Evaluation of different embedding methods for JPEG authentication watermarking

A A Egorova[1], V A Fedoseev[1,2]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
[2]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: varlamova.anna.95@mail.ru

Abstract. This paper considers the applicability of different data embedding methods for semi-fragile watermarking systems used for JPEG image authentication. The methods include Least Significant Bit watermarking and various versions of Quantization Index Modulation. In our investigations, we tested the semi-fragility property against JPEG and compared the visual quality of the watermarked images. We also checked the watermark fragility to unacceptable modifications like median filtering, blurring, and adding Gaussian noise. Finally, we analyzed the provided tampering localization error.

1. Introduction

One of the ways to protect an image from tampering is to embed a fragile or a semi-fragile digital watermark, a barely visible and removable component, whose presence in the image may testify its authenticity [1]. Fragile watermarks are destroyed after any image modifications and are usually used for the data integrity verification. If a specific set of modifications is considered to be acceptable, semi-fragile watermarks are applied to authenticate the data. They are robust against permitted transformations and fragile to any other. As a rule, these permitted transformations include modifications that do not affect image content and structure, for example, weak distortions caused by lossy compression.

The most common standard for lossy image compression is JPEG. More than 20 semi-fragile JPEG watermarking systems have been developed since 2000. The most widespread among them are those that embed a watermark in the frequency domain, namely in the Discrete Cosine Transform (DCT) coefficients before or after quantization [2-15]. The watermarks embedded by such systems are visually imperceptible and JPEG-resistant even at low values of the quality factor.

The effectiveness of a particular semi-fragile system depends mostly on its data embedding method. For this reason, in this paper, we investigate the influence of different embedding methods on the performance of the JPEG semi-fragile watermarking. We consider and compare the methods that are commonly applied in JPEG-resistant watermarking. They include Least Significant Bit (LSB) watermarking [1], Quantization Index Modulation (QIM) [16], and its versions (Sign-QIM [17], MOD-QIM [18], and DM-QIM [16]). In the experimental part, we test their applicability to semi-fragile JPEG watermarking and compare the Peak Signal-to-Noise (PSNR) values of the obtained watermarked images. We also verify the fragility of the considered embedding methods to

unacceptable distortions (exemplified by median filtering, blurring, and adding white Gaussian noise) and analyze the tampering localization error.

The rest of this paper is organized as follows. In Section 2, the lossy JPEG compression scheme is described. In Section 3, the description of considered data embedding methods is given. Section 4 presents the experimental results.

2. Lossy JPEG compression scheme

The JPEG lossy compression algorithm consists of the following key steps (also shown in Figure 1) [17].

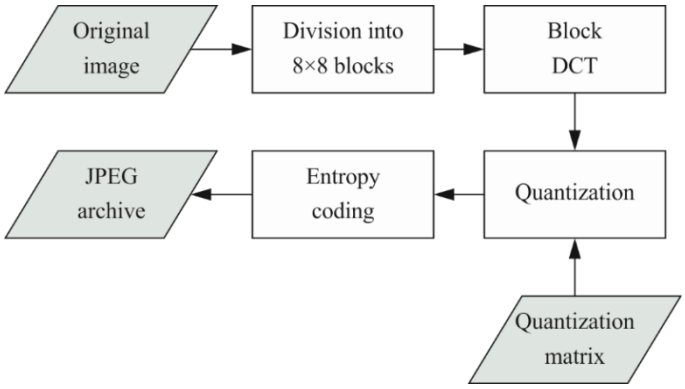

Figure 1. Lossy JPEG compression scheme.

1. Division of the original $N_1 \times N_2$ image I into 8×8 nonoverlapping blocks I_i, where $i = 1,..,N$ and $N = N_1 N_2 / 64$ is the total number of nonoverlapping blocks in the image.
2. Calculating blockwise DCT. DCT decomposes the image values into different frequencies. We denote each obtained block of DCT coefficients as $B_i(m_1, m_2)$. The coefficients in the upper left corner (Figure 2) characterize the low frequency component.
3. Quantization of each block B_i using the quantization matrix Q_{QF} of size 8×8, corresponding to the predetermined compression quality factor QF (from 1 to 100).

$$D_i(m_1, m_2) = round\left(\frac{B_i(m_1, m_2)}{Q_{QF}(m_1, m_2)}\right). \quad (1)$$

The smaller QF is, the higher the values of coefficients of the quantization matrix Q_{QF}, more zeros among quantized DCT coefficients $D_i(m_1, m_2)$, and the smaller the size of the resulting archive.

4. Scanning each block $D_i(m_1, m_2)$ in zigzag order, as shown in Figure 2, and entropy coding. Further, we denote DCT coefficients shortly as $D_i(j)$, where $j = 1..64$ is the index of an element in zigzag order.

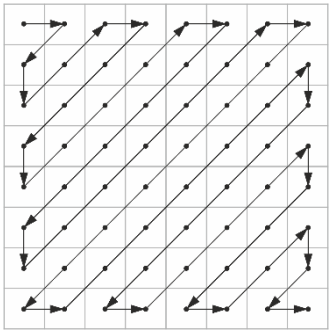

Figure 2. Zigzag scanning of a DCT block.

3. Data embedding methods used in JPEG-resistant watermarking

Data embedding is a key step of any JPEG semi-fragile watermarking system. It determines the way of modifying the spectral components. The most common data embedding methods actively used in JPEG-resistant watermarking are the LSB [1] and QIM variations [16-18].

It is worth mentioning the loss of information in the JPEG algorithm occurs at the DCT coefficients quantization stage. For this reason, frequency domain watermarking systems embed the watermark either at the quantization step or immediately after it.

LSB embedding in JPEG semi-fragile systems is performed after the quantization by replacing one or more of the quantized DCT image coefficients with watermark bits [4, 5]. Let us assume that the number of modified DCT coefficients in each block is equal to the number of bits to be embedded per block N_W. We also assume that inter-coefficient relationships are not taken into account during the watermark embedding process. We denote the positions of the modified DCT coefficients in zigzag order as j_k, where $k = 1..N_W$. In general, they are defined by the secret key. Then the LSB method embeds the watermark by changing the quantized DCT coefficients located in the j_k positions:

$$D_i^W(j_k) = 2\lfloor D_i(j_k)/2 \rfloor + W_{i,k}, \tag{2}$$

where $W_{i,k}$ is the k^{th} bit of information that is embedded in the i^{th} quantized DCT block. All coefficients excluding j_k remain unchanged. The watermark extraction procedure for this method is obvious.

LSB is actively used [3-5] because it does not require high computational cost, simple to implement, and makes possible to hide a sufficiently large amount of information. However, its application in the DCT frequency domain may cause significant distortions.

The QIM-based methods usually lead to smaller distortions of the watermarked image. Unlike LSB, QIM-based techniques embed a watermark while quantizing DCT coefficients. They modulate the DCT coefficients by the watermark bits [16]. In the JPEG semi-fragile systems, various versions of QIM are in use [6-9]. At first, we consider the method applied in the Preda & Vizireanu watermarking system [7]:

$$B_i^W(j_k) = round\left(\frac{B_i(j_k)}{2Q_{QF}(j_k)} - W_{i,k}\right)2Q_{QF}(j_k) + W_{i,k}Q_{QF}(j_k), \tag{3}$$

$$W_{i,k} = round\left(B_i^W(j_k)/Q_{QF}(j_k)\right)(\mathrm{mod}\,2). \tag{4}$$

Note that (3) is similar to LSB. As in the case of LSB, the components $B_i^W(j_k)$ are multiples of the quantization steps $Q_{QF}(j_k)$. The second QIM-based method we consider in this paper is Sign-QIM – a simple modification of the method by Preda & Vizireanu. Its distinctive feature lies in the fact that the watermark component sign depends on the direction to which the modified DCT coefficient is rounded off at the quantization stage. Due to this, the error in the coefficient j_k caused by information embedding does not exceed $Q_{QF}(j_k)$:

$$B_i^W(j_k) = Br_i(j_k) + S_i(j_k)\cdot W_{i,k}\cdot Q_{QF}(j_k), \tag{5}$$

where

$$Br_i(j_k) = round\left(\frac{B_i(j_k)}{2Q_{QF}(j_k)}\right)2Q_{QF}(j_k). \tag{6}$$

The third method is DM-QIM, which is the most known QIM version [16]. It subtracts the noise-like component, which is previously added to the host image components, to avoid a mean value shift instead of adding the remainder of dividing by the quantization step:

$$B_i^W(j_k) = round\left(\frac{B_i(j_k) + d_{W_{i,k}}(j_k)Q_{QF}(j_k)}{2Q_{QF}(j_k)}\right)2Q_{QF}(j_k) - d_{W_{i,k}}(j_k)Q_{QF}(j_k), \tag{7}$$

where $d_0(j), d_1(j) \in \mathbf{R} \cap [-1;1)$ are two pseudorandom arrays used to modulate watermark bits, and

$$d_1(j) = d_0(j) - sign(d_0(j)).$$

One more QIM-based embedding method was proposed in paper [18] by Glumov & Mitekin. It has a wider range of obtained values than Preda & Vizireanu. This method is not intended to provide robustness against JPEG compression, so the embedding is performed in the spatial domain. Another distinction of [18] from [7] is that it conducts the *floor* operation $\lfloor x \rfloor$ instead of $round(x)$. Thus, the embedding a single bit w into a single image component x by [18] is as follows:

$$x^W = \lfloor x/2\delta \rfloor 2\delta + w\delta + x(\mathrm{mod}\,\delta), \quad (8)$$

where δ is the quantization step. In (8), the last summand provides exactly the extension of the range of x^W values. We denote this QIM-based method as MOD-QIM.

To incorporate MOD-QIM method with the JPEG compression procedure, we bring in modifications to the rounding function keeping the remainder as follows:

$$B_i^W(j_k) = Br_i(j_k) + S_i(j_k) \cdot W_{i,k} Q_{QF}(j_k) + M_i(j_k), \quad (9)$$

where

$$Br_i(j_k) = round\left(\frac{B_i(j_k)}{2Q_{QF}(j_k)}\right) 2Q_{QF}(j_k),$$

$$S_i(j_k) = sign(B_i(j_k) - Br_i(j_k)) = \begin{cases} 1, & Br_i(j_k) \geq B_i(j_k) \\ -1, & else \end{cases},$$

and $M_i(j_k)$ is the value $Br_i(j_k) \mathrm{mod}(Q_{QF}(j_k))$ shifted to the range $[-Q_{QF}(j_k)/2, Q_{QF}(j_k)/2 - 1]$. Watermark extraction is carried out by (4).

4. Experimental part

In the experimental research, we implemented and tested the selected embedding methods using different criteria. All experiments were carried out using the images from the University of Waterloo repository [20].

4.1. Efficiency of the embedding methods in JPEG semi-fragile watermarking

The first experiment assesses the efficiency of the considered methods in JPEG-resistant semi-fragile watermarking. In this experiment, we embedded $N_W = 4$ bits into the DCT coefficients in the fixed positions (low, medium and high frequency coefficients were modified). For data embedding, we used $QF = 50$. Then the watermarked images were compressed to JPEG using various quality factors QF^*, both lower and higher than QF.

After that, we extracted the hidden bits from each obtained image and estimated the bit error rate (BER) as:

$$BER = \frac{1}{N \times N_W} \sum_{i=1}^{N} \sum_{k=1}^{N_W} XOR(W_{i,k}, W_{i,k}^R). \quad (10)$$

The results of the experiment averaged by the dataset are presented in Figure 3 and Table 1.

Table 1. Integral BER deviations from theoretical values (after JPEG compression with all possible QF^* values).

Embedding method	err_{FN}	err_{FP}
LSB	8.486	0.042
Preda-QIM	8.543	0.041
Sign-QIM	8.540	**0.037**
DM-QIM	**5.446**	0.066
MOD-QIM	8.282	1.286

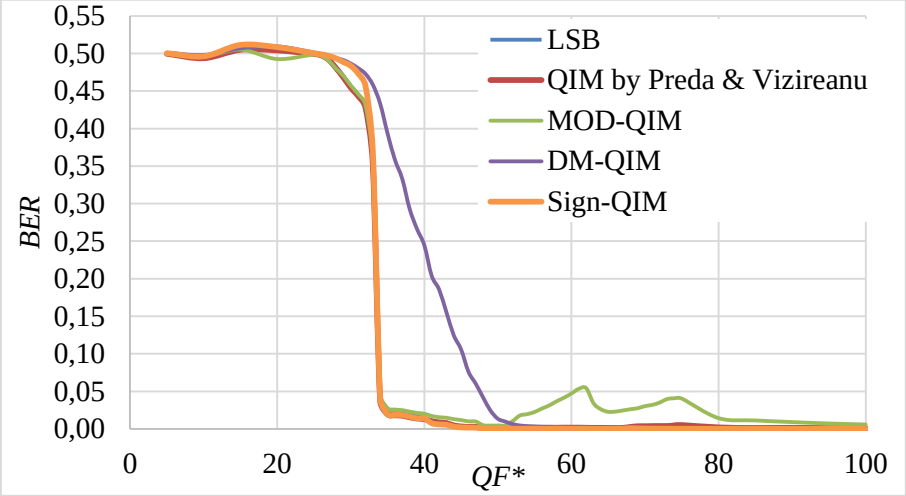

Figure 3. BER after JPEG compressions with different QF^*.

In the ideal case, BER should be close to 0 for $QF^* \geq QF$. For smaller QF^*, BER should be close to 0.5 that corresponds to random guessing. In practice, there is an inevitable transition phase where BER gradually decreases from 0.5 to 0. Figure 3 shows that it is true for all the considered methods. The shorter the transition phase, the better the method.

Besides, there may be nonzero values of BER at $QF^* \geq QF$ due to the rounding of pixel values after the inverse DCT that cause distortions of the spectral components.

To assess the deviation of the obtained curves from the ideal case (a step function), we used the following heuristic measures:

$$err_{FN} = \sum_{QF^*=QF-25}^{QF-1} \left(0.5 - BER\left(QF^*\right)\right), \tag{11}$$

$$err_{FP} = \sum_{QF^*=QF}^{QF+24} BER\left(QF^*\right). \tag{12}$$

These two expressions characterize the integral BER deviation from their theoretical values. The obtained err_{FN} and err_{FP} values are presented in Table 1. The table demonstrates that in terms of err_{FN} measure, DM-QIM considerably outperforms the rest methods. However, DM-QIM provides a high err_{FP} value. LSB, Preda-QIM, and Sign-QIM are very close in err_{FN} values, while MOD-QIM has a large number of errors at $QF^* \geq QF$.

4.2. Investigation of introduced distortion level

In the second experiment, we estimated how the quality of the resulting image depends on the number of embedded bits and the positions of the modified coefficients. For this purpose, we calculated the Peak Signal-to-Noise (PSNR) measure.

The numbers of modified coefficients in each frequency domain were predetermined, but their positions were random. We considered coefficients 2-14 in the zigzag scan as the low frequency domain, 15-35 coefficients as the medium frequency domain, and 36-64 coefficients as the high frequency domain. As in the first experiment, QF was equal to 50. The results of the second experiment are presented in Table 2.

Table 2 shows that LSB, MOD-QIM and Preda-QIM provide quite close quality of the watermarked images. DM-QIM showed the best results. Analysis of various configurations of modified frequency domains showed that it is better to embed information into low frequency components. For instance, if 10 bits are embedded in the low frequency coefficients using DM-QIM,

the quality of the resulting image is higher than if we embed one bit in the high frequency coefficients by any method.

Table 2. Averaged PSNR of watermarked images after watermark embedding by different methods.

Number of bits per block, N_W	Number of modified AC coefficients per domains (LF-MF-HF)	PSNR				
		LSB	Preda-QIM	Sign-QIM	DM-QIM	MOD-QIM
1	1-0-0	43.95	42.92	45.87	47.03	44.74
1	0-1-0	33.85	33.74	34.68	37.72	33.97
1	0-0-1	28.93	28.94	29.23	31.67	28.96
2	2-0-0	41.55	40.55	43.56	44.12	42.17
2	0-2-0	31.84	31.70	32.69	34.68	31.98
2	0-0-2	26.42	26.43	26.74	28.76	26.45
4	4-0-0	38.85	37.87	40.92	41.15	39.42
4	0-4-0	29.32	29.18	30.23	31.71	29.50
4	0-0-4	23.68	23.68	24.02	25.75	23.72
4	1-1-2	25.64	25.62	26.05	28.13	25.70
10	10-0-0	34.90	33.93	36.99	37.22	35.45
10	0-10-0	25.78	25.62	26.69	27.70	25.97
10	0-0-10	20.04	20.03	20.40	21.80	20.09
10	2-3-5	22.12	22.08	22.56	24.15	22.55
10	3-3-4	22.85	22.80	23.31	24.88	22.19
10	2-4-4	22.68	22.62	23.15	24.73	22.76
10	1-3-6	21.49	21.46	21.91	23.50	21.56
Mean		**29.05**	**28.77**	**29.94**	**31.45**	**29.27**

4.3. Investigation of watermark fragility to unacceptable distortions

The considered data embedding methods should be fragile to typical distortions corrupting image content. To verify this property, we performed median filtering and image blurring with a sliding window of size from 3×3 to 15×15, and additive white Gaussian noise with variance values from 400 to 1000. The results are presented in Figures 4, 5, and 6, respectively ($QF = 50$, the number of embedded bits per block $N_W = 4$). Since all these distortions are unacceptable, the relative extraction error (BER) should ideally be close to 0.5, which corresponds to the probability of random guessing of the correct bit.

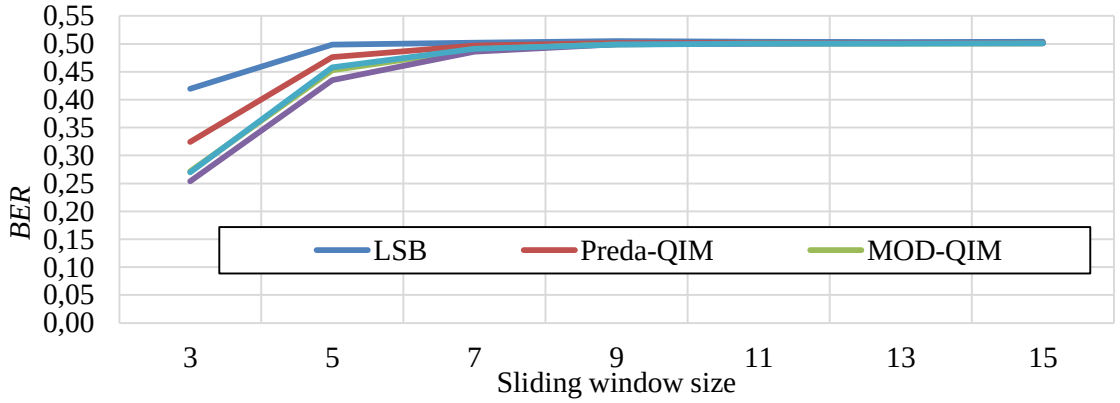

Figure 4. The effect of median filtering on the extraction error.

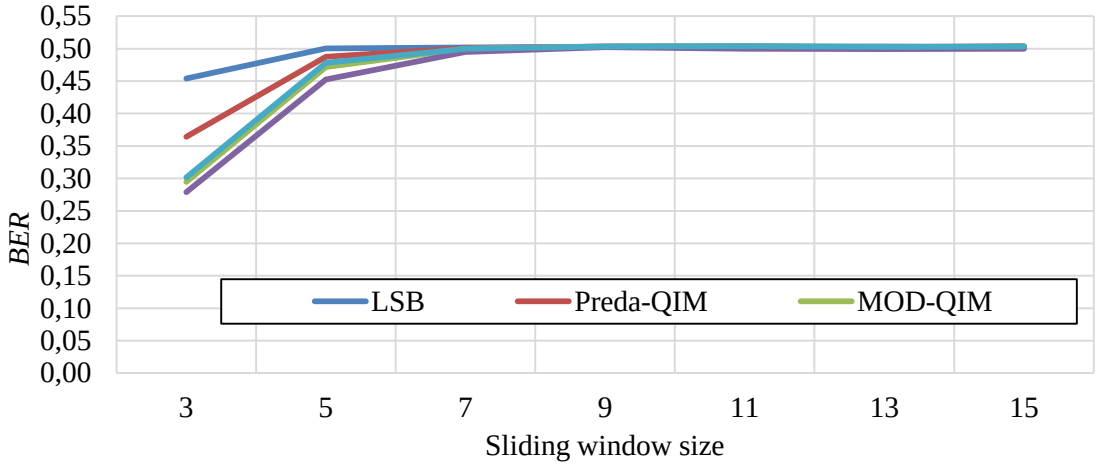

Figure 5. The effect of blur on the extraction error.

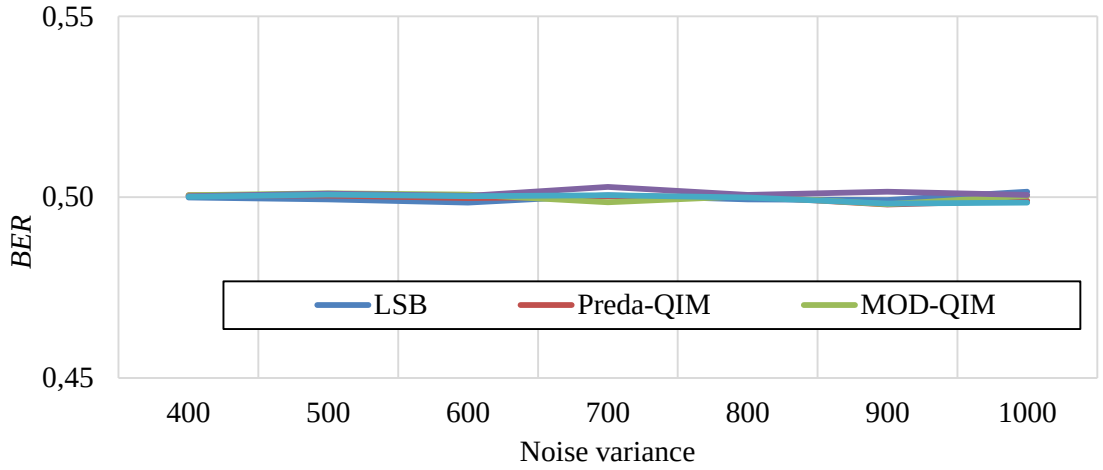

Figure 6. The effect of additive noise on the extraction error.

Figures 4-5 show that LSB slightly outperforms QIM-based methods. However, QIM-based methods also provide the *BER* that exceeds 0.4 after nonlinear and linear filtering even with a window size of 5×5, which is a very good result. With a 3×3 window, the error is also high enough, so the considered methods are fragile to the distortions. After adding noise to the watermarked image, almost all methods behave perfectly.

Thus, according to the experimental results, it can be concluded that the considered data embedding methods are fragile to these three types of distortions.

4.4. Investigation of tampering localization error

Some watermarking systems used for authentication perform content-based watermark generation aimed to raise tampering localization accuracy. One of such systems is proposed in paper [7] Preda & Vizireanu. For each block, it calculates a hash value of a pseudo-random sequence and block coordinates and uses the obtained code as a watermark. This technique protects the image from copy-move attacks. However, in this research, we did not apply any technique improving localization accuracy, because we just aimed to compare the embedding methods.

Therefore, we constructed the watermark in a pseudo-random manner, so the localization error was overestimated. Theoretically, the probability of skipping a distorted block, in this case, should be close to $1/2^N$, where N is the number of bits embedded in each block. Thus, for example, if $N_W = 4$, the percentage of error should be about 6.25%.

The dependence of the fraction of falsely detected blacks on the fraction of tampered blocks is presented in Figure 7. It illustrates that the graphs for different methods, as expected, are very close to

each other and correspond to the theoretical estimation. The only exception is MOD-QIM that provides a high error and does not depend strongly on the number of tampered blocks. Consequently, this method cannot be applied for JPEG semi-fragile watermarking.

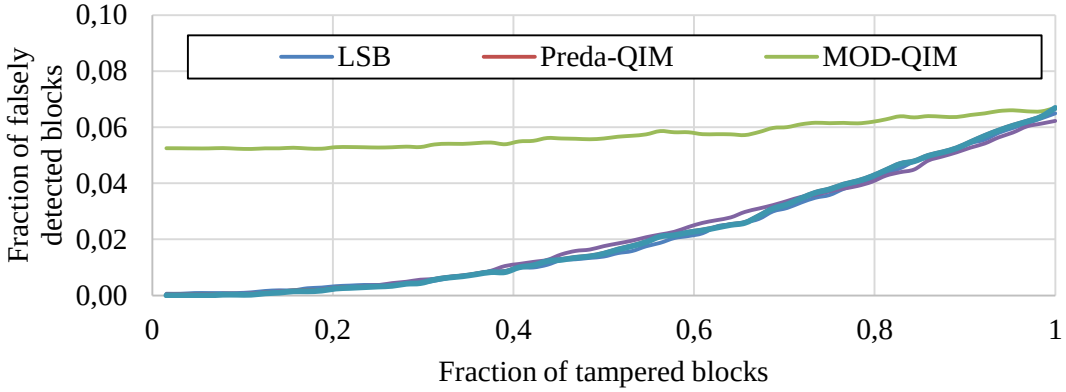

Figure 7. The influence of the number of modified blocks on the tampering localization error (the watermark is randomly generated).

5. Conclusion

In the paper, we investigated the different data embedding methods that are usually used in JPEG semi-fragile watermarking systems such as LSB and various QIM versions (MOD-QIM, Sign-QIM, and DM-QIM). The evaluation of their performance showed that all considered methods could be applied to embed a JPEG semi-fragile digital watermark in the frequency domain apart from MOD-QIM. The study of the quality of the images formed by the watermark embedding process showed the superiority of the DM-QIM method over the others. We also showed that visual distortions of the watermarked image are more visually imperceptible when the watermark is embedded in the low frequency DCT coefficients. As JPEG semi-fragile watermarks must be destroyed with any image modifications, apart from JPEG, we checked the fragility of the considered methods to the distortions: median filtering, blurring and white Gaussian noise. Finally, we carried out that the error of tampering localization coincides with theoretical value for all considered methods excluding MOD-QIM.

6. References

[1] Cox I 2008 *Watermarking and Steganography* (Morgan Kaufmann) p 624
[2] Lin C Y and Chang S F 1999 Issues and solutions for authenticating MPEG video *Proceedings of SPIE* 54-65
[3] Lin C Y and Chang S F 2000 Semifragile watermarking for authenticating JPEG visual content *Security and Watermarking of Multimedia Contents* II **3971** 140-151
[4] Ho C K and Li C T 2004 Semi-fragile watermarking scheme for authentication of JPEG images *ITCC* **1** 7-11
[5] Huang L Y 2013 Authentication watermarking algorithm resisting JPEG compression based on preliminary quantization *Information Technology Journal* **12** 3723-3728
[6] Ye S, Zhou Z, Sun Q, Chang E and Tian Q 2003 A quantization-based image authentication system *Proceedings of the 2003 Joint* **2** 955-959
[7] Preda R O and Vizireanu D N 2015 Watermarking-based image authentication robust to JPEG compression *Electronics Letters* **51** 1873-1875
[8] Wang H, Ho A and Zhao X 2011 Novel fast self-restoration semi-fragile watermarking algorithm for image content authentication resistant to JPEG compression 72-85
[9] Fan C H, Huang H Y and Hsu W H 2011 A Robust Watermarking Technique Resistant JPEG compression *J. Inf. Sci. Eng* **27** 163-180
[10] Fallahpour M and Megias D 2016 Flexible image watermarking in JPEG domain *ISSPIT* 311-316
[11] Mursi M, Assassa G M R, Aboalsamh H and Alghathbar K 2009 A DCT-based secure JPEG image authentication scheme *World Academy of Science, Engineering and Technology* **53** 681-

687
[12] Lin E T, Podilchuk C I and Delp E J 2000 Detection of image alterations using semifragile watermarks *Security and Watermarking of Multimedia Contents* II **3971** 152-164
[13] Al-Mualla M E 2007 Content-adaptive semi-fragile watermarking for image authentication *14th IEEE International Conference on Electronics, Circuits and System* 1256-1259
[14] Wong P H W, Au O C L and Wong J W C 2001 Data hiding technique in JPEG compressed domain *Security and Watermarking of Multimedia Contents* III **4314** 309-320
[15] Xiao J, Ma Z, Lin B, Su J and Wang Y 2010 A semi-fragile watermarking distinguishing JPEG compression and gray-scale-transformation from malicious manipulation *IEEE Youth Conference on Information, Computing and Telecommunications* 202-205
[16] Chen B and Wornell G 2001 Quantization index modulation: a class of provably good methods for digital watermarking and information embedding *IEEE Transaction on Information Theory* **47** 21
[17] Egorova A A and Fedoseev V A 2019 A classification of semi-fragile watermarking systems for JPEG images *Computer Optics* **43** (in print)
[18] Glumov N I and Mitekin V A 2011 A new semi-fragile watermarking algorithm for image authentication and information hiding *Computer Optics* **35(2)** 262-267
[19] Wallace G K 1992 The JPEG still picture compression standard *IEEE Transactions on Consumer Electronics* **38** xviii–xxiv
[20] Image repository *The waterloo fractal coding and analysis group* URL: http://links.uwaterloo.ca/Repository.html

Acknowledgment
The work was partly funded by the Russian Federation Ministry of Science and Higher Education within a state contract with the "Crystallography and Photonics" Research Center of the RAS under agreement 007-ГЗ/Ч3363/26 (in part of JPEG implementation) and by the RFBR grant # 18-71-00052 (in parts of review and investigations).