Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 133 (2018) 348–356

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Robotics and Smart Manufacturing (RoSMa2018)

Reversible Data Hiding Scheme During Encryption Using Machine Learning

V. M. Manikandan[a], V. Masilamani[a]

[a]*Indian Institute of Information Technology Design and Manufacturing Kancheepuram, Chennai-600127, India*

Abstract

Reversible data hiding (RDH) is a recent research field of information security for secured digital data transmission. The advancements in communication technology and the invention of new medical robotics are very much useful in telemedicine applications. The transmission of medical image and electronic patient records (EPR) is a common process in telemedicine. The images captured by robots may need to be authenticated, the RDH schemes can be used to authenticate data and/or the owner of the data. In addition, the RDH techniques provide a way to embed EPR data into medical images before transmission. The EPR data extraction and recovery of the original image can be carried out by the receiver. This manuscript proposes a new RDH scheme to embed EPR data during image encryption process. A block-wise image encryption technique has been used in the proposed scheme to obtain the encrypted image with hidden EPR data bits. The novelty of the proposed scheme is that a support vector machine (SVM) based classification scheme has been used for data extraction and image recovery process from the encrypted image. Experimental study of the proposed scheme on standard medical images from OsriX dataset shows that the proposed scheme performs better than the existing schemes in terms of embedding rate and bit error rate.

Keywords: reversible data hiding; medical image transmission; image encryption; support vector machine; feature extraction; medical robotics;

1. Introduction

The process of embedding message bits into the digital content such as images or videos for secured data transmission is known as data hiding. Reversible data hiding (RDH) techniques allow the sender to embed secret information into a digital content like images or videos for secure transmission. The extraction of the secret message and the recovery of the original digital content is also possible by the receiver. The RDH methods are very much useful in telemedicine applications, where we need to transfer medical images from one place to another place along with the electronic patient records (EPR). The RDH methods provide a way to hide EPR data into the medical images itself

* Corresponding author. Mob.: +91-9846503205
E-mail address: coe14d001@iiitdm.ac.in

1877-0509
10.1016/j.procs.2018.07.043

as it is easy to maintain the correspondence between images and respective EPR data. In general medical images are transmitted in DICOM format, and DICOM images will have a predefined header section to keep basic information about the image. The RDH scheme provides a way to transmit additional information in a secure way by embedding it into the image itself instead of sending it as a separate file. If the additional message and the medical image are sending as separate files, then it is very difficult to handle the correspondence between image and data.

The existing RDH schemes are mainly categorized in three: lossless compression based RDH [6, 7, 2, 3], difference expansion based RDH [20, 19, 10, 1] and histogram shifting based RDH [11, 21, 22, 17, 17, 14, 13]. In lossless compression based RDH, a portion of the cover image will be compressed using a lossless image compression technique and the vacant space created through compression will be used for the data hiding purpose. In the difference expansion based technique, the difference between the pixel intensity value is used to embed a single data bit through a difference expansion process. The histogram shifting based RDH scheme embeds additional data bit in the peak pixel intensity in the image. The visual quality of the image after data hiding process is also a major concern in RDH. The image quality assessment techniques can be used to measure the visual quality of the image after hiding the secret messages into an in an image [5, 4].

A few other entirely different techniques are also reported for the purpose of RDH in digital images [12, 23]. Image encryption is an efficient tool to ensure confidentiality of the medical images during transmission. The RDH techniques on encrypted images are also widely studied by the researchers. The popular RDH schemes for encrypted images have been reported in [18, 25, 9, 15]. Almost all the existing RDH methods perform image encryption and data hiding as two different phases. In general, the embedding rate from existing RDH schemes is very less for medical images and so it is not useful for the transmission additional messages.

In the proposed scheme, we combined both RDH and image encryption into a single process. The data hiding will be carried out during image encryption process itself. There is no restriction on the image encryption technique which can be used in the proposed scheme, any secure image encryption can be used in the proposed scheme. The proposed reversible data hiding through image encryption scheme achieves data hiding through the image encryption process.

The proposed scheme ensures the confidentiality of the content through image encryption and allows to embed additional secret message (i.e. EPR) into the digital content. The key idea behind the proposed scheme is that the given cover image will be partition into non-overlapping image blocks and in each block, one-bit additional message can be embedded. The encryption key used to encrypt a particular image block will be determined by the data bit which is going to be embedded in a specific block. In the proposed scheme three encryption keys have been used. The authorized receiver can decode the hidden message along with the recovery of the original image. A trained support vector machine (SVM) model has been proposed for data extraction and image recovery.

2. Proposed Scheme

The proposed algorithms are discussed in this section. A overview of the proposed RDH scheme through encryption is shown in Fig. 1. The proposed data embedding and image encryption processes are briefly discussed in section 2.1.

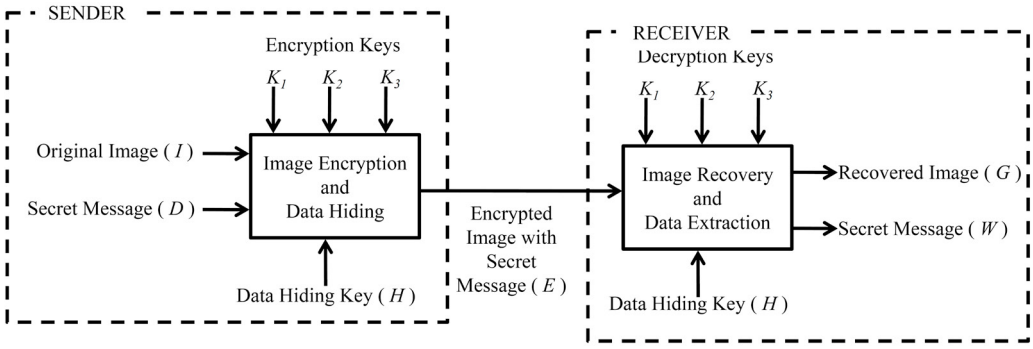

Fig. 1. Overview of the proposed RDH scheme through encryption

As per the proposed scheme, a trained SVM model is required at the receiver side. The SVM training procedure and the feature extraction methods used in the proposed scheme are described in section 2.2. The extraction of data and image recovery process with the help of a trained SVM model is discussed in section 2.3. The frequently used notations in the proposed Algorithms are given below:

Nomenclature

I	Original grayscale image of size $M \times N$ pixels
D	A sequence of bits of length l which represents the additional data to be embedded into the image
D_x	x^{th} bit from D
K_i	Encryption/Decryption keys
H	Data hiding key
I_v	v^{th} block having size of $S \times S$ pixels from image I when scanning linearly in the row-wise order
B	Total number of blocks having size of $S \times S$ pixels in image I
R	Pseudo-random sequence consists of B number of unique integer values in the range $[0, B-1]$ generated based on the data hiding key H
R_x	x^{th} integer value from the pseudo-random sequence R
E	Encrypted image of size $M \times N$ with embedded data
E_v	v^{th} block having size of $S \times S$ pixels from image E when scanning linearly in the row-wise order
G	Recovered image after data extraction
G_v	v^{th} block having size of $S \times S$ pixels from image G when scanning linearly in the row-wise order
W	Extracted bit sequence of length l from E

2.1. Data Hiding Through Image Encryption

The proposed data hiding scheme through image encryption is given in Algorithm 1. As per the proposed scheme, a given grayscale image I of size $M \times N$ pixels will be divided into non-overlapping blocks of size $S \times S$ pixels. For better security, the blocks in the original image I will be accessed in a pseudo-random way based on the data hiding key H. The pseudo-random sequence of integers for encrypting a selected image block C will be determined by the secret message bit D_x which is going to be embedded into C. If D_x is 1 then C will be encrypted using the encryption key K_1. If D_x is 0 then C will be encrypted using the pseudo-random sequence of integers generated by the encryption key K_2. Sometimes, all the image blocks may not be embedded with some message bits, so, if the sender doesn't want to embed any secret message bits into a selected block then that block will be encrypted by using encryption key K_3. This operation will be carried out for all the image blocks in the image I to obtain the final encrypted image. It should be noted that there will be some regions in the image remained after the dividing the image I into blocks, such regions will be encrypted using the pseudo-random integers generated by the encryption key K_3.

Algorithm 1 : Proposed reversible data hiding scheme through encryption

Input : I, D, K_1, K_2, K_3, H
Output : E

1. Divide the given image I into non-overlapping blocks of size $S \times S$ pixels
2. Initialize a matrix E of size $M \times N$ with zeros
3. Find pseudo-random integer sequence R using data hiding key H
4. For $x = 1$ to l
5. $\quad v = R_x,\ C = I_v$
6. $\quad$ If $D_x == 1$

7. $E_v = Encrypt(C, K_1)$
8. Else
9. $E_v = Encrypt(C, K_2)$
10. EndIf
11. EndFor
12. For $x = (l + 1)$ to B
13. $v = R_x,\ C = I_v$
14. $E_v = Encrypt(C, K_3)$
15. Copy E_v into the v^{th} block of E
16. EndFor
17. Define a region A_R from the image I such as $A_R = \{I(i, j) \mid \lfloor M/S \rfloor \times S \leq i < M \text{or} \lfloor N/S \rfloor \times S \leq j < N\}$ by ordered row-wise scanning on image I
18. $E_A = Encrypt(A_R, K_3)$
19. Copy E_A into the remaining portions of E
20. Return E

2.2. *Support Vector Machine (SVM) Training*

The key idea behind the proposed scheme is that depending on the secret message bit, a specific encryption key among the encryption keys K_1, K_2 and K_3 will be used to encrypt a specific image block. At the receiver side, the receiver knows that a given block is encrypted using encryption key K_1 or K_2 or K_3, but don't know specifically which one among those three. In general, the encrypted image blocks will have noise-like structure and there will not be any correlation between the adjacent pixels [24]. The attempt to decrypt an image block with a wrong encryption key will lead to obtaining an image block having a random structure again. But only the originally decrypted version will have the natural image properties (high correlation between the pixels). In the proposed scheme the SVM classifier is used to classify an image block into any one of the two classes: *natural* (class label: 0) or *encrypted* (class label: 1). The block-size $S \times S$ determines the embedding rate of the proposed data hiding scheme. In this work, we have considered image blocks of size 9×9 during training and further data hiding process. The block-size 9×9 is empirically decided to achieve good embedding rate without compromising recoverability of the original image during data extraction and image recovery. We propose a learning scheme to learn the properties or features of the encrypted block. The overview of the training process is given in Fig. 2.

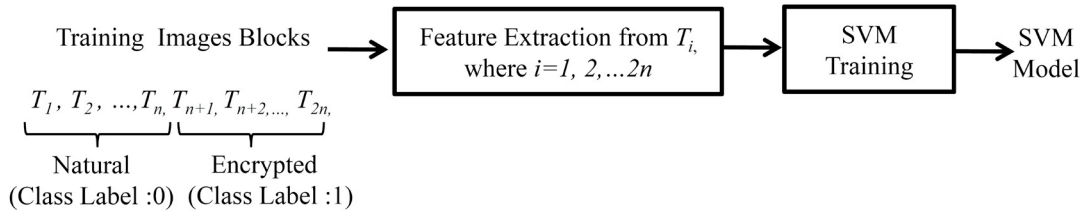

Fig. 2. Training process in the proposed scheme

The SVM in the proposed scheme has been trained using 10000 natural image blocks and 10000 encrypted image blocks. During the training phase, the image blocks have been selected randomly from the 5000 images to capture the different properties of the images. The trained SVM model should be shared with the receiver for data extraction and image recovery process.

From each image block selected for training, 6 different features have been extracted for training an SVM model. The methods used in the proposed scheme to extract 6 features from a selected image block B of size $S \times S$ are detailed below:

1. **Standard deviation (SD):** Standard deviation is a measure to quantify the variation of a set of data. Here, the set of pixel values in a given block is considered as data. The SD value for an image block B of size $m \times n$ can be defined as follows:

$$SD = \sqrt{\frac{1}{S^2}\sum_{i=1}^{S}\sum_{j=1}^{S}(B(i,j)-\mu)^2} \tag{1}$$

where the μ is the mean of the pixel values in the image block B. In general, the SD value from a natural image block will be lower as compared to the SD value from the corresponding encrypted image block.

2. **Peak in the histogram after quantization (PK):** To extract the PK feature from the image block B, firstly, B should quantize into 64 levels. Let us denote the quantized image block by B_Q. The definition of the histogram for B_Q is as given in equation (2).

$$\begin{aligned} &H(P) = \#\{(i,j)|B(i,j) = P\} \\ &\quad \text{where } P = 0, 1, \ldots 63, i = 1, 2, \ldots m, j = 1, 2, \ldots n \end{aligned} \tag{2}$$

The number of pixels in the peak of the histogram corresponds to B_Q is considered as the feature PK. In an encrypted block, the pixel values will be uniformly distributed. Due to this, the histogram corresponds to the encrypted image block will be almost flat. But in a natural image block (after quantization), there is a high probability to find a peak having a larger size in the histogram.

3. **Maximum vote from histogram bins (VH):** To extract this feature, the histogram has been divided into 32 bins, where each bin will have 8 different grayscale values. The total number of pixels belongs to one bin is considered as the vote from that bin. The maximum vote from any of the bin is considered as the VH feature value. It should be noted that in encrypted image, the pixel values may be uniformly distributed. So the VH from an encrypted image will be less as compared the same from a natural image block. The pixels within a natural image block will be very close so that most of the pixels may belong to the same bin. Due to this, vote from that bin will be high.

4. **Entropy (EP):** Entropy is a statistical measure used to analyze the randomness of the pixel values in an image block [8]. The pixel values in an image block have been quantized into 64 levels before finding the randomness. The quantization process helps to bring the very close pixel values into the same value, which in turn will help to get low entropy measure from a natural image block with the close pixel values. The definition of entropy for a block of pixels with values 0 to 63 is given in equation (3).

$$EP = -\sum_{K=0}^{63} P(K)\log_2 P(K) \tag{3}$$

where EP is the entropy measure obtained from the image block, $P(K)$ indicates the probability of the pixels with intensity value K in the given block. The entropy measure for an encrypted image block will be higher as compared to the same for a natural image block due to the high randomness of pixel values in the encrypted image blocks.

5. **Correlation between adjacent pixels (CP):** This is also one of the main features that we can use to classify an encrypted image block from a natural image block. The correlation between adjacent pixels, say CP for the image block B is defined as follows:

$$CP = \frac{|\sum_{i=1}^{S}\sum_{j=1}^{S}(B(i,j)-\mu)(B(i,j+1)-\mu)|}{\sqrt{\sum_{i=1}^{S}\sum_{j=1}^{S}\Big(B(i,j)-\mu\Big)^2}\sqrt{\sum_{i=1}^{S}\sum_{j=1}^{S}\Big(B(i,j+1)-\mu\Big)^2}} \tag{4}$$

where μ is the mean value of the pixels in an image block B. Note that there will be a high correlation between the pixels in the natural image block, which will lead to getting a higher CP value for a natural image block as compared the same measure for an encrypted image block.

6. **Smoothness measure (SM):** The smoothness of an image block can be used as a feature to classify an natural image block from an encrypted image block. A smoothness measure SM from an image block B of size $S \times S$ by considering the 4-neighborhood pixel values is defined as follows:

$$SM = \frac{1}{S^2}\sum_{i=2}^{S-1}\sum_{j=2}^{S-1}\left|B(i,j) - \left(\frac{B(i-1,j)+B(i,j+1)+B(i+1,j)+B(i,j-1)}{4}\right)\right| \tag{5}$$

For a natural image block the smoothness measure SM as per the equation (5) will be very low as compared the same measure for an encrypted image block.

The feature vector corresponds to the k^{th} image block is represented by $F_k = (SD, PK, VH, EP, CP, SM)$. The SVM model has been trained using the feature vectors extracted from 10000 natural image blocks and 10000 encrypted image blocks. The trained SVM model is a binary classifier, it will classify a given image block into any one of the two classes: natural or encrypted. The trained model should be available to the receiver for data extraction and image recovery.

2.3. Data extraction and Image Recovery

Algorithm 2 shows the steps for data extraction and the image recovery process. For data extraction and image recovery, the receiver should know the decryption keys K_1, K_2, and K_3 and the data hiding key H. The key idea behind the data extraction and image recovery process is that the receiver will try to decrypt each of selected block using all the three different keys K_1, K_2 and K_3. Let us denote the three decrypted versions of an image block C by C_1, C_2 and C_3. The 6 features (discussed in section 2.2) will be extracted from each of the image blocks C_1, C_2 and C_3. These features of the image blocks will be given to the trained SVM model and it will identify the correctly decrypted version of the image block. The decryption key used to recover the original image block gives the secret message bit.

Algorithm 2 : Data extraction and image recovery

Input : E, K_1, K_2, K_3, H
Output : G, W

1. Divide the image E into non-overlapping blocks of size $S \times S$ pixels
2. Initialize a matrix G of size $M \times N$ with zeros to keep the recovered image
3. Find pseudo-random integer sequence R using data hiding key H
4. $B = \lfloor M/S \rfloor \times \lfloor N/S \rfloor$
5. For $x = 1$ to B
6. $\quad v = R_x\ ,\ C = E_v,\ \ k = 1$

7. $C_1 = Decrypt(C, K_1)$, $C_2 = Decrypt(C, K_2)$, $C_3 = Decrypt(C, K_3)$
8. Classify C_1, C_2 and C_3 using the trained SVM Model, denote the results as c_1, c_2 and c_3 respectively. Note that c_1, c_2 and c_3 are labels which will be either 0 or 1 (0 corresponds natural image block and 1 corresponds to encrypted image block).
9. If $(c_1 == 0)$
10. $G_v = C_1$, $W_k = 0, k = k + 1$
11. ElseIf $(c_2 == 0)$
12. $G_v = C_2$, $W_k = 1, k = k + 1$;
13. Else
14. $G_v = C_3$
15. EndIf
16.
17. EndWhile
18. Define a region A_E from the image E such as $A_E = \{I(i, j) \Big| \lfloor M/S \rfloor \times S \leq i < M \text{or} \lfloor N/S \rfloor \times S \leq j < N\}$ by ordered row-wise scanning on image E
19. $A_D = Decrypt(A_E, K_3)$
20. Copy A_D into to the remaining portions of G
21. Return G and W;

3. Experimental Study and Result Analysis

For experimental study, 5000 images selected from OsriX dataset have been converted into 8-bit grayscale images of size 512×512 pixels. The algorithms have been implemented and tested using Matlab2017a in a workstation having 32 GB RAM with Intel(R) Xeon(R) CPU, 3.46 GHz. For the image encryption purpose the well-known *RC*4 image encryption scheme has been used [16]. The proposed scheme has been evaluated using the efficiency parameters such as embedding rate, bit error rate (BER) between embedded data and extracted, and computational time complexity. The proposed scheme has been compared with the well-known reversible watermarking schemes reported in [25, 9]. All the comparison results are given in Table 1.

Table 1. Comparison between the existing schemes [25, 9] and proposed scheme

Scheme	Embedding rate (in bpp)	BER	Execution time (in Seconds)		Theoretical Time complexity	
			DH&IE	DE&IR	DH&IE	DE&IR
Scheme in [25]	9.76×10^{-4}	2.61×10^{-1}	3.68×10^{-1}	6.64×10^{-1}	$O(N^2)$	$O(N^2)$
Scheme in [9]	9.76×10^{-4}	2.53×10^{-1}	3.68×10^{-1}	1.74	$O(N^2)$	$O(N^2)$
Proposed Scheme	1.23×10^{-2}	4.74×10^{-3}	4.76×10^{-2}	5.26	$O(N^2)$	$O(N^2)$

3.1. Embedding Rate

The number of data bits that can be embedded into one single pixel of the image is termed as embedding rate of a data hiding scheme. In general, embedding rate will be measured by bits per pixels (bpp). As per the new scheme, one secret message bit can be embedded into a 9×9 image block, therefore the embedding rate of the new scheme will be $\frac{1}{(9\times9)}$, which is 0.01234 bbp.

3.2. Bit Error Rate (BER)

The BER can be used as a measure to find the extraction capability of a RDH scheme as it compares the extraction secret message bits the embedded secret message bits. The BER is defined in Equation 6.

$$BER = \frac{E}{T} \tag{6}$$

where E indicates the the number of bits extracted wrongly, and T is the total number of bits inserted in the given image. The ideal case is a BER of 0. From Table 1, it can be observed that the BER from the proposed scheme is very less as comapred to BER from the existing schemes reported [25, 9].

3.3. Time Complexity Analysis

Theoretical time complexity and execution time from the proposed scheme and existing schemes [25, 9] are given in Table 1. In Table 1, the comparisons have been given for data hiding and image encryption (DH&IE), and data extraction and image recovery (DE&IR).

A training process is involved in the proposed scheme to obtain the SVM model. But here training is a one-time off-line process, therefore we are not considering it for theoretical time complexity analysis. The same trained model can be used by any number of users and it does not have any dependency on the encryption/decryption keys used by the users. The data hiding through image encryption can be carried out with $O(N^2)$ time when the input image has a size of $N \times N$ pixels. The reason is that in the proposed scheme, all the pixels in the image need to be accessed at most once, and the constant time is required for the encryption purpose. Similarly, data extraction and image recovery process can also be done with the time complexity of $O(N^2)$. From Table 1, it can be seen that the data extraction and image recovery time is higher than the time required for the existing reversible data hiding schemes.

4. Conclusion

A reversible data hiding scheme through image encryption and support vector machine based data extraction and image recovery scheme have been proposed in this paper. The proposed scheme provides a way for secure medical image transmission. The electronic patient record which contains basic details about patients can be embedded into the medical images itself, instead of sending it as a separate file. The sender and receiver should agree on three different encryption/decryption keys K_1, K_2 and K_3, and a block-wise image encryption process will be performed on the cover image. The selection of encryption keys is determined by secret message bit which we need to embed to into a specific block. A trained SVM model helps the receiver to extract the secret message along with the recovery of the image. The experimental study of the proposed scheme on the standard OsriX medical image dataset shows that the proposed scheme outperforms the existing schemes in terms of embedding rate and bit error rate. The RDH schemes assume that the images will be transmitted through the lossless channels, but in many cases the data may transmitted through the lossy channels. The future works can be concentrated to propose robust RDH schemes for secure data transmission.

References

[1] Al-Qershi, O.M., Khoo, B.E., 2011. High capacity data hiding schemes for medical images based on difference expansion. Journal of Systems and Software 84, 105–112.

[2] Celik, M.U., Sharma, G., Tekalp, A.M., Saber, E., 2002. Reversible data hiding, in: International Conference on Image Processing, IEEE. pp. II–II.

[3] Celik, M.U., Sharma, G., Tekalp, A.M., Saber, E., 2005. Lossless generalized-lsb data embedding. IEEE Transactions on Image Processing 14, 253–266.

[4] De, K., Masilamani, V., 2013. A new no-reference image quality measure for blurred images in spatial domain. Journal of Image and Graphics 1, 39–42.

[5] De, K., Masilamani, V., 2017. No-reference image contrast measure using image statistics and random forest. Multimedia Tools and Applications 76, 18641–18656.

[6] Fridrich, J., Goljan, M., Du, R., 2001. Invertible authentication, in: Security and Watermarking of Multimedia contents, International Society for Optics and Photonics. pp. 197–209.

[7] Fridrich, J., Goljan, M., Du, R., 2002. Lossless data embedding: new paradigm in digital watermarking. Journal on Applied Signal Processing 2002, 185–196.
[8] Gonzalez, R.C., Woods, R.E., et al., 2002. Digital image processing.
[9] Hong, W., Chen, T.S., Wu, H.Y., 2012. An improved reversible data hiding in encrypted images using side match. IEEE Signal Processing Letters 19, 199–202.
[10] Hu, Y., Lee, H.K., Chen, K., Li, J., 2008. Difference expansion based reversible data hiding using two embedding directions. IEEE Transactions on Multimedia 10, 1500–1512.
[11] Hwang, J., Kim, J., Choi, J., 2006. A reversible watermarking based on histogram shifting, in: International Workshop on Digital Watermarking, pp. 348–361.
[12] Li, F., Mao, Q., Chang, C.C., 2016. A reversible data hiding scheme based on iwt and the sudoku method. International Journal of Network Security 18, 410–419.
[13] Li, X., Zhang, W., Gui, X., Yang, B., 2015. Efficient reversible data hiding based on multiple histograms modification. IEEE Transactions on Information Forensics and Security 10, 2016–2027.
[14] Lu, T.C., Tseng, C.Y., Deng, K.M., 2014. Reversible data hiding using local edge sensing prediction methods and adaptive thresholds. Signal Processing 104, 152–166.
[15] Ma, K., Zhang, W., Zhao, X., Yu, N., Li, F., 2013. Reversible data hiding in encrypted images by reserving room before encryption. IEEE Transactions on Information Forensics and Security 8, 553–562.
[16] Mousa, A., Hamad, A., 2006. Evaluation of the rc4 algorithm for data encryption. IJCSA 3, 44–56.
[17] Ou, B., Li, X., Zhao, Y., Ni, R., 2013. Reversible data hiding based on pde predictor. Journal of Systems and Software 86, 2700–2709.
[18] Puech, W., Chaumont, M., Strauss, O., 2008. A reversible data hiding method for encrypted images, in: Electronic Imaging, SPIE/IS&T. p. 68191E.
[19] Thodi, D.M., Rodriguez, J.J., 2004. Prediction-error based reversible watermarking, in: International Conference on Image Processing, IEEE. pp. 1549–1552.
[20] Tian, J., 2003. Reversible data embedding using a difference expansion. IEEE Transactions on Circuits and Systems for Video Technology 13, 890–896.
[21] Tsai, P., Hu, Y.C., Yeh, H.L., 2009. Reversible image hiding scheme using predictive coding and histogram shifting. Signal Processing 89, 1129–1143.
[22] Wu, H.T., Huang, J., 2012. Reversible image watermarking on prediction errors by efficient histogram modification. Signal Processing 92, 3000–3009.
[23] Wu, H.T., Huang, J., Shi, Y.Q., 2015. A reversible data hiding method with contrast enhancement for medical images. Journal of Visual Communication and Image Representation 31, 146–153.
[24] Wu, Y., Zhou, Y., Saveriades, G., Agaian, S., Noonan, J.P., Natarajan, P., 2013. Local shannon entropy measure with statistical tests for image randomness. Information Sciences 222, 323–342.
[25] Zhang, X., 2011. Reversible data hiding in encrypted image. IEEE Signal Processing Letters 18, 255–258.