Cover Selection Steganography Method Based on Similarity of Image Blocks

Pramneet Kaur

***Abstract*—** The primary function of Steganography is to hide information in a cover image such that the data is not detected easily. In a previously proposed method, a technique based on block texture similarity was introduced where blocks of cover image were replaced with the similar secret image blocks; then indices of secret image blocks were stored in cover image. In this method, the blocks of secret image are compared with blocks of a set of cover images and the image with most similar blocks to those of the secret image is selected as the best candidate to carry the secret image. Also work has been done to embed the information in the noisy region of the image. Using appropriate features for comparing image blocks, guaranties higher quality of stego images and consequently, allows for higher embedding capacity, less detect ability and, enhanced security. Based on this idea, in this paper, an adaptive cover selection steganography method is proposed, that uses statistical features of image blocks and their neighborhood. The method is examined with feature based and wavelet based steganalysis algorithms. The results prove the effectiveness and benefits of the proposed method

I. INTRODUCTION

Steganography is the art and science of writing hidden messages in such a way that no one, except the intended recipient, knows of the existence of the message. Consider that a transmitter consists of a host image H, and the message M that a sender hopes to communicate confidentially. The message can be text, images, or anything that can be represented by a bit stream. The host image H is used to embed the message by using a stego-encoder controlled by a key K. The key is a shared secret with the intended recipient whose knowledge of the key enables him to decode the message from the stego image. The decoding param- eters are known to both sender and receiver as a shared secret. The resulting stego image, SI = f (H,M,K), is transmitted over a channel to the receiver where it is processed by the stego-decoder using the same key K.

Successful steganography depends upon the carrier medium not to attract attention. When presence of stego-content is suspected, the main goal of steganography is defeated [1]. There is a tradeoff between the invisibility (imperceptible to naked eye) and the amount of information that can be hidden in a given cover image [2]. Steganographic security is mostly influenced by the type of cover media; the method for selection of places within the cover that might be modified; the type of embedding operation; and the number of embedding changes that is a quantity closely related to the length of the embedded data. Given two embedding schemes that share the first three attributes, the scheme that introduces fewer embedding changes will be less detectable. The rest of this section reviews some of the existing proposed steganography methods.

DCT domain embedding techniques are very popular due to the fact that JPEG which is a DCT- based image format is widely used in the public domain in addition to being the most common output format of digital cameras. Some of steganographic embedding in DCT domain are Outguess [3], F5 [4], model-based [5], perturbed quantization (PQ) [6], and Matrix embedding [7].

A cover selection method [9,10], like an image retrieval method, retrieves images based on their fitness to carry a given secret image. Cover selection problem was studied in [10] by investigating three scenarios in which the embedder has either no knowledge, partial knowledge, or full knowledge of the steganalysis technique. The main idea in [9] is based on dividing the secret image into blocks of size 4×4 where for each secret block, the most similar block in the host image is found and the secret block is placed there. The host image is found from an image database, in such a way that it has the most number of similar blocks to those of given secret image. To find the similarity between blocks, they used texture analysis measures based on Gabor filter. Then, the location addresses of the blocks in host image which are replaced by blocks of secret image are saved. Then, this data is converted to a bit string and coded by Hamming code.

This bit string is embedded in determined DCT coefficients of the modified host image and the blocks for embedding are selected using a key which is the seed of a random sequence generator. One of the advantages of this method is increasing the embedding capacity by hiding only blocks indices in DCT coefficients of host image.

One of the difficulties in texture based similarity measure presented in [9] is that, they compare only the content of two 4×4 blocks without considering the effect of pixels in close neighborhood to the blocks.

Therefore, by replacing similar blocks with each other, virtual edges may appear in borders and corners of a replaced block.

In this paper to remove the blocking effect and hence, improve the quality of stego images, we used the neighborhood information beside block texture information. Each block in the secret image is taken as a texture pattern for which the most similar block is found among the blocks of the host image. The embedding procedure is carried on by replacing these small blocks of the secret image with blocks in host image in such a way that least distortion is imposed. We have used block texture combined with neighborhood information to measure the similarity of blocks. Our experimental results showed a high level of capacity and minimum distortion on images. In this way, we present a method for image hiding that uses the concept of block similarity between host and secret images. We used texture information to obtain a mean for summarizing the information of each image. Also block neighborhood information is used to avoid generating virtual

edges and furthermore, K-means algorithm is applied to improve the speed of finding the best host image. The stego and restored secret images have high quality and we verified that, using recent powerful blind steganalyzers, one cannot discriminate between clean and stego images reliably.

In section 2, the proposed steganographic method for gray level images is discussed. Performance of the proposed technique is analyzed in Section 3 and finally, section 4 outlines some concluding remarks.

II. PROPOSED STEGANOGRAPHY METHOD

Given a secret image, the proposed steganography algorithm finds the best candidate host image from an image database. Selecting the best cover is based on the similarity of secret and cover image blocks. Like the method which was proposed in [9], the main idea is to find texture pattern blocks of the secret image, in the host image, and save their addresses in host image. Some statistical features of a block and its neighbor- hood information are used to measure similarity of image blocks. The structure for the proposed steganography algorithm is shown in Figure 1.

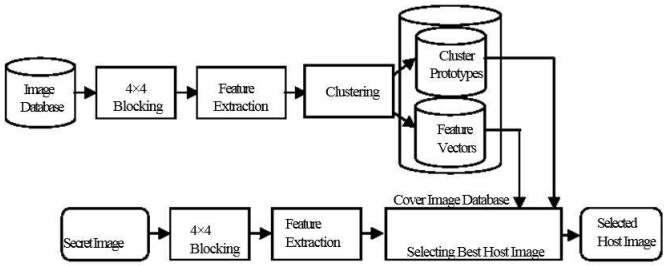

Fig.1: Structure for the proposed steganography algorithm

A. Feature Extraction

Image properties such as image texture could be used to categorize the images. The crude measures of image texture would be the mean, variance, and skewness of image blocks which are simple and can efficiently be computed. In order to compute the similarity between secret and cover images blocks, all the images are divided into blocks of size 4×4. For each block, the statistical values such as mean, variance, skewness of 2×2 sub-blocks and, the neighborhood information, as shown in Figure 2, are calculated.

Four 2×2 sub-blocks are considered in up-right, up- left, down-right and, down-left corners in a 4×4 block.

The average intensity value of four pixels that are adjacent to each side of a 4×4 block, are computed and considered as neighborhood information of that block side. In this way, a 16 dimensional feature vector is obtained (12 statistical values and 4 neighborhood mean values). By searching a host image from image database, the image which provides the best similarity to the secret image will be selected.

B. Finding the best host image

Due to the existence of large number of feature vectors for each image in the database an efficient method must be used for indexing and searching the blocks. Therefore, K-means is applied to each image of the database to cluster feature vectors. For selecting a block of host image, similar to a secret image block, the feature vector extracted from the secret image is compared to the cluster indices. The most similar block is found from the cluster whose cluster prototype is the nearest to the selected secret block feature vector.

For each block in secret image, its feature vector is calculated and compared to cluster prototypes in database. Each cluster prototype indicates the index of a group of similar blocks. Having determined the most similar cluster prototype, the most similar block is searched in that cluster and is selected as the best choice in this group. Euclidean distance is then applied to measure the closeness of image blocks in database to the blocks of secret image. This procedure is carried on for all blocks of secret image and finally, the image with the most number of similar blocks in database is chosen to be the host image.

C. Encide and Decode a secret image

After host image selection, we used the same approach as in [9] to embed the secret image to the selected host image that is as follows: Each block of secret image is replaced with the most similar block to it in host image. The positions of secret image blocks in host image are saved. In next stage, the sequence of mentioned block positions is changed to a bit string. Then a seed for generating a random sequence of location addresses (key) is considered. The position address bit string will be hided in middle DCT coefficients of host image.

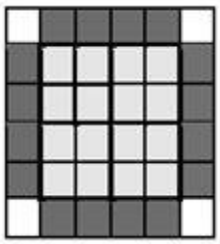

Fig.2: Regions for computing features

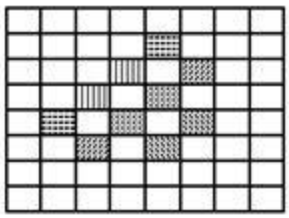

Fig.3: Middle DCT coefficients of an 8×8 block

The way of doing steganography in the DCT domain is to modulate the relative size of DCT coefficients. The algorithm is described in [11] as splitting the image into 8×8 blocks and calculating the DCT of each block. Then two middle-frequency coefficients are chosen and agreed upon by both send and receive parties. A block encodes a 1 if DCT(a,b) > DCT(c,d) and 0 otherwise.

In the encoding step, the coefficients are swapped if their relative size does not match with the bit to be encoded. Determined DCT coefficients are shown in Figure 3.

The sender sends the modified stego image SI to the recipient. Decoding is straightforward because the recipient first forms the random sequence by using the same key K and then, retrieves the embedded bit string from the DCT coefficients of the blocks. The extracted bit string is simply Hamming decoded and then indices of the secret image blocks are extracted. Through the knowledge of block indices, the secret image can be reconstructed.

D. Steganalysis

Each steganalyzer is composed of feature extraction and feature classification components. In this context, two techniques which are studied in this work, take distinct approaches in obtaining distinguishing statistics from images. One of these techniques is wavelet-based steganalysis (WBS) proposed by [12,13]. In feature extraction part of this method, statistics such as mean, variance, skewness, and kurtosis are calculated from each wavelet decomposition subband. Additionally, the same statistics are calculated for the error obtained from a linear predictor of coefficient magnitudes of each sub band. Feature-based steganalysis (FBS) [17] obtains a set of distinguishing features from the DCT and spatial domains. It is shown in [16] that FBS technique outperforms other techniques such as WBS and binary similarity measures.

III. AVERAGE CAPACITY AND TIME COMPLEXITY EVALUATION

Fridrich in [17] showed that the average steganographic capacity of grayscale JPEG images with quality factor 70 is approximately 0.05 bits per non- zero DCT coefficient. In the proposed method, each 64×64 size secret image has 256 blocks of size 4×4. For embedding this secret image in a host image, we should find 256 corresponding similar blocks. If each host image size is 256×256, it has 4096 blocks of size 4×4. Therefore, for addressing 4096 blocks, we need 12 bit addresses. After applying Hamming code algorithm, for each 4 bits, 7 bits, and totally, for addressing each block 21 bits and for the whole 256 blocks, 5376 bits are needed. In this way, a 64×64 secret image is embedded in a 256×256 host image and the embedding rate is 0.06 per bit. Experimental results are carried out on a 2046 MB PIV processor using MATLAB 7.1 and image processing toolbox 5.0.2. It should be noted that MATLAB codes are usually 9 or 10 times slower than their C/C++ equivalents [18]. Table 1 shows the results of the time evaluation of the proposed method. 4(d) show the host image in which the secret image is hidden and the restored secret image, respectively. From Figure 4(c), we can see that the quality of the stego image is high, and unintended observers will not be aware of the existence of a hidden image in it. Indeed, it is impossible to distinguish between Figure 4(b) and (c) or between Figure 4(a) and (d) using naked eye. This indicates that the value and normal usage of the secret image are preserved. The average of PSNR of the stego images is more than 39 dB which shows that the cover images have excellent imperceptibility after the secret image is embedded in them. Table 2 shows the effect of using only block texture information as suggested by [9] and using neighborhood information as suggested by our method. The results show that using neighborhood information increases the quality of stego and restored secret image.

IV. STEGANALYSIS

In this section, we evaluate the security of the proposed algorithm using two blind steganalyzers that use features constructed in wavelet domain (WBS), and features calculated in the DCT domain (FBS). In WBS, a Fisher Linear Discriminator (FLD) and in FBS a nonlinear Support Vector Machine (SVM) is trained to discriminate clean and stego images. One hundred images from database were chosen randomly for testing, while the remaining images were used for training. This partitioning was repeated a total of ten times, with different random subsets used for training and testing each time. For each of the ten partitions, the SVM/FLD was trained with the statistics from the training image subset. Finally, the trained classifier was tested against the previously unseen images. The average of detection accuracy is shown in Table 3.

As can be seen, the proposed method with payload of approximately 0.067 bits per cover image bit cannot be reliably detected by any of the two steganalyzers.

Fig.4: (a) Secret image. (b) Selected cover image. (c) PSNR Between the stego and cover image. (d) PSNR between the original and extracted secret image (Units are in dB).

Table 2: Comparison of stego and restored secret image PSNR with and without considering neighborhood information

Considering Neighborhood Information	Average Stego Image PSNR	Average Restored Secret Image PSNR
No	36	34.5
Yes	39.5	35

Table 3: Blind Steganalyzers detection accuracy

Average Secret Image Size	Average Cover Image Size	Steganalyzer	False Positives	True Positives	Detection Accuracy
4.3 KB	63.5 KB	WBS (FLD) FBS (Non-Linear SVM)	37.11 % 44.6%	53.77 % 37.2%	58.33% 46.3%

V. CONCLUSION

In this a data hiding scheme that is imperceptible while a big secret image is concealed in a cover image. The main idea is based on dividing the secret image into blocks and considering these blocks as units for embedding. Then, using the similarity measure, provided by feature vector, the most similar block in the host image is found and the entire secret block is replaced there.

Considering the neighborhood information of blocks, prevents the method presented in [9] to outbreak the virtual edges in sides and corners of replaced blocks. The main achievements of the proposed steganography method are: (i) reduction of the host image distortion, and (ii) increased security. In addition our method benefits from the advantages of increasing the embedding capacity by hiding only blocks indices (location addresses) in DCT coefficients of host image as suggested in [9].

The approach aims to reduce the risk of detection, while keeping a high embedding capacity. This gain is more important for long messages than for shorter ones because longer messages are, in general, easier to detect. We showed that cover selection method provides good embedding efficiency and its relative embedding capacity densely covers the range of large payloads, making it suitable for practical applications. The experimental results show that applying the steganalysis methods on stego images can not reliably detect stego and clean images.

VI. REFERENCES

[1] N. Johnson, S. Jajodia, "Steganalysis of images created using current steganography software", in Proc. Int. Workshop on Info. Hiding, Germany, 1998, pp. 273-289.

[2] G. Greg, "Steganalysis Gets Past the Hype", IEEE Distributed Sys. Online, April, 2005, vol. 6, no. 4, pp. 1-5.

[3] N. Provos, "Defending against statistical steganalysis", in Proc. 10th USENIX Security Symp., 2001.

[4] A. Westfeld, "F5-a steganographic algorithm: high capacity despite better steganalysis", in Proc. 4th Int. Workshop on Info. Hiding, 2001.

[5] P. Sallee, "Model-based steganography", in Proc. Int. Workshop on Digital Watermarking, Seoul, Korea, 2003.

[6] J. Fridrich, M. Goljan, D. Soukal, "Perturbed quantization steganography with wet paper codes", in Proc. ACM Multimedia Workshop, Germany, 2004.

[7] J. Fridrich, D. Soukal, "Matrix Embedding for Large Payloads", IEEE Trans. on Info. Forensics and Security, vol. 1, No. 3, 2006.

[8] P. Su, C. Kuo, "Steganography in JPEG 2000 compressed images", IEEE Trans. Consum. Electron. 2003, Vol. 49 Issue 54, 824-832.

[9] Z. Kermani, M. Jamzad, "A Robust Steganography Algorithm Based On Texture Similarity Using Gabor Filter", IEEE Int. Symp. on signal processing and Info. Technology, 2005, pp. 578-582.

[10] M. Kharrazi, H. Sencar, N. Memon, "Cover Selection for Steganograpic Embedding", in Proc. ICIP, 2006, pp.117- 121.

[11] S. Katzenbeisser, F. Petitcolas, "Information Hiding techniques for steganography and digital watermarking", 2000, ISBN 1-58053-035-4.

[12] S. Lyu, H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines", in Proc. 5th Int. Workshop on Info. Hiding, 2002.

[13] S. Lyu, H. Farid, "Steganalysis using color wavelet statistics and one-class support vector machines", in Proc. SPIE 5306, 2004, pp. 35-45.

[14] J. Fridrich, "Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes", in Proc. 6th Info. Hiding Workshop, Toronto, 2004.

[15] M. Kharrazi, H. Sencar, N. Memon, "Performance study of common image steganography and steganalysis techniques", Journal of Electronic Imaging 15(4), 041104, 2006.

[16] http://www.cs.washington.edu/research/imagedatabase/groundtruth/