Image cover selection for steganography based on run length matrix

Zahra Jalili[1], Hedieh Sajedi[2], Maryam Hasanzadeh[3]

Receive :2016/04/20 Accepted: 2016/07/26

Abstract

In this paper, we proposed a data-hiding scheme based on Run length matrix. In a previously proposed method, a technique based on texture classification was introduced where four statically features extracted from run length matrix; then best cover images are selected based on these features. Using appropriate features for comparing images from undetectability viewpoint, guarantees, less detectability of stego images and consequently, enhances security of the steganography algorithms. Based on this idea, in this paper, more features are extracted from run length matrix to select the best covers. Our method is examined with feature based and wavelet based steganalysis algorithms. The results illustrate the effectiveness and benefits of the proposed method.

Keywords: Steganography; Steganalysis; Cover Selection; Run Length Matrix.

1 INTRODUCTION

Steganography is the science of hiding a secret information. The goal of steganography is to embed secret messages in such way that no one except the intended recipients can detect presence of secret messages. It can embed secret datain a digital media such as text, image, audio, video, and multimedia. Conversely, steganalysis attempts to expose the existence of hidden data[1]. Image with JPEG form is one of the most popular image steganography. The steganography algorithms for JPEG images can be divided into non-adaptive and adaptive steganography[2].Non-adaptive algorithms do not consider the image content characteristics of cover objects while adaptivealgorithmsembede secret data in complicated image texture and edge regions[2]. Content of cover images that is used for steganographyhave significant impact in efficiencyofsteganalzerss. Textured images havea lot of redundancy data, that is invisible to the human visual system. Thereforeredundancy of the data helps to hide the presence of a secret message and helps modeling of complicated images be difficult[3].

An ideal steganography system should be secure enough against various ateganalyser systems. So steganographer should embed secret data into a cover media in such a way that no severe visual artifacts and are not significantly distrube statistical features of the cover media [4]. The type of cover media, the number of chenges, the places that used for embededing and the type of embedding method are Factors that affect the security of steganographic algorithm [5, 6].In order to having secure communication, most currently steganographic methods focus on designing the data embedding algorithms, but cover medias are selected randomly without considering any information about their suitability as carrier.

Cover selection steganography is a new approach that considred as a way to improvesecurity of steganography algorithm over the recent years. The cover object in steganography acts as a carierof the secret data. Also, the embedder is free to choose any cover object from a database using a cover selection approach.A cover selection approach, can suggest suitable covers based on some criteria. Therefore ranked objects can help the embedder to decide whether to transmit the stego object or to select an alternative one[5]. Therefore, the embedder can minimize the detectability of stego objects by selectingbest cover for hiding the secret data.

In this paper,run length matrixis used to extract the textured characteristics of cover images. combination of these extracted fearures are used to elicit suitable images from a database as proper covers.Then, in order to generate the stego version of a selected images, the secret message is embeded into the selected image. Analysis of results show that the proposed cover selection increases the security of stego images against the steganalyzers.

[1] Department of Computer Science, College of Engineering, University of Shahed Tehran, Iran. Email:z.jalili@shahed.ac.ir
[2] Department of Computer Science, College of Science, University of Tehran, Tehran, Iran. Email: hhsajedi@ut.ac.ir
[3] *Department of Computer Science, College of Engineering, University of Shahed,* m.hasanzadeh@shahed.ac.ir

The remainder of the paper is organized as follows. Section 2 describes gray level run length matrix. Our proposed method is introduced in Section 3. Performance of the proposed technique is analyzed in Section 4. Finally, the conclusion is described in Section 5.

2 Related Work

The cover selection problem was studied in [7] by investigating three scenarios in which the embedder has either no knowledge, partial knowledge, or complete knowledge of the steganalysis method. In addition, [7] introduced some measures for cover selection issue and divided these measures into two groups as cover based and cover-stego based measures.

Another cover selection technique for hiding a secretimage in a cover image is introduced in [8].Thismethod operates based on image texture similarity. For each cover image in database, the method replaces some blocks ofthe cover image with similar secretimage blocks; then, coverimage with most similar blocks is selected as the best candidate to carry the secret image.An improvement on this method is proposed in [6]thatuses statistical features of image blocks and their neighborhoods to select best blocks forreplacing.Use of the blockneighborhood information preventsappearance of virtual edges in the sides and corners.

The research in[4]presented a cover selection steganography method based on capacity as a property of images. An ensemble system that uses different steganalyzer units, capacity of cover image. So for embedding a secret data, the embedder can select the best cover image(s) with considering the capacityproperty of images in the database [4, 9]. Moreover, Ref. [4] analyzed the relation between the complexity and embedding capacity of images. The results show that middle and high complex images have higher embedding capacity [4]. An improvement on this method is proposed in [1]thatuses a preprocessing stage before calculating embedding capacity to increasing details of images. Images with high details have higerembededing capacity. This improvement is due to the properties of contrast enhancement and histogram equalization methods, Successive Mean Quantization Transform (SMQT) enhancement, brightening and darkening, blurring and sharpening [1].

The problem of spreading secret data to embed into multiple cover images is called batch steganography.The work in[10] presents the Adaptive Batch Steganography (ABS)manner and similar to [4] calculate capacity of image to select proper covers. ABS is an approach, which adaptively spread secret data among multiple cover images based on their embedding capacity. To reduce the estimation time in ABS, Rule-based Adaptive Batch Steganography (RABS) is proposed by[11]. In RABS, capacity of cover images is estimated using ‘Signature of Clean Images’ by evolutionary algorithm.

The work of[5]introduced some measures for cover selection steganography issue, and impact of these measurs on visual quality and security of stego images is investigated . The work in[5] divided these measuresinto two categories, fast and exact measures. In addition, a combination of both fast and exact methods can be used for cover selection.

In order to have secure communication in the presence of steganalysis Ref. [12] presented cover selection based on contrast measurement. The image withhighest contrast is selected for data embedding[12].

The main idea in [13] is based on image texture features and human visual system. It firstly calculates run length matrix for each cover image, then extracts four statically features such as Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-uniformity (GLN), and Run Length Non-uniformity (RLN). Second cover images with maximum SRE, GLN, RLN, and minimum LRE are selected.

More features in [13] are only functions of the total number of runs of length j, without considering the gray level information.These features alonewould not be able to detect the variation in gray levels[14]. “In addition, GLN feature is defined in terms of $g(i)$, it is the sum of $g(i)$ that determines its magnitude. In essence, this means that GLN measures the power of the distribution but cannot detect the possible variation in the shape of a distribution of given power” [15]. Thus, to distinguish exactly the textures, it is apparent that one must use not only the number of runs but also the gray values associated with them.

2.1 GRAY LEVEL RUN LENGTH MATRIX (GLRLM)

In this paper, we take a different look at cover selection issue from the texture classification point of view.Textureis the term used to characterize the surface of a given object or region and it is one of the main features utilized suitable for steganography, consequently the steganalysis on textured images is challenging, and recent steganalysis method are focused on textured images[16, 17]. Therefore, it becomes clear that the technologies developed for textured images classification should be able to play an important role in cover selection method. Gray Level Run Length Matrix (GLRLM) is a tools to classification of textured image. The basic idea of run length statistics is to extract information of an image from its gray level runs. A gray level run lenght is the number of adjacent pixels that having the same gray value in expected diraction. For a given image, we can compute a gray level run length matrix to four direction[18].The example in Fig. 1 shows a 4 x 4 image having four gray levels (0-3) and the resulting gray level run length matrices for the $0°$ direction.

0	1	2	3
0	2	3	3
2	1	1	1
3	0	3	0

(a)

Gray Level	**0°** 1	2	3	4
0	4	0	0	0
1	1	0	1	0
2	3	0	0	0
3	3	1	0	0

(b)

Figure 1.Sample of (a)apicture and (b) its Run length matrix.

The dimension of each GLRL matrix is $N_g \times N_r$array, where N_r and N_g are the largest possible run length and highest possible gray level value in the image, respectively. The matrix element (i,j)defiened as the number of times that gray level i in the image is repeated with runlength$j$, in the given direction. Many numerical texture features can be computed based on run length matrix. These features are summarized in Table 1. The first five features are only functions of $P_r(j)$, without considering the gray level $P_g(i)$, the next two features extract gray level information from the matrix. Other four features extract joint statistical measure of gray level and run length

in image processing and pattern recognition. It has been reported that when the data is hidden inside the textured images it is difficult to be detected, in other words, the textured images are

Table1- Statistically features extracted from run length matrix [15, 18,19].

No.	Features Extracted From GLRLM	Definition
1	Short run emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{j^2}$
2	Long run emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}j^2P(i,j)$
3	Gray level non-uniformity	$\frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N}P(i,j)\right)^2$
4	Run length non-uniformity	$\frac{1}{n_r}\sum_{j=1}^{N}\left(\sum_{i=1}^{M}P(i,j)\right)^2$
5	Run percentage	$\frac{n_r}{n_p}$
6	Low-gray-level-run-emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{i^2}$
7	High-gray-level-run-emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}i^2P(i,j)$
8	Short Run Low Gray-Level Emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P(i,j)}{i^2j^2}$
9	Short Run High Gray-Level Emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{i^2P(i,j)}{j^2}$
10	Long Run Low Gray-Level Emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}i^2j^2P(i,j)$
11	LongRun High Gray-Level Emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{j^2P(i,j)}{i^2}$

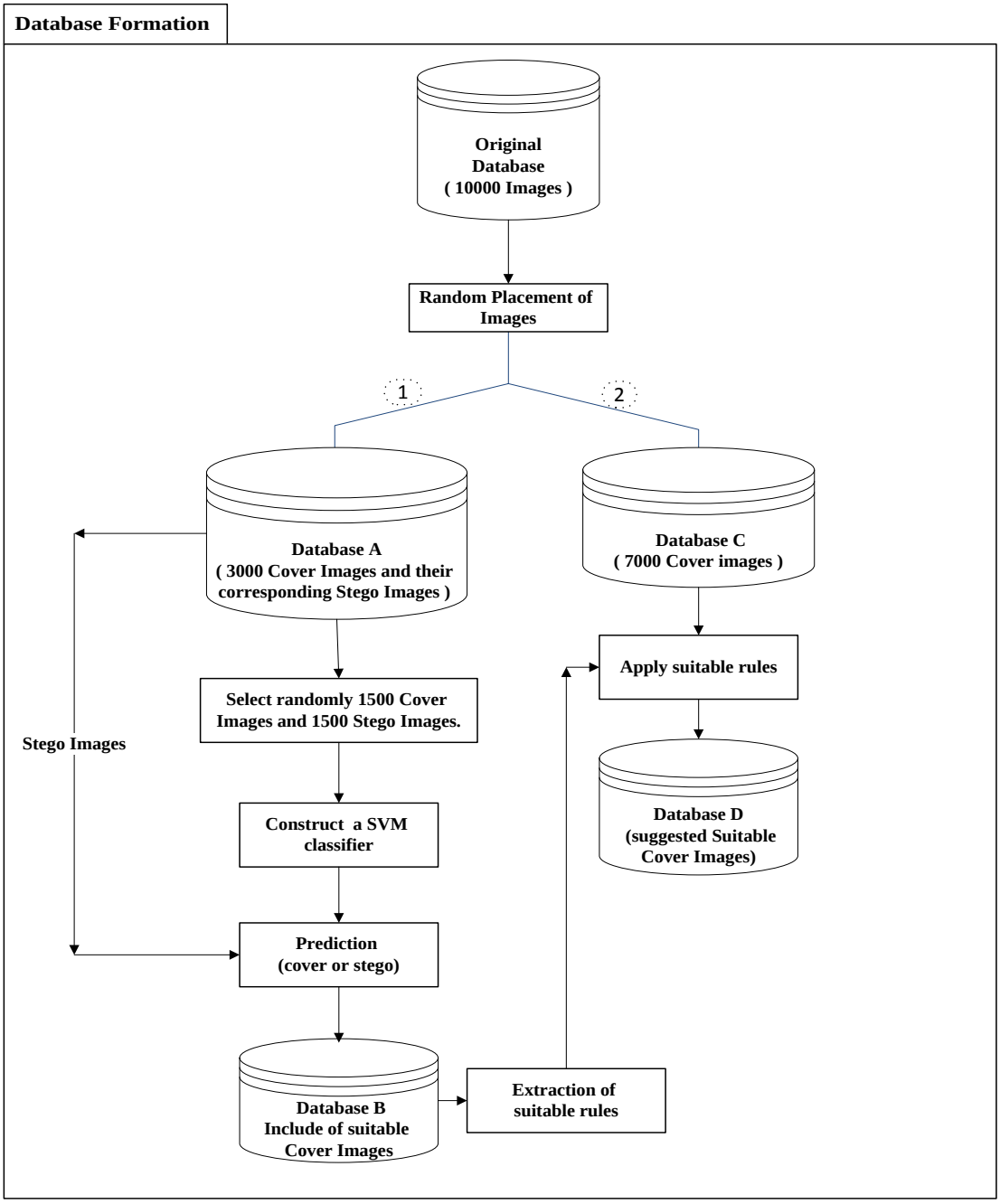


Figure 2. Block diagram of generating databases in our work.

3 THE PROPOSED ALGORITHM

In cover selection based steganography, to have a secure covert communication, one can select a cover image with low delectability to carry a secret.Detection accuracy of hidden secret message is low in textural and complex region of images and new steganalysis methods are using textural features of images to increase their accuracy. Therefore these discussionsledustoincrease security of hidden secret message using textural featuresto select best covers.

Proposed cover selection based steganography includes of some features that extracted from GLRLMleadus to select the best cover to carrythe secret message. Our proposed algorithm works on four different databases:A, B, C and D as shown in Figure 2. The block diagram in Figure 2 shows how these databases are generatedAt the first, we randomly permuted all of cover images in the main database, and then divided it into two parts A and C. A includes of 3000 images and C includes of 7000 images. DatabasesB and D arecreatedinthemiddlephase of the proposed algorithm as explained in the following of the paper. Database D is include of rules that are used to detect signature of stego images that are detected wrongly as cover images. These rules are applied on database C to determine best covers. Selected cover images construct database D.

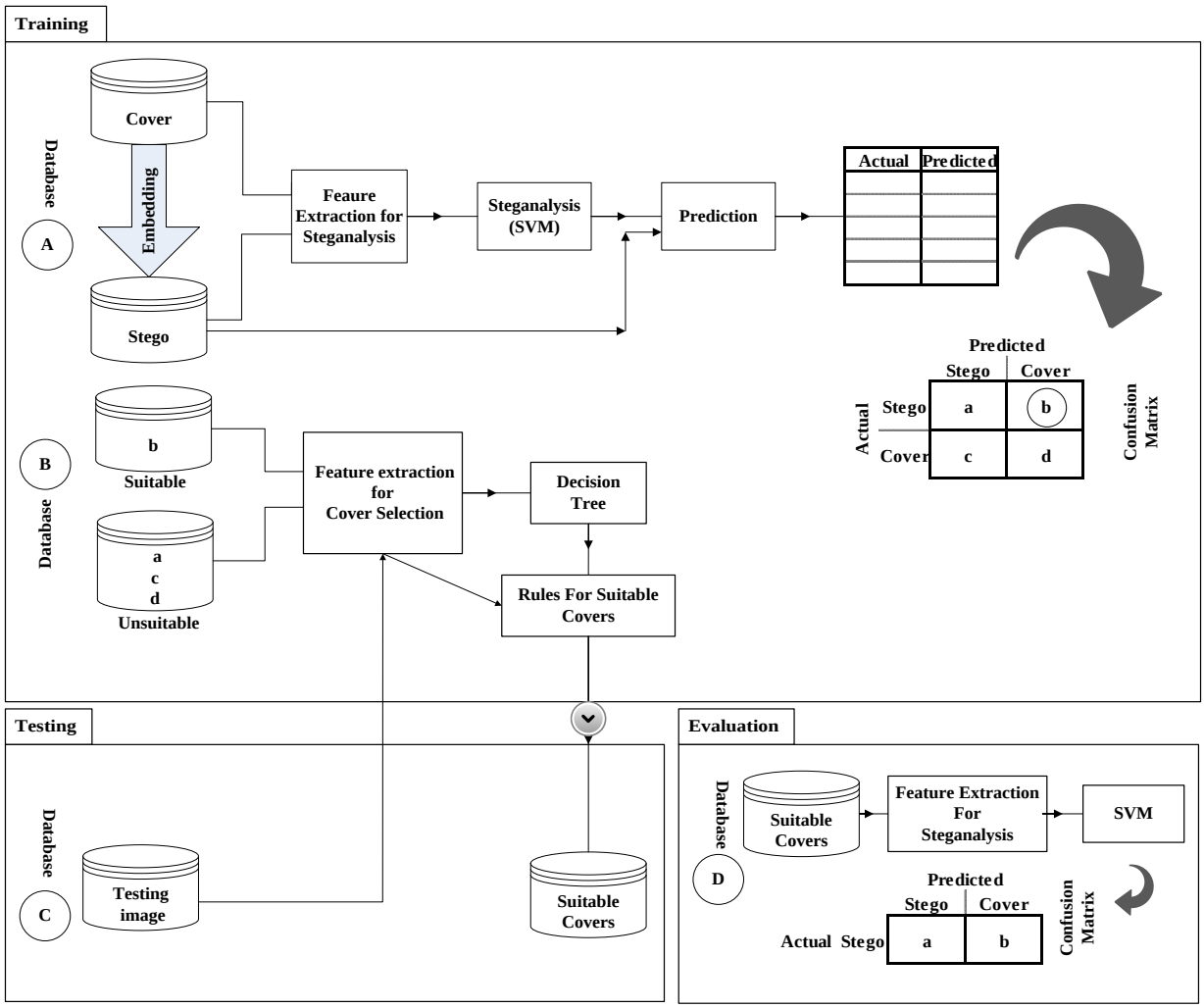


Figure 3. Block diagram of the proposed algorithm for cover selection.

The block diagram of proposed algorithm is shown in Figure 3. This algorithm has three phases as bellow:

1. Training phase: First,a random secret messageembedded into cover images of database A by a steganography method.Next, some features for each cover and stego images from database are extractedusing a blind steganalyzer.A random set steganalysis features of dataset A contains cover and stego images, whichare used to training a Support Vector Machine (SVM). Description of steganalysis features that used in proposed algorithm are as follow:

Block-based steganalysis in [20] presented an effective markov process based JPEG steganalysis scheme, which utilizes both the intra-block and inter-block correlations among JPEG coefficients. It computes relation between JPEG 2-D array to utilize the intra-block correlation, and inter-block correlations arerelation between coefficients located in the same position withinthe 8×8 blocks with ion probability matrices for those difference mode 2-D arrays to utilize the [21]. All the elements of these matrices are enhanced by Cartesian calibration [22] and used as features for steganalysis. Block-based method is a 972 dimentionsfiature steganalyzer.

– DCTR-based steganalysis in [23]proposed a feature set for JPEG steganalysis with low complexity and relatively small dimension. Its dimention is 8000. The features are built as histograms of residuals obtained using the basis patterns used in the DCT. To computing the features first 64 convolutions of the decompressed JPEG image with 64 8×8 kernelsshould be done and then forming histograms of these convolutions are required. The proposed features are called DCTR features (Discrete Cosine Transform Residual).

Stego images that misclassificated as cover images are suitablr for transfering, sotheisstageall

stego images of database Afed into SVM to predicting their class.Cover version of stego imagesthatpredicted as cover image are labeled as good covers,others,arelabeled as bad covers.Inthe next stage, some statistically features that mentioned in Table 1 are extracted from each cover images in database B. At this stage, texture feature of each image and it'slableare provided. Therefore we need a way to extract information to recommended appropriate cover imagesSo at the end of this phase some $(if \dots then \dots)$ rules will be extract by usingadecision tree such as J48(C4.5)[24]algorithmtorecognitionsuitablecovers. In fact, decision tree acts as a classifier based on rules.Desision tree give us information with high Interpretability. Decision tree is like a white box and resulted knowledge are suitable to extract good information, but a clssifiers such as SVM acts as a block box. Blox box algorithm don't present any information and are used just to classify the label and class of objects.The rules are composed from two parts, left hand side and right hand side. Left hand side and right hand side is called the conditions part and the result part respectively.Extracted rules are as follows:

$$Rule\ R_j: If(x_1\ is\ y_1\ and \dots\ x_n\ is\ y_n)\ then\ Image\ is\ a\ good\ cover.$$

2. Testing Phase: in this phase, some statically features that mentioned in table 1 are calculated for all cover images in database C, then extracted rules in the priviousephaseareappliedto these images. By using these rules best covers selected to carry the secret message. We call these selected cover images as database D.

3. Evaluation Phase: in this phase, first the secret messageembeded into the selected cover images in the previous phase (database D); then steganalysis features are extracted for all createdstego images. At the end,the SVM that created in the training phase is employed to predict the class of thestegoimages.

1 Experimental Results

To evaluate the proposed algorithm,someexperimentshave been done. Comparison experiments were conducted with 10000 grayscaleimages from BOWS2 database. All images were in size of 512×512 and PGM format. All of them are converted to JPEG format.Experimentswere performed on a PC with core 2.70 GHZ processor and 2GB main memory.

To make stegodatasets, random binary secret data is embedded into images using PQT and YASS steganography methods with payloads of 1024 and 2048 bits for each steganography algorithm. The quality of the input images and output images in PQT steganography is 85 and 70 respectively, and in YASS method both of them are 99.

Intheevaluationphase, detection accuracy (True Positive detection) of Block–Based and DCT-Basedsteganalyzerscalculatedagainst each used embedding technique.The results are shown in Table 2for PQT and YASS steganography methods.The columns of the table represent the result of classification using Block–Based and DCT-Based steganalysis method. Each row demonstrates the performance of different steganalyzers on a specific method with the determined payload.The results obviously show that the stego images, which are produced by the proposed approach, are less detectable than the stego images constructed by the classical use of steganography methods. Therefore, the proposed cover selection schemeimproved the robustness of PQT and YASS steganography against steganalysis attacks.In addition, it shows that PQT steganography is a powerful method and in low payload is undetectable by steganalyzers.Also, the results show that the Block-Based steganalysis is Stronger than DCT-Based steganalysis.

Table2- True Positive detection accuracy (%) of Block-Based and DCTR-Based steganalyzers on PQT and YASS steganography methods.

Steganalysis detection accuracy (%)					
steganography methods	Average payload(bits)	Classical steganography method		Proposed approach using steganography method	
		Block-Based	DCTR-Based	Block-Based	DCTR-Based
PQ	1024	49.96%	4.34%	35.85%	2.43%
	2048	50.88%	20.40%	39.37%	12.81%
YASS	1024	70.72%	57.86%	16.22%	10.34%
	2048	67.20%	73.62%	17.16%	14.92%

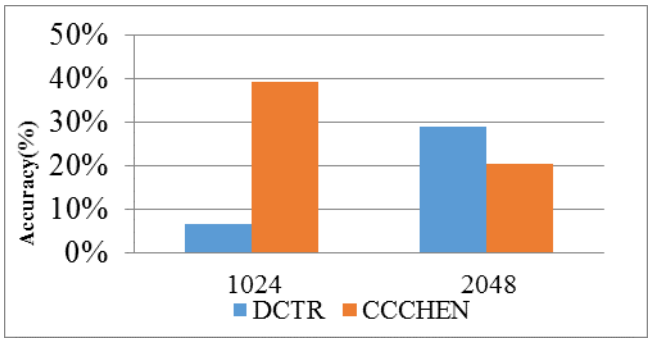
If inthetraining phaseinasted of extract rules from decision tree, it used as a classifierto recommending proper stego images.Therefore at the end of training phase we have a decision tree. In the testing phase decision tree applyied to stego images. Those stego images that classified as cover images create database D. At the end, SVM used to evaluation performance of selected images (database D).Performance of this method showed in table 3. As shown, security of selected images is inhanced than costum steganography, but it's security is lower than when embedder used rules to select proper images. As showed use of rules increase security of stego images, so at the following, experiments will continue for the way that rules have been used.

Table 3- True Positive detection accuracy (%) of steganalyzers in front of PQT and YASS steganography methods, if decision tree used as a classifier instade of extract rulre form it.

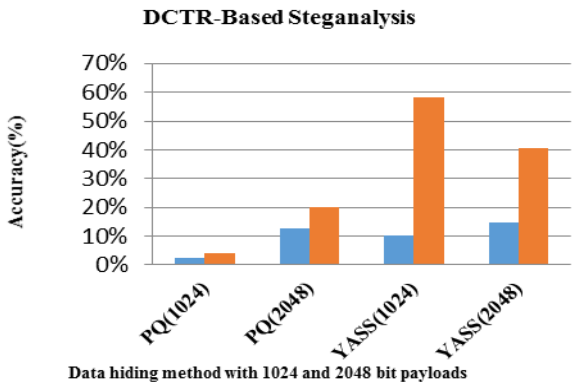
Steganalysis detection accuracy (%)			
steganography methods	Average payload(bits)	Proposed approach (use decision tree as classifier)	
		Block-Based	DCTR-Based
PQ	1024	42.29%	3.44%
	2048	44.07%	13.67%
YASS	1024	24.5%	17.92%
	2048	25.19%	22.06%

We exemined performance of proposed methodagainst another SVM.for example, if rules are extracted by using Block-Based steganalyzer, evaluation will done with DCT-Based steganalyzer and vice versa. Because PQT is a powefull method, we do this only on YASS steganograpgymethod.Figure 4 shows security of recommended cover images in front of different SVM. As showed in this figure, our proposed algorithm is a universal method, because detection accuracy of steganalyzers is lower than random guessing (50%)

.



At the end, figure 4compares the results of the proposed approach with the previous GLRLM approach against Block-Based and DCT-Based steganalysis methods. As the figure shows, our method works better than previous GLRLM methods. Our propsedalgorithm use all of features that extracted from GLRL matrix, but previous work use only four of these features. Therefore the result show by using rule-based algorithm, suggested cover images are better than previous GLRLM-based algorithm.



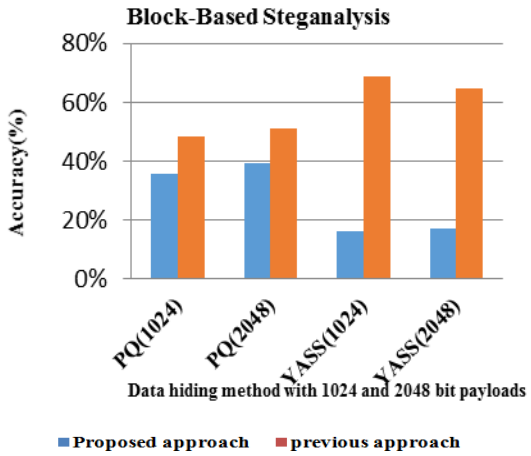


Figure 4-Comparison of the proposed approach with previous GLRLM approach against some steganography methods.

1. 5 Conclusion

In this paper, we defined rule-based steganography as a measure for cover selection. The main idea is based on extracted features from GLRLM.By using appropriatemeasures and extraction features from GLRL matrix, we can select the best covers exactly. Therefore, for embedding a secret data, the embedder can select the best cover image(s) regarding to signature of properimages in the database. PQT steganography is a powefull embedding method, and it's undetectable by exiting steganalysis methods. So for high payloads we can use this algorithm to select suitable images from database.Also, rules that extracted by using Block-Based features are more accurate than DCTR-Based, however DCTR-Based dimention is higher than Block-Based steganograpgy.Therefore, by using a blind steganalizer with high detection our proposed method enhances the security of the steganography algorithms considerably.

For feature work we want to use ensambleclassifier in training phase. Because each one of steganalysers use a different features to detect stego images and Each also have weaknesses, which use they with together cover these weaknesses. Therefore, by using ensamble classifier we can boost our proposed prformance.

2. References

[١]H. Sajedi and M. Jamzad, "BSS: Boosted steganography scheme with cover image preprocessing," *Elsevier*, vol. 37, p. 8, 2010.

[٢]X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, "Steganalysis of Adaptive JPEG Steganography Using 2D Gabor Filters," presented at the Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security, Portland, Oregon, USA, 2015.

[٣]K. Bailey and K. Curran, "An evaluation of image based steganography methods," *Multimedia Tools and Applications*, vol. 30, pp.٢٠٠٦ ,٨٨-٥٥ .

[٤]H. Sajedi and M. Jamzad, "Secure steganography based on embedding capacity," *International Journal of Information Security*, p. 13, 2009.

[٥]H. Sajedi and M. Jamzad, "Using contourlet transform and cover selection for secure steganography ",*Int. J. Inf. Secur.*, vol. 9, pp. 337-352, 2010.

[٦]H. Sajedi and M. Jamzad, "Cover Selection Steganography Method Based on Similarity of Image Blocks," presented at the Computer and Information Technology Workshops, 2008. CIT Workshops 2008. IEEE 8thInternational Conference on Sydney, QLD, 2008.

[٧]M. Kharrazi, H. T. Sencar, and N. Memon, "Cover Selection for Steganographic Embedding," presented at the Image Processing, 2006 IEEE International Conference on, Atlanta, GA, 2006.

[٨]Z. Z. Kermani andM. Jamzad, "A robust steganography algorithm based on texture similarity using Gabor filter," presented at the Signal Processing and Information Technology, Athens, 2005.

[٩]H. Sajedi and M. Jamzad, "Secure Cover Selection Steganography," in *Advances inInformation Security and Assurance*, ed Seoul, Korea: IEEE, 2009, pp. 317-326.

[١٠] H. Sajedi and M. Jamzad, "Adaptive batch steganography considering image embedding capacity," *Optical Engineering*, vol. 48, p. 10, 2009.

[١١] H. Sajedi, "RABS: Rule-Based Adaptive Batch Steganography," in *Recent Advances in Steganography*, H. Sajedi, Ed., ed: Computer and Information Science-Communications and Security, 2012, p. 18.

[١٢] M. S. Subhedar and V. H. Mankar, "Performance Evaluation of Image Steganography Based onCover Selection and Contourlet Transform," presented at the International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, 2013.

[١٣] S. Nazari and M.-S. Moin, "Cover Selection Steganography Via Run Length Matrix and Human Visual System," *Journal of Information Systems and Telecommunication*, vol. 1, p. 8, 2013.

[١٤] X. Tang, "Texture information in run-length matrices," *Image Processing, IEEE Transactions on*, vol. 7, p. 8, 2002.

[١٥] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recogn. Lett.*, vol. 11, pp. 415-420, 1990.

[١٦] R. Wang, M. Xu, X. Ping, and T. Zhang, "Steganalysis of JPEG images by block texture based segmentation," *Multimedia Tools and Applications*, vol. 74, pp. 5725-5746, 2015.

[١٧] Y. Q. Shi, P. Sutthiwan, and L. Chen, "Textural features for steganalysis," presented at the Proceedings of the 14th international conference on Information Hiding, Berkeley, CA, 2013.

[١٨] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing* vol. 4, p. 8, 1975.

[١٩] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level-run length distributions," *Pattern Recogn. Lett.*, vol. 12, pp. 497-5.١٩٩١ ,٠٢

[٢٠] J. Kodovsky and J. Fridrich. (2011). *CC-CHEN972 Steganalisys*. Available: http://dde.binghamton.edu/download/feature_extractors/download/ccchen972.m

[٢١] C. Chen and Y. Q. Shi, "JPEG image steganalysis utilizing both intrablock and interblock correlations," presented at the International Symposium on Circuits and Systems, Seattle, WA 2008.

[٢٢] J. Kodovsk and J. Fridrich, "Calibration revisited," presented at the Proceedings of the 11th ACM workshop on Multimedia and security, Princeton, NewJersey, USA, 2009.

[٢٣] V. Holub and J. Fridrich, "Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT," *Information Forensics and Security, IEEE Transactions on*, vol. 10, p. 11, 2014.

[٢٤] J. R. Quinlan, *C4.5: programs for machine learning*: Morgan Kaufmann Publishers Inc., 1993.