International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

March 2013 | Issue 14 | Volume 5 | Number 1 | Pages 108-113

A REVIEW ON DATA HIDING UPON DIGITAL IMAGES

N.S. Soleimani Zakeri M.A. Balafar

Department of Information Technology, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, n.zakeri90@ms.tabrizu.ac.ir, balafarila@tabrizu.ac.ir

Abstract- Data hiding is one of the most important techniques for protecting communications in the Internet. It provides a secure method to distribute data through a public and unsafe channel. In this paper, we studied different data hiding methods. These methods were classified into two groups according to their reversibility; and finally, after arguing on each group based on their different features, we compared the performance of representative methods due to their capacity of secret data and bit rates. At the end, we concluded irreversible methods can be embedded by more secret data than a reversible method. However, increasing the capacity of embedded secret data can be done by some encoding methods.

Keywords: Data Hiding, Reversible, Irreversible, VQ Based Images.

I. INTRODUCTION

The use of Internet is increasing and file sharing among users in the Internet is become more than past, so it is necessary to make this communication as secure as possible. Many security problems may result in identity theft because of some hackers trying to steal user's personal data. Data can be hidden in an image in different ways; for example, it can be hidden in noise [1] and inhomogeneity [2] of an image.

There are three major applications in image data hiding era namely steganography, watermarking, and cryptography [3]. Unlike the cryptography, both steganography and watermarking are using a cover image. Despite this similarity, watermarking is used for copyright protection and steganography is focused on anonymity issues and capacity of embedded data. The different image coding methods will be described in the next sections.

II. VQ BASED IMAGE CODING SYSTEM

Because of the popularity of using digital images in the Internet, the necessity for using data hiding methods is dramatically growing. By considering the high compression rate and high quality, VQ based images are more practical for sharing images over the Internet and digital image compression techniques that reduce the size of the digital image are also popular [4, 5].

- Each VQ image is partitioned into series of non-overlapping rectangular blocks and in turn each block maps is divided to a finite subset of VQ blocks. Set *Y* is a codebook generated by a special clustering algorithm. Each member of *Y* is called a codeword $Y = \{y_i \mid i = 1, 2, ..., N\}$, *N* is considered as the size of set *Y*. If the size of the cover image is $H \times W$, it is partitioned into $h * w$ a block. So the size of the index table is $(H/h) \times (W/w)$. Figure 1 shows the encoding process which is describing the mentioned sentences clearly .

There are different ways to generate the codebook, Zongbo and Jiuchao [6] represented that if the separation boundaries are nonlinear, the performance can be decreased fast. In other words they present a kernel fuzzy learning (KFL) algorithm to create a codebook automatically.

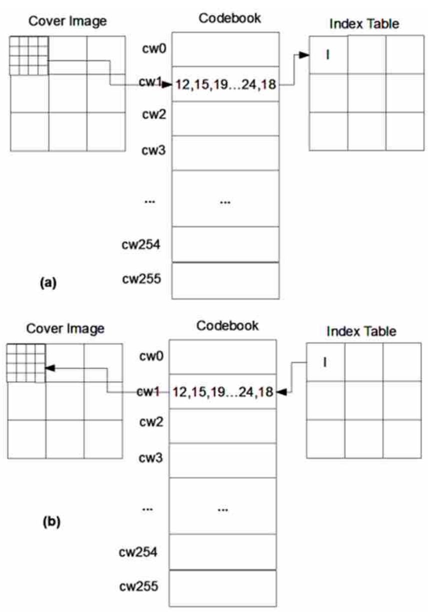

Figure 1. VQ based encoding process [4]

- Search Order coding (SOC) is another image format that is widely used in VQ based image data hiding. The SOC algorithm is proposed to compress the index table and search the nearby blocks flowing a spiral search path.

The n is a pre-defined variable that specifies the size of encoded bits for each block. According to Figure 2, the algorithm begins from a starting positioned at the left side of the current processing block, and flowing in an spiral way as the blocks are processed from left to right and formed a top down process (yellow blocks are excluded because they are not encoded yet), which assigns an n bit code to every indices and finds the same index as current processing blocks [4].

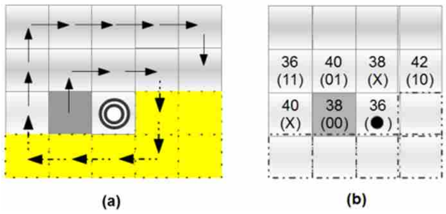

Figure 2. SOC algorithm structure [4]

- Side match vector quantization (SMVQ) [7], is using a different VQ based image format for data hiding. Figure 3 shows the structure of SMVQ algorithm that its detailed structure will be described below.

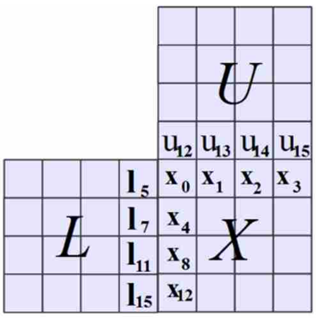

Figure 3. SMVQ algorithm structure [4]

By assuming the current processing block as *X*, and the upper block (i.e., *U*) and the left block (i.e., *L*) have been encoded lately, encoding *X* has to be started. In order to encode *X*, the border values are assigned by its upper and left adjacent blocks *U* and *L*. These values are used to search for n closest code words to form the cod book by Equation (1).

$$D\left(X,Y_i\right)=\sum_{j=0}^{3}\left(x_j-y_{i,j}\right)^2+\sum_{j=1}^{3}\left(x_{4j}-y_{i,4j}\right)^2 \qquad (1)$$

According to this formula, n closest code words construct the state codebook and codewords of the corresponding code which books the minimum Euclidean distance from *X* is used for encoding the current processing block *X*.

A new approach of VQ encoding was proposed [8] which improved the quality of the VQ decompressed images. The main difference between this method and the VQ compression is related to the codewords which are selected for the codebook. In the proposed method, each codeword of its corresponding codebook is a different block rather than an image block. The input is denoted by the original image *O* and a pretrained different codebook *C* and output are denoted by the image difference index table and mean value mi of all blocks in *O*, respectively.

In order to obtain an output from input, taking some consecutive steps are necessary. In the First step, the image will be partitioned in to blocks, and then the parameter mi is computed by taking an average among the pixel values in each block, then difference matrix are derived and that matrix will be partitioned into equal sized blocks. At the final step, the closest codeword will be searched and found in the different codeword. To reconstruct the original image, image difference index table and mean value of all blocks are used. A fast VQ codebook search is presented in [9] which propose an enhanced double test scheme with a file initialization that reduce the time computation. This algorithm takes both a gray scale image and a codebook and in turn returns an index table.

III. CLASSIFICATION OF VQ BASED DATA HIDING METHODS

Data hiding methods get a cover image *C* and a secret data string *S* as an input and gives a Stego image as an output. For recovering purpose, these methods are classified into two groups based on their reversibility. In reversible ones, during recovering process both the secret data and the original image can be obtained, while in the irreversible methods we can't achieve to the original image as the original image is not available. The remaining parts of this paper will describe these two groups and compare the performances of other existing methods.

IV. REVERSIBLE DATA HIDING USING VQ BASED CODES AS OUTPUTS

Now, different methods of reversible data hiding will be presented and their functionality will also be evaluated.

A. Data Hiding for VQ Based Images

An example of the first group is a method that uses one codeword to represent an image block. This method uses SMVQ concepts and creates three state codebook for each block *X*. In the first step G_0 is created, four closest codewords are selected from the main codebook using the distance formula (i.e., the distance formula was introduced in the above part). Then, for each codeword cw_i of G_0, the closest codeword will be found from the same cluster to be assigned as the corresponding codeword of G_1. If there is no close codeword, the corresponding code word will be set to "NULL".

After that, for the aim of embedding the secret data, if *X* is equal to the *i*th codeword of a state cluster and the *i*th codeword of the next cluster is available, one secret bit can be embedded in *X*. Now, if the secret bit is 0, *X* will be remained unchanged; otherwise, *X* is replaced with the *i*th codeword of the next cluster. Another example for this kind of structure, called hit map [4], is to save/record if a codeword is a candidate to replace *X*.

Another research presents the improved data hiding [10] which makes the previous methods perform better than their original versions. To embed the secret data, Chang et al proposed data hiding scheme uses selforganizing sequence list denoted by *L* to design their hiding strategy, but it has some limitations. One of its limitations is to allow just one bit to be embedded and the second one occurs only when the secret bit 0 is embedded; this will compress the index value while this method can compress the input index value where bit 1, bits 10, and bits 00 are embedded.

Another reversible data hiding method which is based on multilevel histogram is introduced here. In this method, they used a multilevel histogram based on different statistics to embed the secret data. The hiding capacity is enhanced compared to the methods based on one and/or two level histograms. To recover the original image, this method uses a sequential strategy and each pixel is reconstructed by its previously recovered neighbor [11].

A prediction based method for data hiding was introduced based on the differences of the vicinity pixels to solve the distortion problem of the previous method. In the previous method, the pixels were altered irrespective of whether it was embedded or not. So this method will firstly calculate the distance between the pixel and its neighbors to determine if it can be embed. Moreover, it uses a predefined threshold and compares that difference with this threshold and if it is higher than the threshold, the algorithms will not modify the pixel and in turn it enhances the quality of the image [12].

Another reversible data hiding method was proposed [13] which were based on mixed base notation and dissimilar pattern strategies for Vector Quantization (VQ) compressed images. Using the declustering process constructs the dissimilar patterns which in turn improves the efficiency of this method. The process of declustering means to cluster the dissimilar code words into the same cluster. After declustering, a mixed based notation is used to embed the secret data.

The results showed that the properties of the image and the number of clustering groups will be an indication for the embedding capacity. The proposed method is also able to extract the secret data and recover the original image without using any additional information**.** Also, a scheme was proposed to recover VQ compressed codes after extracting the data [14]. It uses the codebook that is classified into three subgroups to provide secret data hiding and data recovery. A method for reversible data hiding was also proposed [15] that is not depending on location map so the cost of computation for embedding and extracting is less than others. By applying a customizing factor, this method would result in big payloads and better qualities for input images.

B. SOC Based Methods

SOC based method embeds the secret bit in the blocks that are encoded in SOC. Also, it can embed an unfixed or fixed number of secret bits in each block. To embed a fixed number of secret data, *P* is used a number of secret bits and to embed an unfixed number of secret data, two parameters k and i are used to generate a random number, ranging from 1 to *i*. BR related to this method must be high, if the hit rate of the SOC code is low [4].

Another algorithm that is based on SOC and hiding strategy for creating a novel reversible SOC based data hiding scheme, uses SOC to manipulate the rather random distributed histogram of a VQ compressed image into locations close to zero. Then, it uses encoding strategies to perform encoding and data hiding simultaneously. To improve compression performance of the encoding process, there is no need for an indicator for indices to identify the type of index. This method is reversible, so the original image is completely restorable even after data extraction. The main specifications that weren't in previous methods are as followings: (i) VQ indices are manipulated close to zero by using SOC , (ii) no indicator is required for SOC indices , (iii) the proposed method has very high Embedding Capacity (EC) and Embedding Rate (ER) values [10].

A novel algorithm was presented [16] in which its output is a legitimate SOC code. For embedding the data, it divides the indices in the SOC path into two parts, and then 0 and 1 are separately embedded by the SOC code of each part. And, also in order to achieve to the reversibility feature, this method doesn't need any side information. Another advantage of this algorithm is its flexibility feature to adjust hiding capacity and stego image quality.

Another algorithm to reversible data hiding for VQ images using SOC was also proposed [17]. As in this method, the indices of the input index table are remained unchanged, while an unauthorized user wants to decode this code using standard SOC decoder, nothing suspicious will happen. According to the problem of transmission of side information and the legitimate VQ codes outputs, this algorithm need no side information to produce them, in turn it makes it better than some previous methods. Also in comparison with the methods that produce non-legitimate VQ codes, this one is more secure.

C. Joint Neighboring Coding

According to JNC method, each adjacent block of target block are set to numbers called position flag bits. Two paths are used in this method. In the first path, two bits of the secret data are hidden in each block, but in the second path, three bits of the secret data are hidden in each block. In order to increase the security of this method, the embedding procedure is done by a pseudorandom number and *M* sequence generation [4].

Another method which is based on side match neighbor [18] takes the advantage of the relations among these neighbors. This method has higher compression rate and larger capacity. To explain the proposed method, it firstly uses a rule called the side match neighboring (SMN) encoding rule. The algorithm chooses the root block, then constructs the state codebook from the main codebook and then embedding and extracting stages are performed. The main idea is to utilize the relationship among neighboring blocks to achieve a higher capacity and a better compression rate.

Another new reversible method using vector Quantization (VQ) compressing images was proposed [19] which used the relationships between neighboring blocks to gain reversibility. It also used the side match concept to generate state codebook. At first, the neighbor block encoding is defined, and then a gray block is gained that is used to construct the state codebook from the main codebook and also reconstruct the original index and neighbor blocks. For encoding step, a rule is derived from external areas to internal areas. The gray block is shown in Figure 4.

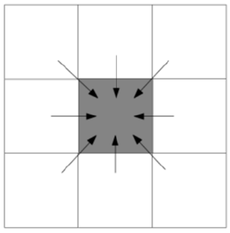

Figure 4. Coding the neighboring blocks concept [19]

In the next step, an algorithm is used to embed and extract data. The input of this algorithm is index table values which are legitimate VQ compresses codes and the output is a side match neighbor code stream. As a result, this method reduces the bit rate and has a high compression rate and capacity.

D. Index Difference

ID method is divided into 3 steps: (i) pre-processing, (ii) embedding, (iii) extracting-recovering.

In the pre-processing step, the codebook will be sorted at first, and then the sorted codebook is used to encode the cover image.

In the embedding step, except the most left-up index which is the datum point in the index value, the difference value for the remainder index is computed by Equation (2):

$$\text{Difference Value of } d_p = \\ = \text{Current Index Value} - \text{Previous Index Value} \quad (2)$$

Then, according to d_p and t (the threshold of a difference value), the output code stream is consisted of different components. At last for extracting and recovering step, the datum point is decoded by VQ immediately and then other three components are being decoded using below rules shown in Figure 5 [4].

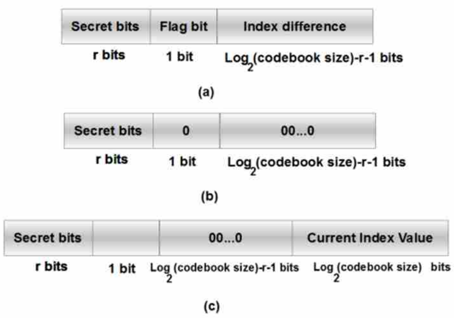

Figure 5. Output code stream in three cases
(a) Case 1 (b) Case 2 (c) Case 3 [4]

A new data hiding method based on VQ compression encoding which is introduced by Ching, Ming et al, [20] uses index difference and reduces the transmission codes. As it restores the index table, it also reduces the length of code stream and each block has the capacity of three bits to embed the secret data. At first, it divides the index table into blocks with a size of $(2m+1)(2m+1)$, and sets I_v as the central point of the block where $v = 2m(2m+1)+1$. The d_x is the difference of the index v and the index x. Then, at the embedding process according to the comparison results of d_x and t (i.e., threshold), there are four cases. Finally, to perform the recovering process, a long code cluster is divided into several groups, each group has the size b and K is the size of the codebook $\left(b = \left[\log_2^K\right]\right)$. Then according to three cases, the recovering process will be computed and completed. As it was explained, this method utilized the vector qualification compression and effectively reduces code stream size.

E. Some Other New Methods for Reversible Data Hiding

Previously introduced methods used Vector-Quantization (VQ) format to embed secret data. In [21], the authors used a modified fast correlation vector quantization (MFCVQ) that enlarges the embedding capacity by embedding multiple bits into a VQ index and by applying the Huffman Code and 0 centered classification, the compressed bit rate was reduced. In comparison with other MFCVQ methods, this method can embed more data and has fewer bit rates.

According to the problem of the most presented methods that the quality of images will be decreased by secret data increasing. The authors in their research [22] used the histogram modification techniques and utilized multiple histograms to increase embedding capacity while marked-image quality is kept reserved.

In another method, the authors used the local binary pattern texture classification approach to achieve a transparent and secure embedded secret data into an image [23]. In another proposed method, they presented a novel reversible data hiding scheme based on combinations of pixel orientations located at two steganographic images. So they enhanced embedding capacity and achieved to a good visual quality. Before embedding step, the algorithm converts the data into a sequence of digits in a base 5 numeral system and by using this method more secret data can be embedded in the image [24].

Paper [25] proposes an information hiding scheme with the least image distortion that can be used in the applications which need high-visual quality. At first, it divides the message into sub-images and each of them is embedded into a pixel vector with three pixels. Comparing with the existing methods that embed only one bit per pixel, this method increases the embedding capacity and as a result modification of a pixel will not be more than one.

V. IRREVERSIBLE DATA HIDING USING VQ BASED CODES

As Irreversible data hiding does not need the precise information of cover image, so it can embed more secret data than reversible methods. Some of these methods are described as below.

A. Vector Quantization

VQ method consists of three procedures. The first procedure is to group codewords. The grouping procedure selects an initial codeword and by considering a threshold, it finds the closest codewords and iteratively continues to put all the codewords into a group.

In the second procedure known as data embedding, the secret data is embedded into the index table. The capacity of secret data depends on the size of the group that the code word belongs to. In the third procedure, the decoder restores the same groups and then extracts the secret bit according to the order of codeword in the group [4].

B. Side Match VQ

SMVQ method is a high capacity image hiding method that uses two thresholds; the first threshold is TH_{sc} in which it determines how the state codebook is constructed and the second one is TH_{smvq} that determines which encoding type (VQ/SMVQ) is used.

If VQ is chosen, three codebooks G_0, G_1, and G_{-1} are constructed in which similar code words are classified into G_0 and G_1 groups and non-similar ones are gathered in G_{-1} group.

If the procedure chooses SMVQ for encoding the data, it constructs a state codebook for encoding block using its upper and left blocks. If the method is set to embed n bits secret data into a SMVQ block, the size of the state codebook will be. After creating the code book, the method uses the embedded secret bits to find the corresponding codeword [4].

Another study applies the concept of side match that for achieving reversibility, it uses extra information called the hit pattern. In order to enhance visual quality of the output stego image, the concept of state codebooks is used. Moreover to partition the codebook into the state codebooks a look-up table is used to minimize embedding and extraction time [26].

C. Search Order Coding

In SOC method, SOC format and multiple codebooks are used. In fact, using a codebook of size 64 will construct four copies of that codebook. Each copy has different indices. For example, CB_1, CB_2, CB_3, and CB_4 ranges are 0-63, 64-127, 128-191, and 192-255 respectively.

These four codebooks are combined to generate a super codebook with size of 256 codewords. So, this method will extract the secret data and embed 6-bits secret messages according to the indices and their coding type [4].

VI. CONCLUSIONS

Data hiding is an important concept to protect data communications through the Internet against the potential attackers. In this paper, we classified the new different methods and their capabilities. It was discussed that reversible methods are more secure than irreversible methods. And we present the new existing methods which have improved the capacity of embedding data in comparison with the irreversible methods. A method based on multilevel histograms was also described and then a prediction based method based on determining pixel's being embedded, was presented to solve the distortion problem of multilevel histograms method. Also some different coding methods and some new formats were introduced in this paper. Recovering the original image without distortion is one of the important concepts in data hiding era which should be taken in to consideration more than before.

REFERENCES

[1] M.A. Balafar, "Review of Noise Reducing Algorithms for Brain MRI Images", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 13, Vol. 4, No. 4, pp. 54-59, December 2012.

[2] M.A. Balafar, "Review of Intensity Inhomogeneity Correction Methods for Brain MRI Images", Issue 13, Vol. 4, No. 4, pp. 60-66, December 2012.

[3] S.S.M. Chow, W.S. Yap, "Partial Decryption Attacks in Security Mediated Certificateless Encryption", IET Inform. Sec., Vol. 3, No. 4, pp. 148-151, Dec. 2009.

[4] W.J. Wang, C.T. Huang, S.J. Wang, "VQ Applications in Steganographic Data Hiding Upon Multimedia Images", IEEE Systems Journal, Vol. 5, No. 4, 2011.

[5] Y.H. Chiou, J.D. Lee, "Reversible Data Hiding Based on Search Order Coding for VQ Compressed Images", Journal of Convergence Information Technology (JCIT), Vol. 6, p. 177, Dec. 2011.

[6] Z. Xie, J. Feng, "Codebook Design for Vector Quantization Based on a Kernel Fuzzy Learning Algorithm", Circuits Syst. Signal Process, Vol. 30, pp. 999-1010, Oct. 2011.

[7] C.C. Chen, C.C. Chang, "High Capacity SMVQ Based Hiding Scheme Using Adaptive Index", Signal Process, Vol. 90, No. 7, pp. 2141-2149, 2010.

[8] J.J. Shen, Y.H. Lo, "A New Approach of Image Compression Based on Difference Vector Quantization", 7th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 137-140, Dalian, China, 2011.

[9] C.C. Chang, Y.P. Hsieh, "A Fast VQ Codebook Search with Initialization and Search Order", Information Sciences, Vol. 183, pp. 132-139, 2011.

[10] C.C. Chang, T.S. Nguyen, C.C. Lin, "A Reversible Data Hiding Scheme for VQ Indices Using Locally Adaptive Coding", J. Vis. Commun. Image, p. 22, 2011.

[11] Z. Zhao, H. Luo. Z.M. Lu, J.S. Pan, "Reversible Data Hiding Based on Multilevel Histogram Modification

and Sequential Recovery", Int. J. Electron. Commun. (AEU), Vol. 65, pp. 814-826, Oct. 2011.
[12] C.C. Chang, Y.H. Huang, H.Y. Tsai, C. Qin, "Prediction Based Reversible Data Hiding Using the Difference of Neighboring Pixels", AEU International Journal of Electronics and Communications, Vol. 66, Issue 9 , September 2012.
[13] C.C. Chang, C.Y. Lin, Y.P. Hseih, "Data Hiding for Vector Quantization Images Using Mixed Base Notation and Dissimilar Patterns without Loss of Fidelity", Elsevier Inc., Vol. 201, pp. 70-79, Oct. 2011.
[14] C.C. Chang, W.C. Wu, Y.C. Hu, "Lossless Recovery of a VQ Index Table with Embedded Secret Data", Journal of Visual Communication and Image Representation, Vol. 18, Issue 3, June 2007.
[15] C.F. Lee, H.L. Chen, "Adjustable Prediction Based Reversible Data Hiding", Digital Signal Processing, Vol. 23, pp. 941-953, Dec. 2012.
[16] P. Rahmani, E. Rahmani, Gh. Dastghaibyfard, "A Location Map Free Reversible Data Hiding Scheme for VQ Compressed Images Based on Search Order Coding", 7th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Dalian, China, Oct. 2011.
[17] P. Rahmani, E. Rahmani, Gh. Dastghaibyfard, "A Reversible Data Embedding Scheme Based on Search Order Coding for VQ Index Tables", 8th International ISC Conference on Information Security and Cryptology (ISCISC), Shiraz University, Shiraz, Iran, Sep. 2011.
[18] C.T. Huang, W.J. Wang, S.J. Wang, "A Novel Data Hiding Scheme Using SMN Blocks in VQ Compressed Images", Opto-Electronics Review, Vol. 19, pp. 225-233, June 2011.
[19] W.J. Wang, C.T. huang, S.J. Wang, "Reversible Data hiding Using Neighboring Block in VQ Compressed Images", ACM Article, New York, 2010.
[20] C.F. Lee, M.C. Shih, "Reversible Data Hiding Based on VQ Compression Code", 5th International Conference On Genetic and Evolutionary Computing, Chaoyang University of Technology, Taichung, Taiwan, Sep. 2011.
[21] C.H. Yang, S.C. Wu, S.C. Huang, Y.K. Lin, "Huffman Code Strategies to Improve MFCVQ Based Reversible Data Hiding for VQ Indexes", Journal of Systems and Software, Vol. 84, pp. 388-396, March 2011.
[22] C.T. Wang, H.F. Yu, "High Capacity Reversible Data Hiding Based on Multihistogram Modification", Multimedia Tools and Applications, Vol. 61, pp. 299-319, Nov. 2011.
[23] E.E. Varasaki, V. Fotopoulos, A.N. Skodras, "Data Hiding Based on Image Texture Classification", Springer-Verlag London Limited, Vol. 15, Jan. 2011.
[24] C.F. Lee, Y.L. Huang, "Reversible Data Hiding Scheme Based on Dual Stegano Images Using Orientation Combinations", Springer Science+Business Media, LLC, July 2011.
[25] C.C. Lin, "An Information Hiding Scheme with Minimal Image Distortion", Computer Standards and Interfaces, Vol. 33, pp. 477-484, Sep. 2011.
[26] C.H. Yang, W.J. Wang, C.T. Huang, S.J. Wang, "Reversible Steganography Based on Side Match and Hit Pattern for VQ Compressed Images", Bibliometrics Data, Vol. 181, pp. 2218-2230, June 2011.

BIOGRAPHIES

Negar Sadat Soleimani Zakeri received her B.Sc. degree in Computer Software Engineering, in 2009. She has been a M.Sc. student in Computer Engineering - Artificial Intelligence at University of Tabriz, Tabriz, Iran, since 2011. Her research interests include artificial intelligence, artificial neural network, image processing, and seismic feature extraction.

Mohammad Ali Balafar was born in Tabriz, Iran, in June 1975. He received the Ph.D. degree in IT in 2010. Currently, he is an Assistant Professor. His research interests are in artificial intelligence and image processing. He has published 11 journal papers and 4 book chapters.