

STEGANALYSIS OF DEGRADED DOCUMENT IMAGES

Ming Jiang, Edward K. Wong, Nasir Memon

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

Xiaolin Wu

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

ABSTRACT

In this paper, a steganalysis technique using compression bit rate as a distinguishing statistic is presented to detect secret messages embedded in document images that are degraded in quality by printing, photocopying, and/or scanning processes. We consider embedding techniques that flip pixels in binary document images that contain characters and symbols. Noise introduced by printing, photocopying, and/or scanning can be modelled by a local optical distortion process. Steganographic embedding is modelled as an additive noise process and we use compression bit rate as a distinguishing statistic to discriminate between stego images and unmarked images. Experimental results showed that the proposed technique can detect stego images with reasonably good accuracy, given the inherent difficulty of the problem.

1. INTRODUCTION

Steganography is the science of inconspicuously hiding data within data. Although steganography is an old subject, its modern version was first 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. A good survey of steganographic techniques can be found in [2].

Steganalysis, in this context, is the art of detecting and sometimes even decoding hidden data within a given medium [3]. The basic idea behind most steganalysis techniques is by computing image features that are typically not "normal" in the input image, and then classifies the image as *stego* or *unmarked* based on the computed features.

Over recent years, many steganalysis techniques for grayscale and color images have been proposed. Although many

techniques for embedding data in binary document images are now known [4], there have only been a few steganalysis techniques proposed for binary document images [5, 6, 7]. In [5], an auto-regressive model was used to detect hidden data embedded along the boundaries of characters and symbols in binary document images. In [6], compression bit rate was used as a distinguishing measure to detect hidden messages in binary document images. In [7], the average pattern for a group of similar patterns in a document is computed and used to detect hidden data.

The techniques proposed in [5, 6, 7] work well only for electronic document images containing fonts and symbols of good image quality. When a document image goes through printing, photocopying, and/or scanning, the quality of the document is degraded and noise is introduced throughout the image, including at locations along the boundaries of characters and symbols. Data embedded by a steganographic technique would mix with noise caused by printing, photocopying and/or scanning, and become harder to detect.

In this paper we propose a steganalysis technique for document images that have been degraded by printing, photocopying, and/or scanning processes. Noise introduced by printing, photocopying, and/or scanning is modelled by a local optical distortion process. Steganographic embedding is modelled as an additive noise process and we use compression bit rate as a distinguishing statistic to discriminate between stego images and unmarked images.

Steganographic techniques for document images include those based on text line, word, or character shifting, boundary modifications, partitioning of images into blocks and selectively flipping image pixels, modifications of run-length patterns, or modifications of half-tone images. Reference [4] provides a good survey of these techniques. In this paper, we are mainly concerned with steganographic techniques that embed data by *flipping* pixels. By flipping we mean changing a white pixel to black and vice versa. Furthermore, due to perceptibility constraints, most techniques flip pixels only along character or symbol boundaries. References [8, 9, 10] are representative techniques that flip pixels along character or symbol boundaries in the embedding process.

THIS WORK WAS SUPPORTED BY AFOSR GRANT F30602-03-C-0091. CONTACT AUTHOR IS EDWARD K. WONG WONG@POLY.EDU

The rest of this paper is organized as follows. In the next section we present a degradation model for printing, photocopying, and/or scanning. In Section 3, we make justification for choosing compression bit rate as a distinguishing statistic. In Section 4, we present our proposed steganalysis method. In Section 5 we present experimental results, and we give our conclusion and future work in Section 6.

2. DOCUMENT DEGRADATION MODEL

Document degradation due to printing, photocopying and/or scanning was modelled in [11, 12] as an optical distortion process. The model accounts for the pixel inversion (from foreground to background and vice versa) that occurs independently at each pixel due to light intensity fluctuations, sensitivity of the sensors, the thresholding level, and the blurring that occurs due to the point-spread function of the scanner optical system.

They modelled the probability of a pixel changing from its ideal value as a function of the distance of that pixel from the boundary of a character. Let d be the distance of a foreground or background pixel from the boundary and let β and γ be the scale parameters. Let $P(1|d, \beta, f)$ and $P(0|d, \beta, f)$ be the probability of a foreground pixel at a distance d from the boundary to remain as 1 and to change to 0, respectively. Similarly, let $P(1|d, \gamma, b)$ and $P(0|d, \gamma, b)$ be the probability of a background pixel at a distance d from the boundary changing to a 1 and remaining as 0, respectively. The functions $P(1|d, \beta, f)$ and $P(1|d, \gamma, b)$ could be different. A random perturbation process then proceeds to change pixel values on a document in an independent manner. The foreground and background conditional probabilities can be described by the equations below:

$$P(0|d, \beta, f) = \beta_0 e^{-\beta d^2} + \eta_f \quad (1)$$

$$P(1|d, \gamma, b) = \gamma_0 e^{-\gamma d^2} + \eta_b \quad (2)$$

where β_0 and γ_0 are the initial values for the exponentials; β and γ control the decay speed of the exponentials; η_f and η_b are the uniform probabilities of a foreground or background pixel flipping, respectively. After pixel inversion, a morphological *closing* operation with a disk structuring element of diameter k is used to account for the correlation introduced by the optical point spread function preceding the thresholding operation which produces the noisy image. Figure 1 shows an example character image before and after degradation using the optical distortion model.

3. COMPRESSION BIT RATE AS A DISTINGUISHING STATISTIC

We adopt the general quantitative steganalysis methodology for digital images proposed by Fridrich et. al. [13]. Based



Fig. 1. The effect of degradation. (a) Original image; (b) After degradation.

on this approach, our goal is to identify a good distinguishing statistic of an image that predictably changes with the length of the embedded secret message, even in the presence of noise caused by printing, photocopying, and/or scanning.

As we pointed out in [14], any data embedding technique can essentially be viewed as an additive process. That is, we can write $S = C + M$, where C , S , and M are the original cover signal, the stego signal, and the embedded message respectively. If we assume the message bits M are *i.i.d.* and independent of C , it is clear that for any embedding process modelled as an additive noise process, the stego signal has a higher entropy than that of the cover signal [15], that is $H(C) \leq H(S)$. Actually, we can say more. In all information hiding systems where the embedded message is independent of the cover signal, the entropy of the stego signal is a monotonically increasing function of the embedded signal strength, that is

$$H(S_{\alpha_1}) \leq H(S_{\alpha_2}) \quad (3)$$

where $H(S_{\alpha_1})$ and $H(S_{\alpha_2})$ are the entropy of stego signals when the embedded message strengths, or embedding rates, are α_1 and α_2 respectively, and $\alpha_1 \leq \alpha_2$.

Now, the entropy of a signal is difficult to determine as often we do not have an adequate model for the signal. However, we do know that entropy and compression are closely related. A perfect compression technique encodes at a rate equal to the entropy. Hence it is reasonable to use compression bit rate as an estimate for signal entropy and consequently as a distinguishing statistic for the purpose of steganalysis. In our proposed method, we specifically use the JBIG-2 compression algorithm to compress document images.

4. PROPOSED STEGANALYSIS METHOD

In the proposed steganalysis method, we argue that noise introduced by printing, photocopying and/or scanning processes cause a certain amount of increase in the compression bit rate of the original electronic document image. Embedding a noisy image with hidden data will further increase the compression bit rate. We assume randomness in the embedding process; i.e., the message bits are randomly spread throughout the entire cover signal. For pixel-flipping data

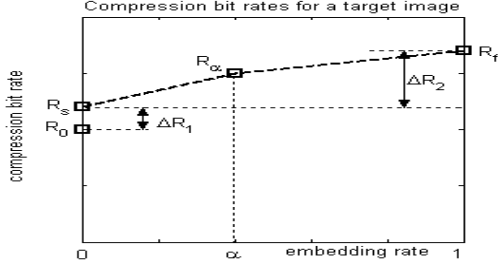


Fig. 2. Compression bit rates of four interesting points computed from a stego image

embedding methods, when a stego image already has a message embedded in it, the flipping rate does not change when we embed into the already embedded portion.

Based on the above, our algorithm computes the compression bit rate at four interesting points as shown in Figure 2, where α is the embedding rate. The meanings of the four points are described below:

- R_α is the compression bit rate of the stego image with an unknown embedding rate of α (possibly zero) that we are trying to determine.
- R_0 is the compression bit rate of the original electronic version of the cover image, before degradation from printing, photocopying, and/or scanning. As we do not have the cover image and its original electronic version, R_0 can only be estimated. We do so by applying a morphological *opening* operation to the given stego image, producing a *reset* image. It was shown in [6] that the compression bit rate of the reset image is approximately the same as the compression bit rate of the cover image, which does not contain noise caused by printing, photocopying, and/or scanning. In the case of degraded document images that we are considering here, the compression bit rate of the reset image approximates that of the original electronic version of the cover image.
- R_f is the compression bit rate of the stego image when it is randomly embedded at full capacity, or when $\alpha = 1.0$.
- R_s is the compression bit rate of the cover image, a degraded version of the original electronic document image after printing, photocopying, and/or scanning. Since we do not have the cover image, we will have to estimate its compression bit rate. We use the simulation software from the authors of [11] to simulate the printing, photocopying, and/or scanning processes and generate degraded document images from the *reset* image. Using different parameter values in Equations 1 and 2, document images with different degree

of degradations can be generated. We call the images thus generated as *synthesized* images. For each synthesized image, we embed data at full capacity ($\alpha = 1.0$) and compress using the JBIG-2 algorithm. We search for the parameter values such that the compression bit rate of the synthesized image embedded at full capacity equals R_f , or the compression bit rate of the stego image at full capacity. The search is done by iterate through the parameter values $\beta_0, \beta, \eta_f, \gamma_0, \gamma,$ and η_b with a step size of 0.02. We argue that when the compression bit rate of the synthesized image embedded at full capacity equals R_f , the synthesized image is a good approximation of the cover image and R_s can be approximated by compressing the synthesized image.

Observe that in Figure 2, ΔR_1 represents increase in compression bit rate due to noise from the printing, photocopying, and/or scanning processes, and ΔR_2 represents increase in compression bit rate due to embedding at full capacity. Denoting the estimate for R_s as \hat{R}_s , we use the following measure for deciding whether a candidate image is marked or unmarked:

$$\lambda = \frac{R_\alpha - \hat{R}_s}{R_f - \hat{R}_s} \quad (4)$$

where $0 \leq \lambda \leq 1$. If $\lambda \geq \tau$, we classify the candidate image as marked, otherwise as unmarked. τ is a predefined threshold obtained empirically. Increasing τ would decrease the false positive rate, but at the same time increase the false negative rate. Denoting the estimate for R_0 as \hat{R}_0 , we can compute ΔR_1 as $\hat{R}_s - \hat{R}_0$. The value ΔR_1 gives us a quantitative measure on the amount of printing, photocopying, and/or scanning noise that are present in the candidate image.

5. EXPERIMENTAL RESULTS

An experiment was conducted to validate the effectiveness of the proposed steganalysis method. We used a set of 50 images from the UW-III English Document Database CD-ROM [16] as test images in our experiment. The images are marked with different embedding rates using a data hiding program from the authors of [9]. The estimated compression bit rate \hat{R}_s of the cover image has error (defined as $(R_s - \hat{R}_s)/R_s$) within the range of -2.2% to $+3.4\%$. Final steganalysis result is shown in Table 1. From the table, we see that 83.3% of the stego images are correctly classified and 75% of the unmarked images are correctly classified. We view the performance as reasonably good considering that the problem is inherently difficult.

Table 1. Steganalysis Results

	Postive	Negative
Stego images (30)	25 (83.3%)	5 (16.7%)
Unmarked images (20)	15(75%)	5(25%)

6. CONCLUSION AND FUTURE WORK

We have presented a steganalysis technique for document images degraded by printing, photocopying, and/or scanning processes. Detection of these stego images would be difficult with other known steganalysis techniques. Our experimental results showed that our method can detect stego images with reasonably good accuracy given that the problem is inherently difficult. We view the proposed technique as a first step in the development of good steganalysis techniques for document images degraded by printing, photocopying, and/or scanning. In future work, we will attempt to improve the accuracy of the proposed technique and find a reliable method to estimate the length of the hidden message.

7. ACKNOWLEDGEMENT

We would like to thank Prof. R. M. Haralick at the City Univ. of New York for providing us the document degradation simulation software and the UW-III English Document Database CD-ROM. We would also like to thank Prof. M. Wu at the Univ. of Maryland, College Park for kindly providing us her watermarking software.

8. 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. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston, London, 2000.
- [3] J. Fridrich and M. Goljan, "Practical steganalysis of digital images - state of the art," in *Proc. SPIE Conf. on Security and Watermarking of Multimedia Contents*, Portland, Oregon, USA, April 2002, vol. 4675, pp. 1–13.
- [4] 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.
- [5] M. Jiang, X. Wu, E. K. Wong, and N. Memon, "Steganalysis of boundary-based steganography using autoregressive model of digital boundaries," in *IEEE Int'l Conf. on Multimedia and Expo*, Taipei, Taiwan, June 2004.
- [6] M. Jiang, N. Memon, E. K. Wong, and X. Wu, "Quantitative steganalysis of binary images," in *Proc. IEEE Int'l Conf. on Image Processing*, Singapore, Oct. 2004.
- [7] J. Cheng, A. C. Kot, J. Lou, and H. Cao, "Steganalysis of Binary Text Images," in *Proc. IEEE ICASSP*, March 2005, pp. IV–689 to IV–692.
- [8] Q. Mei, E. K. Wong, and N. Memon, "Data hiding in binary text documents," in *SPIE Proc Security and Watermarking of Multimedia Contents III*, San Jose, CA., Jan 2001, vol. 2.
- [9] M. Wu, E. Tang, and B. Liu, "Data hiding in digital binary images," in *Proc. IEEE Int'l Conf. on Multimedia and Expo*, New York, Aug 2000.
- [10] H. Lu, A. C.Kot, and J. Cheng, "Secure data hiding in binary document images for authentication," in *Proc. IEEE ISCAS*, May 2003, vol. 3, pp. 806–809.
- [11] T. Kanungo, R.M. Haralick, and I. Philips, "Global and local document degradation models," in *Proc. Second Int'l Conf. Document Analysis and Recognition*, Oct. 1993, pp. 730–734.
- [12] T. Kanungo, *Document Degradation Models and a Methodology for Degradation Parameter Validation*, Ph.D. thesis, 1996, Univ. of Washington, Seattle.
- [13] J. Fridrich, M. Goljan, D. Hoge, and D. Soukal, "Quantitative steganalysis of digital images: Estimating the secret message length," *ACM Multimedia Systems Journal: Special issue on Multimedia Security*, vol. 9, no. 3, pp. 288–302, 2003.
- [14] M. Jiang, E. K. Wong, N. Memon, and X. Wu, "A simple technique for estimating message lengths for additive noise steganography," in *Proc. IEEE Int'l Conf. on Control, Automation, Robotics and Vision*, Kungming, China, Dec. 2004.
- [15] M. U. Celik, G. Sharma, and A. M. Tekalp, "Universal image steganalysis using rate-distortion curves," in *Proc. SPIE Conf. on Security and Watermarking of Multimedia Contents*, San Jose, CA, Jan. 2004.
- [16] I. T. Phillips, "UW-III English document image database," <http://documents.cfar.umd.edu/resources/database/3UWCdRom.html>, 1996.