COVER SELECTION FOR STEGANOGRAPHIC EMBEDDING

Mehdi Kharrazi [a], *Husrev T. Sencar* [b], *Nasir Memon* [b]

[a] Dept. of Electrical and Computer Eng., Polytechnic University, Brooklyn, NY, USA.
[b] Dept. of Comp. and Inf. Science, Polytechnic University, Brooklyn, NY, USA.

ABSTRACT

The primary goal of image steganography techniques has been to maximize embedding rate while minimizing the detectability of the resulting stego images against steganalysis techniques. However, one particular advantage of steganography, as opposed to other information hiding techniques, is that the embedder has the freedom to choose a cover image that results in the least detectable stego image. This resource has largely remained unexploited in the proposed embedding techniques. In this paper, we study the problem of cover selection by investigating three scenarios in which the embedder has either no knowledge, partial knowledge, or full knowledge of the steganalysis technique. For example, we illustrate through experiments how simple statistical measures could help embedder minimize detectability, at times by 65%, in the partial knowledge case.

Index Terms— Steganography, steganalysis, cover selection

1. INTRODUCTION

In the past few year, there has been a large number of research done on image based steganography due to the wide availability of images and better understanding of their properties as opposed to other digital contents. In turn the development of such techniques has given rise to a number of steganalysis technique proposed in the literature. The reader is referred to [1] for a comprehensive review of steganography and steganalysis techniques.

Modern steganography is formulated in terms of the simmon's prisoner problem [2], were two inmates, Alice and Bob, communicate a message M to hatch a scape plane, with the warden Wendy monitoring all communications between the two. Much work has been done on how Alice would embed the secret message M in the cover image (i.e. embedding techniques), and in general one would argue that the "better" the embedding technique, the less likely that the stego image would be detected by warden Wendy, where an embedding technique is assumed to be "better" (as compared to others) on the basis of benchmarks such as the one presented in [3]. But in steganography, unlike other information hiding techniques (i.e., watermarking, fingerprinting, etc.), the cover image only acts as a carrier for the message, therefore Alice has the freedom to choose any cover image for the embedding process.

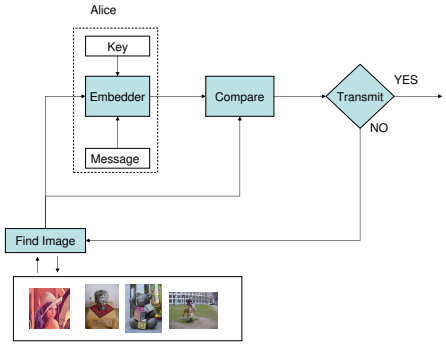

Fig. 1. A block diagram for cover selection.

That is, other than the choice of the embedding technique, Alice is free to select amongst a set of cover images. A general block diagram for cover selection is given in figure 1, in which, after obtaining the stego image, Alice compares the stego and cover images in order to decide whether she would like to transmit the stego image, or opt to select an alternate cover image. By doing so, she could choose cover images with which the resulting stego image would be misclassified (i.e., false negative) by the steganalyzer. Therefore even in the presence of a powerful steganalyzer she has improved her chances of going undetected.

In what follows we go over the problem setting in section 2, and discuss a set of measure which are used to compare the cover and stego images in section 3. We illustrated through a number of experiments in section 4 how employing the discussed measures help the embedder select less detectable cover images. We conclude in section 5 with a discussion of obtained results and future work.

2. PROBLEM SETTING

As noted above, Alice can minimize her chances of detection by choosing stego images which are more likely to be misclassified as cover images by the Warden. To do so, she would need to understand the behavior of the steganalyzer, and imitate its decision making process. The steganalyzer's classification function distinguishes between cover and stego images, based on the location of images with respect to the decision hyper-plane in the feature space. The distinguishability in the feature space is due to the displacement of calculated statistics

from stego images due to the embedding process, or a lack of such displacement in the case of cover images.

In fact since Alice is the embedder, she has available to her both the cover and stego images. Assuming that the embedder has some number of images available and willing to use them in an order, rather than randomly selecting them for the embedding process; We propose a simple but effective ranking technique based on a set of distance measures between the cover and stego images, which would help Alice in choosing stego images that are more likely to yield false negatives. Furthermore, we discuss cover image properties (i.e. compression rate, size) which independent of the embedding process could effect the detectability. One could think of three possible scenarios in which Alice has different levels of information about the steganalyzer:

- *No knowledge:* Here, Alice has no knowledge about the steganalyzer, but she could still increase her chances of remaining undetectable by looking at general features which are likely to *change* during embedding process. For example minimizing the number of JPEG DCT coefficients changed, due to the embedding process.

- *Partial Knowledge:* This would be the case if Alice only had a limited understanding of the steganalyzer. We assume Alice has some form of access to the Warden's steganalyzers, but her access is limited in two ways. First, she has only access to the input and output of the steganalyzer. Thus, she can not find out anything specific about its inner working (i.e., type of features used). Secondly her access is limited in time, although she is allowed to obtain the steganalyzers' decision with any input image she chooses. Here she could be more precise than in the *No Knowledge* case by using the steganalyzer's decisions corresponding to the set of *text* stego images to calculate certain threshold values to be used for cover selection.

- *Full knowledge:* In this case, Alice knows exactly the steganalyzer and the statistical features being used by the Warden. Therefore she could use a data set of cover and stego images to train a similar steganalyzer. Then Alice could test her stego images against the trained steganalyzer and identify images which are misclassified as cover images. In this case, our approach would be irrelevant.

At this point we note that since in our approach we exploit the errors (i.e. false negatives) made by the steganalzyer, the accuracy of our ranking technique depends on it's error rate. If the steganalyzer is error free, then no stego image will go undetected, and our approach will be of no use. In the next section, we discuss a number of cover image properties as well as cover-stego based measure which could be used for cover selection.

3. MEASURES

We propose using a set of standard measures, to capture image properties before or after the embedding process, which effect the performance of steganalysis techniques. These measure are divided into two categories. First cover image properties, and second cover-stego based distortion measures. Below we will go over each category and motivate the selection of each measure.

3.1. Cover based

Independent of the embedding operation, properties of the cover image used, will effect the performance of steganalyzer. Below we will review two of such properties:

- *Changeable Coefficients* are the set of coefficients which will be utilized by the embedding process. Since the message is fixed in the cover selection problem, images with larger number of changeable coefficient will relatively have a smaller number of changes induced by the embedding operation.

- *JPEG Quality Factor* as we have observed trough experimentation, as continuation of our previous benchmarking study in [3], JPEG quality factor is inversely correlated with the performance of steganalyzers. In other words the higher the JPEG quality factor the less is the performance of the studied steganalyzers.

3.2. Cover-stego based

Since we have available to us both the cover and stego images, we are able to measure the embedding artifacts directly. Thus, we are interested in measures which are able to quantify such artifacts. Below we will introduce the cover-stego based measures which we have employed in our work and motivate their selection:

- *Number of Modifications* to the cover image could be thought as the most intuitive. The smaller the number of changes made the less detectable the resulting stego image should be.

- *Mean Square Error (MSE)* is a simple non-perceptual error metric which is obtained from the cover-stego image pairs where lower MSE values are assumed to be indicative of lesser detectability.

- *Prediction Error* is a local measure which we have used in our experiments by looking at the difference between the mean prediction error of the cover and stego image using the prediction model proposed in [4]. Similar to MSE here prediction error is assumed to be correlated with detectability.

- *Watson's* metric [5] is a perceptual measure, which is used in quantifying the quality of JPEG images, therefore detectability should be lower as the difference in Watson's metric between the cover and stego is less.

- *Structural Similarity Measure (SSIM)* [6] is used to quantify the similarity between the cover and stego images. Therefore the larger the similarity measure, one would expect the smaller detectability.

4. EXPERIMENTS

From the large data set of gray scale images obtained in [3], we obtained images which had quality above 85 and a minimum width of 1000 Pixels. These images were then downsized to a size of 640x480 and saved in JPEG format at quality factors of 95 and 75. Thus obtaining 2 data sets with different compression rates, each consisting of about 13000 images. The stego images were created using Outguess(+) [7] embedding techniques with fixed message lengths, which is relative to the number of pixels in the images. Two steganalyzers were employed FBS [8], and WBS [9]. The steganalyzers were trained using a random 30% of stego images and equal number of cover images.

A message length of 0.04 bits per pixel was used with JPEG images compressed at quality factor of 95. A message length of $.05$ bits per pixel was used with JPEG images compressed at quality factor of 75. (The reason that the two stego data sets were created with different parameters was due to the data set dependency of of the steganalysis techniques, which was observed in [3].)

4.1. No Knowledge Case

In the case that the embedder has no knowledge of the steganalyzer in distinguishing between stego and cover images, she is only able to use her intuition in selecting images with minimum embedding artifacts with the hope that the measure she is using is well correlated with the decisions of steganalyzer. In the previous section, we discussed some possible measures which are thought to be correlated with the steganlayzer's output. To test their reliability in selecting covers, we experimented with the discussed measures using the designed steganalyzers and only images unseen to the steganalyzer at the time of training. The experimental procedure is as followed:

1. Randomly select 10% of images, and calculate the number of TP in the chosen set.
2. Sort the images based on a given measure.
3. Select the top or bottom 10% of sorted images, depending on whether the measure is to be maximized or minimized, and calculate the TP rate.

We should note that the 10% value is chosen based on the assumption that the steganalyzer has error rate larger or around 10%. If we where to choose a value smaller (i.e. 5%), then we would have decreased the size of our sample set. Alternatively if we had chosen a value larger like (i.e. 20%), and the classifier had an error rate smaller, then we would have decreased the probability of correctly selecting misclassified images. In step 1, we are calculating the number of images which will be correctly classified as stego if 10% of image were chosen randomly. In step 3, we calculate the number of correctly identified stego image, if we employ our simple ranking scheme. Results obtained for two embedding steganalysis pairs, FBS-Outguess and WBS-Outguess, is presented in table 1.

	Out+(.04)	Out+(.05)
Type	FBS	WBS
AV TP	85.62	82.98
Changeable Coef.	29.11 *	86.55 *
Number of changes	60.44	76.47
MSE	99.14	52.10
Pred. Error	99.48	52.52
Watson	80.99	62.81
SSIM	58.22 *	81.51 *

Table 1. Improvements in choosing a "good" cover in the no knowledge case. The entries marked with (*), were sorted in ascending order. Therefore in such cases, the larger the calculated value the less detectable the stego image. All other entries were sorted in descending order.

From table 1, we could make the observation that the correlation of measures to the decision of the classifier is not the same in the two studied cases. For example by maximizing the number of changeable coefficient the FBS technique performs less accurately, but maximizing that measure has no effect on the performance of the WBS technique.

4.2. Partial Knowledge Case

The second scenario which we have investigated is the partial knowledge case. In this case, we have assumed that the embedder has obtained the steganalyzer's decisions on a number of stego images. Therefore, Alice could use those images, as a training set to obtain threshold value on the previously discussed measures. For example any image which has less that N number of modifications to it as part of the embedding process will be considered as less detectable by the steganalyzer, where N is calculated from the test data set. Using the same data set as in the previous section, the following steps were executed in order to investigate the accuracy of our approach:

- Training stage
  - Using images for which steganalyzers decision are obtained, select the top or bottom 10% of images, which ever set has the lower TP rate
  - Calculate the mean value of the measures obtained from the selected set of images (i.e. 10%)
- Test stage
  - Using only images which were not on the training set, threshold the images for the given measure, and calculate the number of TP in the resulting set

Available pairs	-	10	100	500
AV TP	85.62	85.66	85.64	85.65
Changeable Coef.	29.11*	33.70	24.56	22.01
Number of changes	60.44	74.69	59.18	50.20
MSE	99.14	30.00	21.57	20.39
Pred. Error	99.48	41.39	36.09	33.87
Watson	80.99	93.05	85.64	85.23
SSIM	58.22*	56.94	46.31	47.42

Table 2. Improvements in choosing a "good" cover in partial knowledge in the case of FBS-Outguess. The entries marked with (*), were sorted in ascending order. Therefore in such cases, the larger the calculated value the less detectable the stego image. All other entries were sorted in descending order.

Repeating the above steps in 10 iterations, we calculated the average improvement in undetectability using the two FBS-Outguess and WBS-Outguess pairs. Table 2 presents the results obtained from the Outguess-FBS pair, given no knowledge, and partial knowledge with 10, 100, and 500 images available to the embedder. For example we observe that if we where to choose images based on MSE, and with no knowledge, then we would have decreased the chances of going undetected to null. But by using even 10 images to fine tune the measure and threshold, we could improve the TP rate from 99% to 30%. Similar results were obtained for the WBS-Outguess case as given in 3.

Available Pairs	-	10	100	500
AV TP	82.98	81.80	81.82	81.82
Changeable Coef.	86.55*	69.61	40.38	40.38
Number of changes	76.47	74.94	75.81	79.40
MSE	52.10	62.68	41.82	39.21
Pred. Error	52.52	61.70	48.39	42.27
Watson	62.81	69.86	62.32	58.05
SSIM	81.51*	82.00	82.04	81.61

Table 3. Improvements in choosing a "good" cover for partial knowledge in the case of WBS-Outguess. The entries marked with (*), were sorted in ascending order. Therefore in such cases, the larger the calculated value the less detectable the stego image. All other entries were sorted in descending order.

5. DISCUSSION

In the following work we addressed the problem of cover selection, and investigated a simple but effective solution to the problem. In our setting, Alice has limited access to Wendy's classifier. But, by exploting the information learned from the classifiers decision for a number of test stego images, she is able to select *better* cover images. For this, she searches for a (distortion) measure (along with a threshold value) that divide the test images into two classes imitating the steganalyzer output. We illustrated through experimentation how a number of simple measures could improve the chances of undetectability by the steganalyzer, and investigated the performance due to the use of each measure.

We observed that common distortion measures like MSE and prediction error are not necessarily good measures in quantifying the steganographic embedding distortion. That is, a decrease in distortion measured between the cover-stego pair does not increase the chances of raising a false alarm. The reverse holds true as well, see table 1. This aspect of the problem has to be considered in developing a theoretical framework for steganoraphy. For example with the FBS steganalysis technique, we see that with the increasing MSE values it is more likely that the stego image would be misclassified. (This may be due to the fact that FBS steganalysis technique estimates the cover from the stego image through cropping and re-compressing. The increased distortion, introduced to the cover image, results with a less accurate estimate of the cover, and correspondingly the classifiers ability to distinguish between cover and stego images diminishes.)

For the no knowledge case, our results indicate that minimizing the number of changes made to the cover image serves as a reliable measure. On the other hand, in the partial knowledge case, despite their simplicity, MSE and the number of changeable coefficients seem to be more effective measures in selecting a less detectable cover image. The superiority of these measures over perceptual measures, like SSIM and Watson's metric, also indicate that steganographic embedding distortion has to be quantified in a statistical (rather than perceptual) manner. In addition, as the partial knowledge available to Alice, in terms of the size of test stego set, increases from 10 to 100 the our cover selection method becomes more effective, and the improvement due to increase to 500 is rather marginal. We are currently expanding our work to cover more steganalysis-embedding pairs as well as investigating alternate measures.

6. REFERENCES

[1] M. Kharrazi, H. T. Sencar, and N. Memon, "Image steganography: Concepts and practice," *to appear in Lecture Note Series, Institute for Mathematical Sciences, National University of Singapore*, 2004.

[2] G.J. Simmons, "The prisoners problem and the subliminal channel," *CRYPTO*, pp. 51–67, 1983.

[3] M. Kharrazi, T. H. Sencar, and N. Memon, "Benchmarking steganographic and steganalysis techniques," *EI SPIE San Jose, CA, January 16-20, 2005*.

[4] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Transactions on Communications, Volume 45, Issue 4, April 1997 Page(s):437 - 444*.

[5] Andrew B. Watson, "Dct quantization matrices visually optimized for individual images," *ed. B. Rogowitz and J. Allebach, Human Vision, Visual Processing, and Digital Display IV, Proc. Vol. 1913, pp.202-216, SPIE, Bellingham, WA*.

[6] Z Wang, A C Bovik, H R Sheikh, and E P Simoncelli, "Perceptual image quality assessment: From error visibility to structural similarity," *IEEE Trans Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[7] N. Provos, "Defending against statistical steganalysis," *10th USENIX Security Symposium*, 2001.

[8] J. Fridrich, "Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes," *Proc. 6th Information Hiding Workshop, Toronto, Canada, May 23-25*, 2004.

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