Journal of Computing and Security

January 2014, Volume 1, Number 1 (pp. 39–46)

http://www.jcomsec.org

The Effect of Correlogram Properties on Blind Steganalysis in JPEG Images

Dariush Alimoradi [a]
Maryam Hasanzadeh [a,*]

[a] *Computer Engineering Department, Shahed University, Tehran, Iran.*

ARTICLE INFO.

Article history:
Received: 15 December 2012
Revised: 23 June 2013
Accepted: 17 July 2013
Published Online: 20 December 2013

Keywords:
Correlogram Properties, Blind Steganalysis, Cover Selection, JPEG Image.

ABSTRACT

Blind image steganalysis is a technique for discovering the message hidden in images in an independent manner than embedding the hidden message. The content of the image contribute to the success of steganalysis drastically. In the past, texture, one of the most basic features in any image processing, has been used for content-based image classification. Correlogram properties as textural based features have numerous applications in this field. Homogeneity, contrast, correlation, energy and entropy are the correlogram properties used more frequently than others for this purpose. In this article, the impacts of these properties, as descriptors for image content, on blind steganalysis in JPEG image are investigated. The results indicate that when correlogram homogeneity increases, the false image detection of blind steganalysis increases accordingly; while, decrease in correlogram contrast and entropy, leads to an increase in error. The energy and correlation of correlogram have unspecified effects on image blind steganalysis.

1 Introcution

Steganalysis is a method for discovering the existence of hidden message in a carrier signal. The hidden message can be embedded into a carrier signal using different method referred to as steganography. If steganalysis method is not used the properties of specific steganography method referred to a blind or universal steganalysis.

The message embedding rate of many steganography methods in JPEG domain depends on the number of non-zero coefficient of Discrete Cosines Transform(DCT). Also,The length of the random massage used in steganography is an image capacity ratio[1–5]. These indicate implicitly that image content influences both the steganography and steganalysis performance.

There is an important question in this field. Which kind of images is more difficult for steganalysis? Two sets of different measures are provided to help select the suitable image. The measures that are stego-cover based and the measures that are cover-based [6]. The first set selects an image which is less affected by message embedding; therefore, image selection is affected by message selection and as the message changes the selected image will change as well. But, the second set is defined on the basis of image content only and independent of the message. Gray level Co-occurrence matrix(GLCM) properties of image is an example of second set of measures, used for image selection from a set of images [7]. Co-occurrence matrix depends on distance and direction which has less impact on steganalysis; therefore, correlogram, generalization for

[*] Corresponding author.

Email addresses: d.alimoradi@shahed.ac.ir (D. Alimoradi), hasanzadeh@shahed.ac.ir (M. Hasanzadeh)

ISSN: 2322-4460 © 2014 JComSec. All rights reserved.

Co-occurrence matrix which depends only on distance, can help discover suitable images for steganography. The motivation behind this selection is its property usage for content based image classification[8] and its similarity to GLCM used in steganalysis[9].

In this article, the effect of correlogram properties on image blind steganalysis is investigated. The set of the article is as follows: the related works in section II, the correlogram and its properties in section III, the experiment tools and methods in section IV, the results in section V and conclusion in section VI are explained.

2 Related works

In this section, the objectives and examples for each of the related topics and their difference with this article are described.

2.1 Steganography

Researchers have proposed various steganography methods to make the least changes on grayscale JPEG image and resist against current steganalysis methods.

“Sallee” provided a model based steganography and named it MB1[3]. It was recognized by a simple blocking measure. For this reason, he improved this method to resist against attack by the blocking measure and named it MB2[4].

Heuristic methods are another category of steganography techniques. These techniques apply the method provided by Jsteg Algorithm. The following generation of heuristic steganography methods includes F3, F4 and F5. ”Fridrich et al.” improved F5 to help increase the message embedding capacity and named it nsF5[2].

Yet Another Steganography Scheme (YASS) is another basic method. It employs first 19 coefficients in the macroblocks to embed the message[10]. “Sarkar et al.” suggested that an 8×8 JPEG block to be selected randomly in a macroblock. The JPEG block is selected based on measures such as total non-zero AC coefficients or block variance [5]. In this article, the total non-Zero AC coefficient measure is applied for YASS.

“Fridrich et al.” proposed PQ method based on perturbing the quantization step of JPEG standard[1] which is detectable by singular value decomposition based features[11]. Then, it was developed by changing the block selection measures to provide its different versions such as PQE, PQT and –PQT [2].

2.2 Steganalysis

Blind steganalysis methods propose a feature set to discover a hidden message using their investigation which is named feature vector. Three widely known feature vectors used in this article, are PEV-274[12], JAN-548[13] and CHEN-390[14].

Another section of the blind steganalysis is classification method. The output of this section can be put in the four group; true assigning of clear images(without hidden message) TN, false assigning of clear image FN, true assigning of stego images(include hidden message) TP and false assigning of stego image FP.

The Steganalysis Evaluation Measures(SEMs) consist of precision, recall, specificity and accuracy are used in evaluating the classification performance. Table 1 shows the SEM definitions. $n(X)$ is defined as a function to represent the number of image of a typical group.

Table 1. Steganalysis Evaluation Measures(SEMs)

Measure	Definition
Precision	$n(TP)/(n\,(TP)+n\,(FP))$
Recall	$n(TP)/(n\,(TP)+n\,(FN))$
Specificity	$n(TN)/(n\,(TN)+n\,(FP))$
Accuracy	$\dfrac{n\,(TP)+n\,(TN)}{n\,(TP)+n\,(TN)+n\,(FP)+n(FN)}$

The precision measure specifies the percentage of true detection of images that have already detected to be a stego. An image will be eliminated soon after it is detected to be a stego. So, the decrease of precision measure can reduce steganalysis applicability[15].

Recall is for recognition of true stego image and its reduction indicates that the hidden information is passing through the system which is quite critical for security systems[15].

Specificity stands for true recognition of clear images. Clearly, a decrease in this measure could lead to dissatisfaction in user not intending to send hidden message[15].

Accuracy is responsible for accuracy of the overall operation of system and it shows the whole system efficiency in average state[15].

All steganalysis methods including those listed at first paragraph of II-B focus on increasing average performance and do not consider their weakness against different images. For this reason, suitable image selection can improve steganography performance, consequently increase the fault probability of steganalysis.

2.3 Image selection

There are two measure categories namely the cover-based measures and the stego-cover-based measures for appropriate image selection. Based on the provided knowledge about used steganography and steganalysis methods; no knowledge, partial knowledge and full knowledge scenarios influence image selection measures[6].

Preprocessing and selecting the suitable image from an image set can speed up the effectiveness of steganography [7]. Fast measures are calculated based on image complexity or texture; while the exact measures are calculated based on image alterations because of the embedded message.

We are going to propose some measures influence steganalysis based on cover image and no-knowledge scenario. Therefore, we must use image properties to propose these features. An important feature category is the texture based one and one of the best tools used in both image indexing based on content[16] and steganalysis[9] is GLCM.

The GLCM is calculated on the basis of relation between pair of points in an image located in a specific distance and direction from each other.

In this article, by deleting direction parameter, GLCM is generalized to correlogram in order to clarify the effect of correlogram properties on blind steganalysis performance. The correlogram and GCLM are discussed more in section III.

2.4 Message and embedding capacity

The embedded message is the aspect of steganography. Lack of special pattern and message length are the important factors affecting steganography and steganalysis. Steganography methods' experiments usually generate random message and embed it to the image in order to avoid special pattern. In many articles, message length is a ratio of image capacity which is affected by the steganography method [1–5, 13]. This is an implicit verification that image content influences the steganography and steganalysis performance. Another subject which brings this hypothesis to mind is the steganography method's dependency on non-zero AC coefficients.

In this research, since the focus is on the effect of image content, the fixed length is used for message instead of image capacity ratio. In fact, using of image capacity ratio is aimed at disregarding the effect of content and concentrating on steganography and steganalysis which would lead to maximum image capacity.

3 Color Correlogram Properties

The Color Correlogram(CC) for grayscale images is very similar to GLCM. GLCM is a tool used in steganalysis feature extracting and identified by(1) where I is a $m{\times}n$ grayscale image and $\triangle x$ and $\triangle y$ are the horizontal and the vertical distance.

$$C_{\triangle x,\triangle y}\left(i,j\right)=\sum_{p=1}^{n}\sum_{q=1}^{m}p\left(I\left(p,q\right)=i,I\left(p+\triangle x,q+\triangle y\right)=j\right) \tag{1}$$

Each index (i,j) is the probability of existence of a pixel with grayscale j in distance identified by $\triangle x$ and $\triangle y$ from a pixel with grayscale i[17]. First, a 1-D histogram is generated using a projection of this matrix along the diagonal line. Next, its three first moments and their characteristic functions are being introduced as the steganalysis features[9, 18].

GLCM properties are used in Content Based Image Retrieval (CBIR) systems as features[19]. These properties are homogeneity, contrast, correlation, energy and entropy(see Table 2). C, i, j, μ and σ are Image, row index, column index, mean and variance for row or column of images, respectively. CC shows the spatial correlation changes of every pair of colors at a specified distance; therefore, we can define it as (2) where d is the distance and $C_{\triangle x,\triangle y}(i,j)$ is defined as in (1).

Table 2. GLCM and CC Properties Identifications[16]

GLCM and CC Properties	Identification
Homogeneity	$\sum_{i,j}C(i,j)/(1+\lvert i-j\rvert)$
Contrast	$\sum_{i,j}\lvert i-j\rvert^{2}C(i,j)$
Correlation	$\sum_{i.j}\left(\left(i-\mu_i\right)\left(j-\mu_j\right)\overrightarrow{C}(i,j)\right)/(\sigma_i\sigma_j)$
Energy	$\sum_{i,j}C(i,j)^{2}$
Entropy	$\sum_{i,j}-\ln\left(C\left(i,j\right)\right)C\left(i,j\right)$

$$Correlogram_{d}\left(i,j\right)=\sum_{\sqrt{\triangle x^{2}+\triangle y^{2}}\le d}C_{\triangle x,\triangle y}\left(i,j\right) \tag{2}$$

The correlogram(d) expression is used to refer to whole CC matrix instead of $Correlogram_d$. Figure 1 shows an image and Figure 2 shows its correlogram(1) according to (2). Although, Figure 2 does not show exact value of each $Correlogram_d\left(i,j\right)$, it shows the ratio of each matrix element to another one. If the number of image color is not decreased, the correlogram dimension for grayscale images would be 256 × 256. Each $pixel(i,j)$ in Figure 2 shows occurrence ratio of color j and color i in one pixel distance for image shown in Figure 1. The white pixel shows the most occurrence and black pixel shows the least. The bar inside the main image in Figure 2 defines the value of each color.

The CC properties are defined in Table 2 extract statistical features of color correlogram matrix. They show the color turbulence of images. An image with more color turbulence is more suitable for steganography and the probability of detecting hidden message in it is decreased.

CC properties are similar to GLCM properties and reflect the color turbulence; but they are defined based on CC matrix instead of GLCM as shown in Table 2.

Correlogram of color decreased images are often used in CBIR[8]. Color decrease is useful for CBIR because we don't need image details there; while, image details are very important in steganalysis. So, in this article, we consider the original image (without color decreasing) and investigate the effects of its CC properties on blind steganalysis. The color correlogram dimension is 256 × 256 which 256 is number of colors in grayscale images.

Figure 1. A sample image

4 Experiments

In order to evaluate the effects of CC properties on steganalysis, we must employ appropriate experiment process, tools and set the involved parameters which will be explained in continue.

4.1 Tools and parameters

BOWS2 image set consist of 10000 grayscale 512×512 image in SGM format[20]. These images are converted to JPEG format with quality factor 98 employed in this research. They are divided into two distinguished parts BOWS2-1 and BOWS2-2 each with 5000 images.

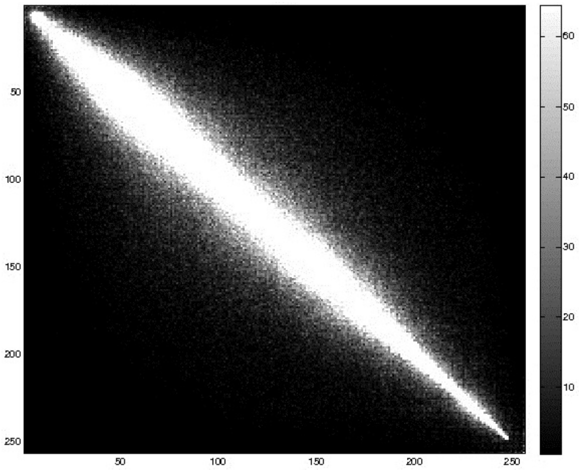

Figure 2. Sample image correlogram(1)

Eight steganography methods: MB1, MB2, nsF5, YASS, PQ, PQE, PQT and –PQT are used for embedding random messages of lengths 256, 512, 768, 1024 and 2048 bytes.

Three feature vectors PEV-274, JAN-548 and CHEN-390 and the Quadratic Support Vector Machine (Q-SVM) as the classification method, are employed for steganalysis.

The correlogram(1) is utilized because of better relation between closer pixels. Homogeneity, contrast, correlation, energy, and entropy are selected as the correlogram(1) properties to investigate their effects on blind steganalysis.

4.2 Experiment steps

The conducted experiment consists of three steps. First step is selection of random message length and applying the steganography method. Second step is choosing BOWS2-1 or BOWS2-2 as training image set and the other one as test image set. In each of training and test steps, to make a composition of both stego and clear images, half of the images are selected randomly and by using a selected steganography method, random messages of selected lengths are embedded in them. In third step, the test images are classified by using a steganalysis vector.

On the other hand, the value of selected correlogram properties is calculated for all clear images. Let d be the interval of the range of value of each property CP and $f(I, CP)$ be a function to calculate the correlogram property CP for the image I. The $H(I, CP)$ shows image scope as (3).

$$H(I,CP)=\begin{cases} 1st\ Scope & 0\le f(I,CP)<d/5 \\ 2nd\ Scope & \frac{d}{5}\le f(I,CP)<2d/5 \\ 3rd\ Scope & \frac{2d}{5}\le f(I,CP)<3d/5 \\ 4th\ Scope & \frac{3d}{5}\le f(I,CP)<4d/5 \\ 5th\ Scope & \frac{4d}{5}\le f(I,CP)\le d \end{cases} \tag{3}$$

So, every image will be allocated to the related steganalysis detection category while it belongs to a specified correlogram property scope. Then, the Steganalysis Evaluation Measures (SEMs) are calculated in each scope separately.

It is clear that, based on tools and parameters used in experiment, the total runs of the process is multiplied by the number of steganography methods, the number of different message length, the number of steganalysis methods and the number of training and test image sets. Here, total runs are 240. In this article, the mean of the obtained results of these runs are used in judging the effects of correlogram properties on the image blind steganalysis.

5 Results

The experiments are conducted based on process and parameters described in the previous section. The obtained results are summarized in Figures 3, 4, 5, 6 and 7.

Figure 3 shows the SEMs values in different homogeneity scopes. Precision, recall and accuracy measures decrease when the homogeneity of correlogram increases. Specificity decreases too when homogeneity of correlogram increases.

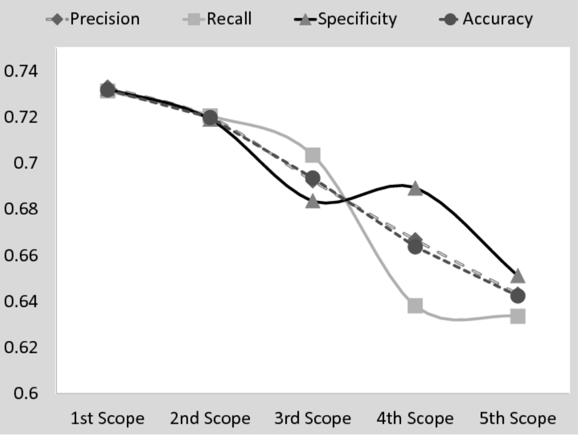

Figure 3. SEMs in homogeneity scopes of correlogram(1)

Figure 4 shows SEMs values in different contrast scopes of correlogram. The SEMs in the 3rd scopes are less than their values in 2nd scopes. In the other scopes, the SEMs values are more than their values in each lower scope.

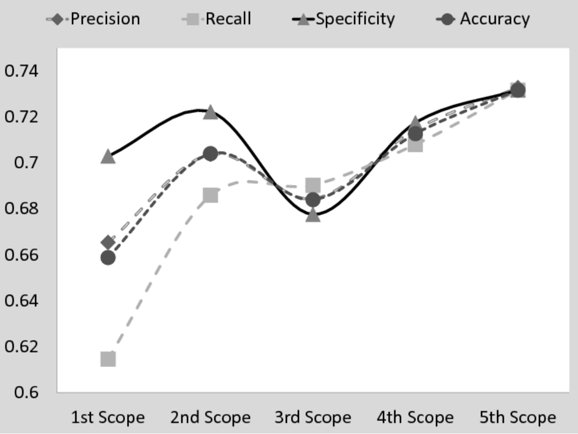

Figure 4. SEMs in contrast scopes of correlogram(1)

Figure 5 shows SEMs values in different correlation scopes of correlogram. In the used image database, there is no image in the 4th scope. Precision and specificity have more values in the scopes have more correlation of correlogram except the 1st scope. Recall and accuracy decrease from the 2nd scope towards the 5th scope. All SEMs values in 2nd scope are less than their values in the 1st scope.

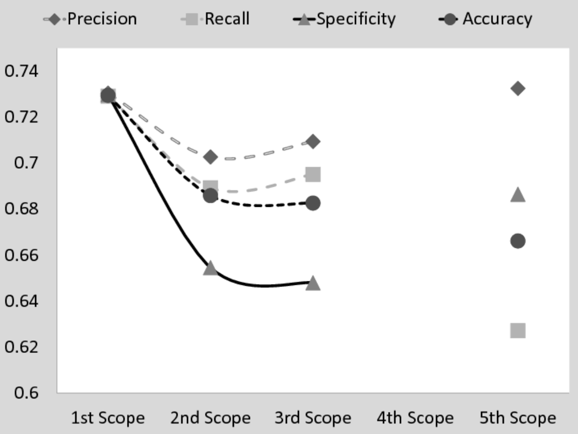

Figure 5. SEMs in correlation scopes of correlogram(1)

Figure 6 shows SEMs values in different energy scopes of correlogram. Precision and specificity have unspecified behaviors. Recall and accuracy values are less than their values in the previous scope from the 3rd towards the 5th scope.

Figure 7 shows SEMs values in different entropy scope of correlogram. All measures have more values in the scopes have more entropy of correlogram. In other words, the probability of steganalysis success would increase by an increasing in correlogram entropy.

The results are summarized in Table 3 with no details. Several relations are used in Table 3 . Inverse

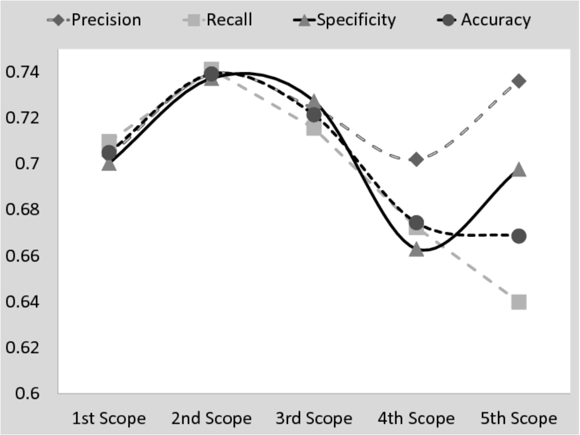

Figure 6. SEMs in energy scopes of correlogram(1)

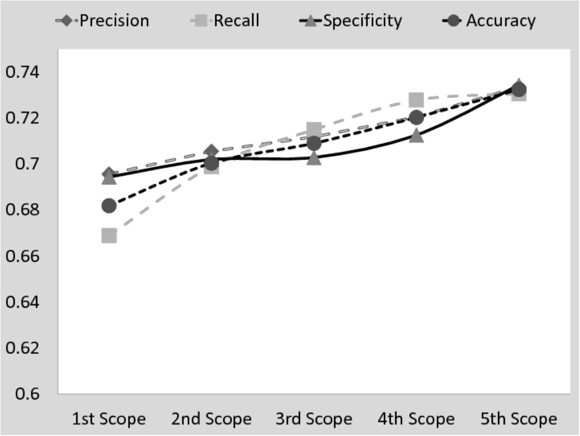

Figure 7. SEMs in entropy scopes of correlogram(1)

relation that means the SEM value would decrease in the scopes where the correlogram property has more value. While the direct relation means SEM value would increase in the scopes where the correlogram property has more value. Semi-direct and semi-inverse relations are similar to direct and inverse relations. But, they do not follow the definition of direct and inverse relation in a number of scopes. Weak relation refers to an unspecified relation.

Table 3. The effect of correlogram properties on SEMs(>>, >,<<, <and ● show direct, semi-direct, inverse, semi-inverse and weak relations respectively)

	Precision	Recall	Specificity	Accuracy
Homogeneity	<<	<<	<	<<
Contrast	>	>>	>	>
Correlation	>	●	>	●
Energy	●	<	●	<
Entropy	>>	>>	>>	>>

6 Conclusion

In this article, the effect of correlogram properties on image blind steganalysis is studied. The results show that an increase in homogeneity and entropy and a decrease in contrast would decrease the SEMs values. The effect of entropy is more obvious than others. Therefore, steganography in the images that their correlogram has more homogeneity and less entropy and contrast would decrease the probability of detection of the hidden message.

Correlogram properties are image texture features. In CBIR, Gabor filter based features, also the texture based features, are more important than correlogram properties. In the continuation of this study, the effects of these features on steganalysis can be useful in finding other properties influence steganalysis. Color based and shape based features which are used for image classification and CBIR systems are other properties that their studying can be useful to find better properties.

References

[1] Jessica J. Fridrich, Miroslav Goljan, and David Soukal. Perturbed quantization steganography with wet paper codes. In Jana Dittmann and Jessica J. Fridrich, editors, *MM&Sec*, pages 4–15. ACM, 2004. ISBN 1-58113-854-7.

[2] Jessica Fridrich, Tomáš Pevný, and Jan Kodovský. Statistically undetectable jpeg steganography: dead ends challenges, and opportunities. In *Proceedings of the 9th workshop on Multimedia & security*, MM&#38;Sec '07, pages 3–14, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-857-2. doi: 10.1145/1288869.1288872. URL http://doi.acm.org/10.1145/1288869.1288872.

[3] Phil Sallee. Model-based steganography. In Ton Kalker, Ingemar Cox, and YongMan Ro, editors, *Digital Watermarking*, volume 2939 of *Lecture Notes in Computer Science*, pages 154–167. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-21061-0. doi: 10.1007/978-3-540-24624-4_12. URL http://dx.doi.org/10.1007/978-3-540-24624-4_12.

[4] PHIL SALLEE. Model-based methods for steganography and steganalysis. *International Journal of Image and Graphics*, 05(01):167–189, 2005. doi: 10.1142/S0219467805001719. URL http://www.worldscientific.com/doi/abs/10.1142/S0219467805001719.

[5] Anindya Sarkar, Kaushal Solanki, and B. S. Manjunath. Further study on yass: Steganography based on randomized embedding to resist blind

steganalysis. URL http://citeseerx.ist.psu.edu/viewdoc/download?p=rep1&type=pdf&doi=10.1.1.139.6691.

[6] Mehdi Kharrazi, Taha Taha Sencar, and Nasir D. Memon. Cover selection for steganographic embedding. In *ICIP*, pages 117–120. IEEE, 2006.

[7] Hedieh Sajedi and Mansour Jamzad. Bss: Boosted steganography scheme with cover image preprocessing. *Expert Syst. Appl.*, 37(12):7703–7710, 2010. URL http://dblp.uni-trier.de/db/journals/eswa/eswa37.html#SajediJ10.

[8] Jing Huang, S. Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, and Ramin Zabih. Image indexing using color correlograms. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pages 762–, Washington, DC, USA, 1997. IEEE Computer Society. ISBN 0-8186-7822-4. URL http://dl.acm.org/citation.cfm?id=794189.794514.

[9] Xiaochuan Chen, Yunhong Wang, Tieniu Tan, and Lei Guo. Blind image steganalysis based on statistical analysis of empirical matrix. In *Proceedings of the 18th International Conference on Pattern Recognition - Volume 03*, ICPR '06, pages 1107–1110, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2521-0. doi: 10.1109/ICPR.2006.332. URL http://dx.doi.org/10.1109/ICPR.2006.332.

[10] Kaushal Solanki, Anindya Sarkar, and B. S. Manjunath. Yass: yet another steganographic scheme that resists blind steganalysis. In *Proceedings of the 9th international conference on Information hiding*, IH'07, pages 16–31, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 3-540-77369-X, 978-3-540-77369-6. URL http://dl.acm.org/citation.cfm?id=1782854.1782857.

[11] Gokhan Gul, Ahmet E. Dirik, and Ismail Avcibas. Steganalytic features for jpeg compression-based perturbed quantization. *IEEE Signal Processing Letters*, 14(3):205–208, 2007. URL http://isis.poly.edu/%7Esteganography/pubs/pq_ieeesp.pdf.

[12] Tomas Pevny and Jessica Fridrich. Merging markov and dct features for multi-class jpeg steganalysis. In *Storage and Retrieval for Image and Video Databases*, volume 6505, 2007. doi: 10.1117/12.696774.

[13] Jan Kodovský and Jessica Fridrich. Calibration revisited. In *Proceedings of the 11th ACM workshop on Multimedia and security*, MM&#38;Sec '09, pages 63–74, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-492-8. doi: 10.1145/1597817.1597830. URL http://doi.acm.org/10.1145/1597817.1597830.

[14] Chunhua Chen, Y.Q. Shi, Wen Chen, and Guorong Xuan. Statistical moments based universal steganalysis using jpeg 2-d array and 2-d characteristic function. In *Image Processing, 2006 IEEE International Conference on*, pages 105–108, 2006. doi: 10.1109/ICI.2006.312383.

[15] Dariush Alimoradi and Maryam Hasanzadeh. Article: The effect of variance difference of dyadic quantized histograms on universal steganalysis. *International Journal of Computer Applications*, 62(8):19–24, January 2013. Published by Foundation of Computer Science, New York, USA.

[16] D S Guru, Y. H. Sharath, and S. Manjunath. Article:texture features and knn in classification of flower images. *IJCA,Special Issue on RTIPPR*, 1(1):21–29, 2010. Published By Foundation of Computer Science.

[17] Wikipedia. Co-occurrence matrix, available at: http://en.wikipedia.org/wiki/; Accessed April 10, 2012. URL http://en.wikipedia.org/wiki/.

[18] Rajarathnam Chandramouli, Mehdi Kharrazi, and Nasir Memon. Image steganography and steganalysis: Concepts and practice. In Ton Kalker, Ingemar Cox, and YongMan Ro, editors, *Digital Watermarking*, volume 2939 of *Lecture Notes in Computer Science*, pages 35–49. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-21061-0. doi: 10.1007/978-3-540-24624-4_3. URL http://dx.doi.org/10.1007/978-3-540-24624-4_3.

[19] Michele Saad. Content-based image retrieval. In *EE381K-14 Multidimensional Digital Signal Processing*. EE381K, 2008.

[20] P.bas and Furon T. Bows2 image database. URL http://bows2.eclille.fr/BOWS2OrigEp3.tgz. available at: http://bows2.eclille.fr/BOWS2OrigEp3.tgz;Accessed April 4, 2012.

Dariush Alimoradi is an IT Expert at the Social Security Organization, Azna, Lorestan, Iran. He received his MS.C. in the field of IT from Shahed University and his B.S. in the field of software engineering from Isfahan University. Now,he is visiting lecturer at Aligoudarz Branch of Payam-Noor University ,too. He teaches OS, Software engineering and computer networks there. His favorite field to research about that is security of Networks, Operating Systems, Applications, Databases, Information and etc. His email addresses are d.alimoradi@shahed.ac.ir and alimoradi.dariush@gmail.com.

Maryam Hasanzadeh was born in 1979 in Mashhad. She received the B.S. degree in computer engineering (software) from Ferdowsi University of Mashhad and MS.C and PH.D degree in computer engineering (Artificial Intelligence) from Sharif university of technology, Tehran, Iran in 2009. She is currently an assistant professor of Shahed University. Her recent research interests include information hiding, image processing, and meta-heuristic search methods.