \frac{(V_i(x, y) - \mu)^3}{\sigma^3} \tag{5}$$

$$\gamma_2 = \frac{1}{|D|} \sum_{(x,y)\in D} \frac{(V_i(x, y) - \mu)^4}{\sigma^4} \tag{6}$$

where D is the domain of (x, y). The mean, variance, skewness and kurtosis for $H_i(x, y)$, $D_i(x, y)$, $EV_i(x, y)$, $EH_i(x, y)$, and $ED_i(x, y)$ can be similarly defined. These statistical features form a multidimensional feature vector. For more details about these statistical features, the reader is referred to [20].

2.3. Classification with Fisher Linear Discriminant Analysis

Using Fisher linear discriminant analysis (FLD), we design a two-class classifier to classify clean and marked images based on the multidimensional feature vector we obtained in Section 2.2. A set of images with and without hidden messages is used to train the FLD classifier. Fisher linear discriminant is a classification method that projects high-dimensional data onto a line and performs classification in the one-dimensional space. The projection maximizes the distance between the means of the two classes while minimizing the variance within each class. That is, we seek to maximize the Fisher criterion $J(\vec{w})$ over all linear projections $\vec{w}$,

$$J(\vec{w}) = \frac{|\mu_1 - \mu_2|^2}{\sigma_1^2 + \sigma_2^2} \tag{7}$$

where μ represents a mean, σ^2 represents a variance, and the subscripts denote the two classes. Maximizing this criterion yields the following closed form solution that involves the inverse of a covariance-like matrix.

$$\vec{w} = S_W^{-1}(\vec{\mu}_1 - \vec{\mu}_2) \tag{8}$$

where S_W^{-1} is the inverse of the total within-class scatter matrix. The total within-class scatter matrix S_W is symmetric and positive semi-definite, defined as

$$S_W = S_1 + S_2 \tag{9}$$

where $S_i, i = 1, 2$ are the individual within-class scatter matrices for classes 1 and 2 defined as follows:

$$S_i = \sum_{\vec{x} \in X_i} (\vec{x} - \vec{\mu}_i)(\vec{x} - \vec{\mu}_i)^t \tag{10}$$

where X_i denotes the set of all class i feature vectors. After the FLD projection axis $\vec{w}$ and the decision boundary is determined from the training set, a test image is projected onto the same 1-D subspace

$$z = \vec{w}^t \vec{x} \tag{11}$$

and then classified as a "clean" or a "marked" image. For more details on FLD classification, the reader is referred to [21].

3. EXPERIMENTAL RESULTS

To validate the effectiveness of our proposed technique, an experiment was conducted to detect hidden messages embedded with the AWST (Authentication Watermarking by Self Toggling) technique described in [17]. Their method embeds a binary logo image into a halftone image at randomly selected locations generated by a pseudo-random number generator. The technique is suitable for embedding dispersed-dot halftone images. However, as the number of pixels being altered is very small, the technique can be applied to any halftone image without causing a noticeable loss of quality.

We used a computer program supplied by the authors of [17] to generate a set of 720 stego images with embedding rates ranging from $0 \leq \alpha \leq 1$. An embedding rate of 0.0 indicates a clean image, and an embedding rate of 1.0 indicates a fully embedded image. Shown in Figure 2 are several example images taken from the test image set. In our experiment, the averaging filter was used because it was shown that it generates better grayscale-like images than the Gaussian filter for steganalysis purposes. The filtered images were then recursively decomposed into four scales (i.e., $P = 4$) using separable quadrature mirror filters. We then applied our FLD classifier to the set of 720 test images. The detection rates were 98.25% for clean images and 97.12% for stego images.

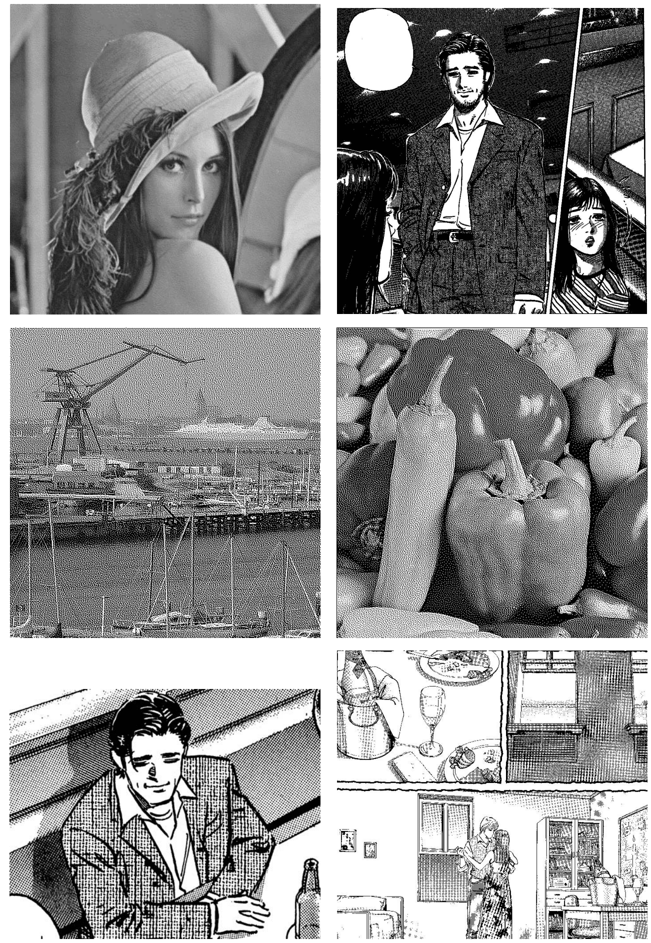

Fig. 2. Sample test images.

4. CONCLUSIONS

Information can be embedded into halftone images in ways that are imperceptible to the human eye, and yet, these manipulations can significantly alter the underlying statistics of the cover images. To detect the presence of hidden messages in halftone images, a novel steganalysis method for halftone images has been developed. Our proposed technique first converts halftone images into grayscale-like images using low-pass filtering. Higher-order statistics are then computed from the multi-scale decomposition of the obtained grayscale-like images for classification. Our experimental results demonstrated that, for steganalyis purposes, it is not necessary to recover the original grayscale image using sophisticated inverse-halftoning techniques. Obtaining a grayscale-like image using a low-pass filtering operation would be sufficient. Our technique is universal (or general,) and can be used for the steganalysis of stego images generated by any halftone image embedding technique.

To further improve the detection accuracy, a better set of statistical features can be designed. The statistical features used in our technique are the same as those used in [8] and were proven to be effective for the steganalysis of grayscale images. However, as pointed out in [8], they are not optimal and it would be beneficial to choose a set of statistics that optimize detection rates. Also, increasing the size of the training set could help in the design of a more accurate classifier.

5. REFERENCES

[1] G. J. Simmons, "Prisoner's problem and the subliminal channel," in *Advances in Cryptology: Proceedings of CRYPTO '83*.

[2] S. Katzenbeisser and F.A.P. Peticolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston, London, 2000.

[3] R. Chandramouli, M. Kharrazi, and N. Memon, "Image steganography and steganalysis : Concepts and practice," in *Second International Workshop on Digital Watermarking*, Seoul,Korea, October 2003.

[4] A. Westfeld and A Pfitzmann, "Attacks on steganographic systems," in *Proc. Third International Workshop on Information Hiding*, Dresden, Germany, Sept. 1999, pp. 61–75.

[5] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of lsb steganography in color and grayscale images," in *Proc. of the ACM Workshop on Multimedia Secuity*, Ottawa, CA, May 2001, pp. 27–30.

[6] Sorina Dumitrescu, Xiaolin Wu, and Zhe Wang, "Detection of lsb steganography via sample pair analysis," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1995–2007, July 2003.

[7] Ismail Avcibas, Nasir Memon, and Bulent Sankur, "Steganalysis using image quality metrics," in *Security and Watermarking of Multimedia Contents*, San Jose, CA, Feb. 2001.

[8] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," in *Proc. 5th International Workshop on Information Hiding*, Noordwijkerhout, The Netherlands, 2002.

[9] Mehmet U. Celik, Gaurav Sharma, and A. Murat Tekalp, "Universal image steganalysis using rate-distortion curves," in *Security and Watermarking of Multimedia Contents*, San Jose, CA, Jan. 2004.

[10] M. Jiang, E. K. Wong, N. Memon, and X. Wu, "A simple technique for estimating message lengths for additive noise steganography," in *The Eighth International Conference on Control, Automation, Robotics and Vision, ICARCV 2004*, Kungming, China, Dec. 2004, accepted.

[11] Z. Baharav and D. Shaked, "Watermarking of dither half-toned images," in *Proc. of SPIE Security and Watermarking of Multimedia Contents*, 1999, vol. 1.

[12] H. Z. Hel-Or, "Watermarking and copyright labeling of printed images," *Journal of electronic imaging*, vol. 10, no. 3, pp. 794–803, 2001.

[13] H-C A. Wang, "Data hiding techniques for printed binary images," in *Proc. Int'l Conf. on Information Technology: Coding and Computing*, April 2001.

[14] E. Koch and J. Zhao, "Embedding robust labels into images for copyright protection," in *Proc. International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies*, Vienna, Aug. 1995.

[15] M. S. Fu and O. C. Au, "Data hiding in halftone images by conjugate error diffusion," in *Proceedings of the 2003 International Symposium on Circuits and Systems*, May 2003, vol. 2.

[16] Soo-Chang Pei and Jing-Ming Guo, "Data hiding in halftone images with noise-balanced error diffusion," *IEEE Signal Processing Letters*, vol. 10, no. 12, pp. 349–351, Dec. 2003.

[17] H.Y. Kim and A. Afif, "Secure authentication watermarking for binary images," in *Proc. SIBGRAPI 2003-XVI Brazilian Symposium on Computer Graphics and Image Processing*, Vienna, Oct. 2003.

[18] Ping Wah Wong and Nasir D. Memon, "Image processing for halftones: Focusing on inverse halftoning, compression, and watermarking," *IEEE SIGNAL PROCESSING MAGAZINE*, July 2003.

[19] M. Chen, E. K. Wong, N. Memon, and S. Adams, "Recent developments in document image watermarking and data hiding," in *Proc. SPIE Conf on Multimedia Systems and Applications IV*, Denver, CO, Aug 2001.

[20] H. Farid, "Detecting hidden messages using higher-order statistical models," in *International Conference on Image Processing (ICIP)*, Rochester, NY, 2002.

[21] Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*, Wiley-Interscience, 2000.