Develop a Detection System for Grey and ColourStego Images

Ahd Aljarf, Saad Amin, John Filippas, and James Shuttelworth

***Abstract*—Recently the concept of 'Image Steganography' is became an important issue in the computer security world. Image steganography simply means hide some secret data into an object. The object can be a text, an image or a sound, but the most popular cover object used for hidden secret message is images.**

A developed detection system is introduced in this paper. The first part of the work includes creating varieties of stego-images. These stego-images are having different image file formats. Also, these stego-images have been done using three steganography tools. They are: OpenStego, S-Tools and F5 algorithm. The created stego-images are used to train the detection system in the next stage.

However, the second part of the work includes detecting the hidden data. In order to do so the co-occurrence matrix is created for all images. Number of image features is extracted from the matrix. Later, a discriminator will be used to validate the features selected. These features are needed it to differentiate between the clean images and the stego-images.

***Index Terms*—Steganography, image steganography, image steganalysis, stego-images.**

I. Introduction

Steganography is an important research issues in computer security field. Steganography is the science of communicating secret data in an appropriate cover object. The word steganography is derived from the Greek words "stegos" meaning "cover" and "grafia" meaning "writing", defining it as "covered writing" [1], [2].

There are four types of the cover objects, which are image, text, audio and video. According to these different covering objects; there are many types of steganography, such as image steganography, steganography in txt files .etc. [1].

The most popular cover object used for hidden secret message is images for many reasons. Images are widespread on the internet; also they can be used as carrier objects without raising much suspicion. In addition, image files have a lot of capacity for modification without noticeable damage to the content [1], [2].

On the other hand, Steganalysis is the art and science of detecting the use of all steganographic techniques. In Steganalysis, the goal is to be able to compare the cover-object, the stego object (the cover message with the hidden data embedded in it) and any possible portions of the stego-key (encryption method) in an effort to intercept, analyze and/or destroy the secret communication [3].

This means the goal of steganography is to hide data into an object; however, the goal of steganalysis is to detect these data.

This paper is focusing on developing a steganalysis method based on some existing techniques. The developing method is working by extracting a group of image features. Then a discriminator will be applied on these features to test the validity of them. In addition, varieties of stego-images have been created in a previous stage to use them later with the developed method. These stego-images are containing one or more hidden file.

II. Category of Steganalysis

Steganalysis has been of interest since the end of 1990's. The aim of steganalysis is to detect secret information hidden in a given image. Steganographic attacks consist of detecting, extracting and destroying hidden object of the stego media. There are few experts in this field; Neil Johnson is one of them. He classifies attacks in six main categories [1].

They are: 1) Stego-only attack: only the stego-image is available for analysis. 2) Known cover attack: the original cover-image and stego- image are both available. 3) Known message attack: at some point, the hidden message becomes known to the attacker. Analyzing the stego-image for patterns that correspond to the hidden message may be beneficial for future attacks against that system. 4) Chosen stego attack: the steganography tool (algorithm) and stego-image are known. 5) Chosen message attack: thesteganalyst generates a stego-image from some steganography tool or algorithm from a chosen message. This goal in this attack is to determine corresponding patterns in the stego-image that may point to the use of specific steganography tool or algorithms. 6) Known stego attack: the steganography algorithm is known and both the original and stego-image are available [1], [4].

III. Related Work

Lyu and Fraid [5] described a steganalysis approach that relies on higher-order image statistics and SVM. This may be the first blind steganalysis for GIF images, but it cannot achieve a good performance against EzStego steganography.

Wang and Gong [6] proposed asteg analysis algorithm based on colours-gradient co-occurrence matrix (CGCM) for GIF images. CGC Miscon structed with colours matrix and gradient matrix of the GIF image, and 27-dimensionalstatisticalfeaturesofCGCM, which are sensitive

Manuscript received May 2, 2013; revised July 2, 2013.

Ahd Aljarf is with Coventry University (e-mail: aljarfa@uni.coventry.ac.uk).

DOI: 10.7763/IJMO.2013.V3.320

tothecolour-correlation between adjacent pixels and the break in go fima get exture, are extracted. This proposed steganalysis algorithm does not requirea lot ofcomputing time.

Arvis *et al.* [7] has proposed a multispectral method considering the correlations between the colour bands. To study the efficiency of their method, they test it in a classification problem on the image databases VisTex and Outex available on the internet. They also extended the co-occurrence method according to the two other approaches, which are: (fusion of texture and colour descriptors and quantization of the colour image) to have a comparison between the three approaches to the texture in colour images.

IV. The Proposed System

The proposed system has been divided into two stages. In stage one, a collection of stego-images is created first as they are needed as applications to test the detection system later as shown in Fig. 1. However, the second stage is focusing on evaluating a set of images (clean and stego-mages) for detecting any hidden data through extracting number of image features as shown in Fig. 2.

A. The Stego-Images Created

Textsand images are embeddedinto clean images as hidden data to produce varieties of stego-images. The three steganographic tools selected for this purpose are OpenStego, StegHide and F5 Algorithm. The three tools are implemented into many clean images to embed data as shown Fig. 1.

The first tool used is OpenStego.It uses (Data Encryption Standard) DES algorithm for data encryption and supports two steganographic plugins, which are LSB (Least Significant Bit of Image Pixels) and Random LSB (Randomized LSB). The second tool used is: S-Tools, it allows audio and image files to be hidden within other audio and image files. However, the third tool used is F5 algorithm; it offers a large steganographic capacity and uses straddling mechanism. It also provides matrix encoding to improve the efficiency of embedding.

Different types of these clean images are used with the three steganographic tools, including colour and grey images. The images file formats used are JPEG and BMP. However, in order to have a different collection of stego-images, we have hidden single and multiple files data to test our detection system later. The purpose of hiding multiple secret data is to test our detection system if it can detect all the multiple hidden files.

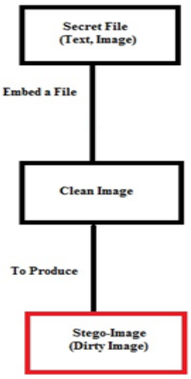

Fig. 1.The process of creating the stego-images.

There are many reasons of choosing three different steganographic tools. OpenStego does support many images formats, but the output stego-image will be in PNG format all the time. However, S-Tools support the BMP file formats only, and the F5 algorithm is working with the JPEG file format.

B. The Detectiom System

The proposed detection system is based on extracting image features from a clean image (cover image) and its stego version. The goal here is to distinguish between the clean image and the stego-image. Therefore, there is a need for comparison between the values of the image features resulted, as well as testing the statistical differences for the two types of images as shown in Fig. 2.

The aim of the proposed detection system is to detect the stego-images created in the previous stage. Some of the stego-images crated containing more than one hidden file. These images could also have different formats. In addition, the stego-images have been created using three different steganography tools using different methods. Therefore, the proposed detection system will be tested on different collection of stego-images.

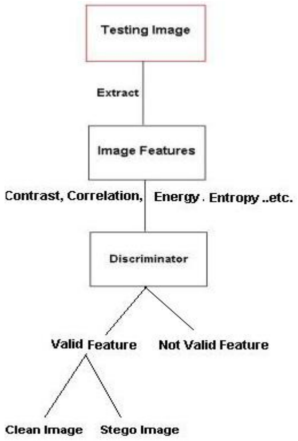

Fig. 2. The process of the proposed detection system.

The detection system is going to be trained on a set of images. Therefore, a co-occurrence matrix will be created for each image. This matrix is useful as it is a statistical approach that helps to provide valuable information about the relative position of the neighbouring pixels in an image. Then, number of image features will be extracted from the matrixes created. Different image features will be selected for extraction, such as: contrast, correlation, energy, colours variance and maximum difference [8].

V. Initial Results

This is a first stage for implementing the detection system. Initial results are done for set ofcolour and gray images. Four examples of the images used and of the image features extracted will be discussed in this section.

To start testing the proposed detection system, four image features are extracted from the co-occurrence matrix of severalgrey and colourclean images, and also extracted from their stego versions. These features are: contrast, correlation, energy and homogeneity. The steganography tool used for creating the stego-images in this initial test is the S-Tools. The image format chosen is BMP for all images. The hidden data file used is the same for the all stego-images created. Table I shows the images used, their sizes, their types and the size of the hidden data.

TABLE I: GREY IMAGES USED AND THIER SIZES IN THE INTIAL RESULTS

TABLE 1: Clean Image	Bit Depth	Size	Type	Stego-Image	Size	Hidden Data (Size)
Gray1	8	257 kb	Medium Grey Level	Stego1	257 kb	1 Image 109 kb
Gray2	8	757 kb	Medium Grey Level	Stego2	768 kb	1 Image 109 kb
Gray3	8	190 kb	Medium Grey Level	Stego3	190 kb	1 Image 109 kb
Gray4	8	226 kb	Dark Grey Level	Stego4	226 kb	1 Image 109 kb
Gray5	8	257 kb	Medium Grey Level	Stego5	257 kb	1 Image 109 kb

The following are the steps of extracting the image features using MATLAB for the grey images:

1) Read an image, using ‘imread( )’ function.
2) Create the co-occurrence matrix for an image, using ‘graycomatrix( )’ function.
3) Extract the ‘contrast’, correlation, energy and homogeneity features from the matrix created, using graycoprops function.

Similar steps are done for the colour images to extract the four features using MATLAB:

1) Read an image, using ‘imread( )’ function.
2) Found the RGB channels for the colour image. For example:

```
Red=clean_image1(:,:,1);
Green=clean_image1(:,:,2);
Blue=clean_image1 (:,:,3);
```

This means there are three matrixes for each image.

3) Create the co-occurrence matrix for the three RGB matrixes, usingthe ‘graycomatrix( )’ function.
4) Extract the ‘contrast’, correlation, energy and homogeneity features from the three matrixes created for each colour image, using graycoprops( ) function.

More details about the four image features as the following:

- **Contrast:** it is a measure of the contrast or the amount of local variations present in an image. It Returns a measure of the intensity contrast.
- **Energy**: Returns the sum of squared elements in the GLCM.
- **Homogeneity:** Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.
- **Correlation:** is a measure of gray-tone linear-dependencies in the image. It returns a measure of how correlated a pixel is to its neighbour over the whole image [8].

Fig. 4 shows some of the clean images used for the initial results. However, figure 6 shows the hidden data files used.

Fig. 4. Some of the clean images used with the initial.

Fig. 5. The hidden data files.

TABLE II: VALUES OF THE FOUR FEATURES EXTRACTED FROM CLEAN IMAGE1 AND STEGO1

Clean Image1 (Gray1)		Stego-Image1	
Contrast	0.1092	Contrast	**3.1177**
Correlation	0.9853	Correlation	**0.6407**
Energy	0.1220	Energy	**0.0429**
Homogeneity	0.9533	Homogeneity	**0.6553**

TABLE III: VALUES OF THE FOUR FEATURES EXTRACTED FROM THE RGB CHANNEL FOR CLEAN IMAGE2 AND STEGO2

Clean Image2 (Coulour2)		Stego-Image2	
Contrast Red Channel	0.1102	Contrast	**0.1114**
Correlation Red Channel	0.9565	Correlation	**0.9560**
Energy Red Channel	0.2653	Energy	**0.2647**
Homogeneity Red Channel	0.9453	Homogeneity	**0.9448**
Contrast Green Channel	0.0366	Contrast	**0.0397**
Correlation Green Channel	0.9848	Correlation	**0.9835**
Energy Green Channel	0.3760	Energy	**0.3741**
Homogeneity Green Channel	0.9817	Homogeneity	**0.9802**
Contrast Blue Channel	0.0899	Contrast	**0.0920**
Correlation Blue Channel	0.9937	Correlation	**0.9936**
Energy Blue Channel	0.3152	Energy	**0.3143**
Homogeneity Blue Channel	0.9558	Homogeneity	**0.9548**

Table II shows the values of the four image features extracted from the clean grey image1and its stego version. The features were extracted after creating the co-occurrence matrix from the clean1 and stego1.

The contrast for stego1 has been incredibly increased. However, other features have been decreased.

Due to the fact that the grey image always having 8 bit depth, thus, the contrast will increase. As a result, the four image features used with the clean1 and stego1 are different from each other.

On the other hand, Table III shows the values of the four image features extracted from the red channel of clean image2 and its stego version. The features were extracted after creating the co-occurrence matrix from the clean2 and stego2. As it can be seen in Table III, the four features have not changed for the red channel.

Large number of different grey and colour images will be tested. Also, more image features will be extracted, such as, variance and maximum error. In addition, the T-test will be implemented as discriminator on the features selected to test how valid they are.

VI. Discussion

Using grey images for steganography has many limitations. First, the capacity of hiding data is low, due to the fact that the image bit depth is always 8. Moreover, most of the grey images are BMP file format. In addition, in case of converting them to another format, they will convert to colour images. Otherwise, the resolution of these images will noticeably affect. However, in regards to the initial test, the co-occurrence function used in MATLAB supports the grey images only. This means this function has to be used with each single colour channel in the colour image. For example, extract the co-occurrence matrix for the red, green and blue channel; therefore, there will be three matrixes for each colour image.

The noticed differences in the image features values for a clean image and its stego version will be an indication to differentiate between them.

However, the proposed system will be evaluated according to its ability in regards to detect the two hidden data files within some stego-images.

VII. Future Work

As it had been mentioned earlier in this paper, more grey and colour images will be used to test the proposed detection system.

In addition, the T-test will be implemented as discriminator on the selected image features to test their validity.

All process that been done will be automate using MATLAB to generate large set ofstego-images and detection system to test and evaluate the system as well as to get the image features values.

VIII. Conclusion

This paper briefly reviewed the definition of steganography and Steganalysis as well. Thesteganalysis categories and some related work form the literature. The paper introduces the proposed system for detecting the hidden data. OpenStego, S-Tools and F5 algorithm have been used to create varieties of stego-images. These stego-images are used to test and evaluate the detection system proposed. The detection system works by creating the co-occurrence matrix for all images. Numbers of image features were extracted from the co-occurrence matrix of these images. The values of the image features for the clean and stego images are compared to distinguish between them. These features will be validated using the T-test discriminator using MATLAB in the next stage.

References

[1] N. Johnson and S. Jajodia, “Steganalysis: The investigation of hidden information,” *IEEE Information Technology Conference*, New York, USA, 1998.

[2] P. Thiyagarajan, G. Aghila, and V. Venkatesan, “Steganalysisusing colourmodel conversion,” S*ignal & Image Processing: An International Journal*, vol. 2, no. 4, 2011.

[3] Z. Duric, M. Jacob, and S. Jajodia, “Information hiding: Steganography and steganalysis,” *Elsevier Science*, 2004.

[4] G. Huayong, H. Mingshenge, and W. Qiana, “Steganography and steganalysisbased on digital image,” in *Proc. 4th International Congress on Image and Signal Processing*, Shngahai, China, 2011.

[5] H. Farid, “Detecting hidden messages using higher-order statistical models,” in *Proc. IEEE International Conference on Image processing*, New York, USA, 2002.

[6] R. Gong and H. Wang, “Steganalysis for GIF images based on colors-gradient co-occurrence matrix,” *Optics Communication*, vol. 285, 2012.

[7] V. Arvis, C. Debain *et al.*, “Generalization of the concurrencematrixforcolour images: Applicationto Colour Texture Classification,” *Image Anal Stereo*, 2004.

[8] I. Avcibas, N. Memon, and B. Sankur, “Steganaly sisusing image quality metrics,” *IEEE Transaction on Image Processing*, vol. 12, no. 2, 2003.

Ahd Aljarf received her B.Sc. degree in Computer Science from Umm Al-Qura University, Saudi Arabia in 2007. She received her MSc degree in Forensic Computing rom Coventry University, Coventry, UK in 2011. She is studying now towards her PhD degree at the Faculty of Engineering and computing in Coventry University.

Saad Amin is a principal lecturer and Computer System subject leader at the Faculty of Engineering and Computing in Coventry University, UK. His main area of research involves the image processing and multimedia applications on parallel computers. Dr Amin is involve in several research projects and has published many journal and conference papers.

John Filippas received his B.Sc. (Hons) in Computer Science from Coventry University. He received his MSc in Software Engineering and his Ph.D. in Biomedical Computing from the same university. His areas of interests are Parallel Computing, Computer Games and Biomedical Computing. He produced a number of journal and conference papers.

James Shuttleworth is an associate Head of Computing at Coventry University. His research is in the areas of image analysis, machine vision and pervasive computing. James has worked on a range of research projects and consultancies, producing a number of journal and conference paper.