Semi-Fragile Watermarking Authentication with Local and Global Watermarks

Jinwei Wang[1,2], Shiguo Lian[2], Zhen Ren[2], Yuewei Dai[1]and Zhiquan Wang[1]
1 Department of Automation, Nanjing University of Sci. & Technol.
Nangjing 210094, P.R. China
wjwei_2004@163.com,
2 France Telecom R& D Beijing
Beijing 100080, P.R. China
shiguo.lian@francetelecom.com

Abstract. In this paper, a novel semi-fragile watermarking authentication scheme is proposed, which is based on two watermarks: the local watermark and the global watermark. The two watermarks complement each other. The local watermark is generated from the quantized ACs in 8×8 DCT block. The global watermark is produced by the lowest frequency subband in DWT-transformed image. In watermark embedding, the dither-modulation quantization rule is adopted to improve the scheme's security. In watermark detection and authentication, a novel authentication principle is proposed, which obtains good authentication performances. Experimental results prove that this scheme can exactly detect and verify the tampered location against JPEG compression and some other attacks.

1 Introduction

Till now, many authentication schemes have been proposed, which can be classified into two categories, i.e. fragile authentication [1] and semi-fragile authentication [2,3,4]. For the former, it can not tolerate any possible modification to the multimedia content, e.g., common signal processing operations (JPEG compression, filtering, noise, etc.). Differently, the semi-fragile authentication scheme is rapidly developed and widely used since almost all the applications allow the minor changes to multimedia works if their content could be proved authentic.

Being suitable for practical applications, many semi-fragile authentication schemes have been reported, which can be classified into two types, i.e. the content-independent scheme [5] and content-dependent scheme [4,6,7,8]. In the former scheme, the authentication data are the authentication sequences or logos independent of the multimedia content. For example, a random sequence is embedded into the DWT transform coefficients by quantizing them to integer

Please use the following format when citing this chapter:
Wang, Jinwei, Lian, Shiguo, Ren, Zhen, Dai, Yuewei, Wang, Zhiquan, 2006, in IFIP International Federation for Information Processing, Volume 204, Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 681–688

multiples of a step size [5]. The drawback of this kind of scheme is that the security can not be guaranteed. For attackers, the unwatermarked coefficients can be modified to make authentication out of work, or the watermark can be estimated according to a watermarked image and then embedded into other images. In the content-dependent scheme, the authentication data are extracted from the multimedia content, which are then signed with digital signature or embedded into the multimedia content. This kind of scheme is emphasized in the following content.

The content-dependent scheme can be classified into signature-based scheme [7,8] and content-based watermarking scheme [4,6]. In [7], two labeling methods are proposed, which are based on the second-order image moments and image edges. In [8], the authentication information is formed based on the relationship of the DCT coefficients in 8×8 block pairs using a pre-determined secret mapping function. The clear drawback of a signature-based scheme is that authentication information's transmission or storing requires extra channels and this increases the scheme's danger. In [6], the authentication information described in [8] is inserted into DCT coefficients by the quantization method. As an improvement [4], the authentication information is inserted into DWT coefficients using random bias and non-uniform quantization. Compared with the signature-based scheme, the content-based watermarking scheme not only makes sure that the authentication information is exclusive but also saves extra channels. The drawback is that it changes images' content and degrades images' quality [8]. Generally, the semi-fragile content-based watermarking authentication scheme satisfies the following requirements [6,9]: imperceptibility, obliviousness, robustness, fragileness, location and security.

In this paper, a semi-fragile watermarking authentication scheme is proposed, which satisfies the general requirements. Two complementary feature watermarks, named the local watermark and the global watermark, are generated and embedded, which obtain good performances. The following content is arranged as follows. In Section 2, the architecture of the proposed authentication scheme is presented. The watermark generation is described in detail in Section 3. In Section 4, the embedding rule is proposed. In Section 5, the authentication process is designed. The security of the proposed scheme is analyzed in Section 6. In Section 7, experimental results are demonstrated. The conclusions are drawn and future work is presented in Section 8.

2 Architecture of the Authentication Scheme

The architecture of the proposed authentication scheme is depicted in Figure 1. In embedding process, two feature watermarks, named the local watermark and the global watermark, are generated and embedded into DCT coefficients. In the verification process, the authentication watermark extracted from the received image is compared with the one generated from the received image.

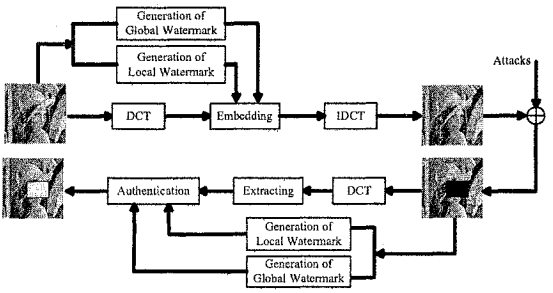

Figure 1. Architecture of the proposed scheme.

3 Generation of the Watermarks

The watermarks are composed of the local watermark and the global watermark. The local watermark is generated from the features of the quantized ACs in 8×8 DCT blocks. This watermark contains the detail information of images, which is sensitive to the changes of such detail components as textures or edges. The global watermark is produced by the features of the lowest frequency subband in the DWT-transformed image. This watermark contains the approximate information of images, which is sensitive to the changes of such approximate components as contours. Both of them will be presented in detail in the following content.

3.1 Extraction of the Local Feature

3.1.1 Pseudo-Random Sequence

Pseudo-random sequence X is generated from a chaotic map, i.e. Logistic map.

$$x_{n+1}=\lambda x_n(1-x_n) \tag{1}$$

The sequence is chaotic when λ is equal to 4. The initial value x_0 is considered as the secret key. Eq. 2 is used to generate the bi-value sequence Y (-1 or 1).

$$Y = \text{sign}(X-0.5) \tag{2}$$

Here, sign(x) is Sign function that gets 1 if x is no less than 0 and -1 if x is less than 0.

3.1.2 Quantization

The original image is transformed using blocked 8×8 DCT, and the transform coefficients are quantized by the quantization matrix Q. Q can adopt the given standard quantization matrix of JPEG, and the quantization and de-quantization rules are shown in Eq. 3 and Eq. 4, respectively.

$$M_k^Q = \text{Round}\left(\frac{F_k(u,v)}{\beta Q(u,v)}\right) \tag{3}$$

$$F_k^Q(u,v) = M_k^Q \cdot \beta Q(u,v) \tag{4}$$

Here $F_k(u,v)$ is the coefficient at position (u,v) in the k-th sub-image, $M_k^Q(u,v)$ is the multiple of quantizing $F_k(u,v)$ by $Q(u,v)$, $F_k^Q(u,v)$ is the quantized coefficient corresponding to $F_k(u,v)$, β is the modulation factor of Q, and Round(·) is to obtain

the integer closest to x. The bigger β is, the less the non-zero quantized transform coefficients (MQTC) are, and the stronger the ability against the noise effect is.

3.2 Generation of the Local Watermark

By Eq. (3), M^Q composed of MQTC is obtained. The sensitivity to the noise can be changed by modulating β, and it is still increased when more MQTCs are selected. Consequently, we select n MQTCs in zigzag order to construct the vector M_k^Q. It is supposed that the original image's size is M×N. Thus, the total number of sub-image is L=(M/8)×(N/8), and the size of M^Q is n×L, that is, $M^Q = [M_0^Q, \cdots, M_k^Q, \cdots, M_{L-1}^Q]'$, k = 0, 1, ···, L-1. Then, by using Logistic map (Eq. (2)), a pseudo-random matrix Y = $[Y_0, \cdots, Y_k, \cdots, Y_{L-1}]$, k = 0, 1, ···, L-1, is generated. Next, the components in M^Q are multiplied by the corresponding one in Y and produce the product module 2, i.e. Eq. (5). Finally, the local watermark W^L is obtained, which is composed of the components W_k^L (k = 0, 1, ···, L-1).

$$W_k^L = (M_k^Q \cdot Y_k) \bmod 2 \tag{5}$$

Five 512×512 images are selected as examples to generate the local watermark. Here, let β and n be 10 and 5 respectively, which obtain strong robustness against the noise. The produced local watermarks of several sample images (Lena, Barbara, baboon, goldhill, and peppers) are shown in Figure 2.

Figure 2. Local watermarks generated from Lena, Barbara, baboon, goldhill and peppers. (from left to right)

3.3 Extraction of the Global Feature

3.3.1 Generation of Dither Quantization Matrix

By Logistic map (Eq. (2)), the integer dither matrix $\boldsymbol{D}$ with the size of M×N is generated, and is then partitioned into 8×8 sub-matrices. Next, every sub-matrix is added to the quantization matrix Q (the standard quantization matrix in JPEG) that is multiplied by α, i.e. a sensitivity factor. The bigger α is, the less sensitive it is to the noise. Finally, the dither quantization matrix Q^d with the size of M×N is obtained.

3.3.2 Extraction of the Global Feature

The original image is transformed using blocked 8×8 DCT, its transform coefficients are quantized by Eq. (3) with the dither quantization matrix Q^d, and the coefficients are then de-quantized by Eq. (4) with the same matrix Q^d. The produced image is named the quantized image. Then, the approximate data LL are extracted after the quantized image is transformed by DWT.

3.4 Generation of the Global Watermark

For most of the coefficients in the lowest frequency band are over 255, the LL coefficients are preprocessed. Then, the preprocessed LL is converted to a binary image by selecting the rational threshold. The ultimate binary image is taken as the global watermark W^G.

The five images mentioned above are taken as examples. Let α be 10, which achieves strong robustness against the noise. Figure 3 shows the produced global watermarks of the five images.

Fig. 3. Global watermarks generated from Lena, Barbara, baboon, goldhill and peppers. (from left to right)

4 Watermark Embedding

The adopted embedding rule is the quantization method [10] whose quantization step size Δ changing with the frequency difference in the embedding region. Here, Δ is a vector with its component as a part of the standard JPEG quantization matrix, and dither modulation (DM) is used to improve the security of watermark embedding.

5 Authentication

In most of the existing papers [5,7], only single feature watermark is embedded into the image. The feature watermark generated from the received image is compared with the one extracted from the received image. As shown in Figure 4, A_1 or A_2 is called verified authentication set, which denotes the set of all the covers that are verified on the basis of the feature watermark, while A is called attack authentication set, which denotes the set of the attacked watermarked covers under the condition of certain feature or some features. Thus, there exist the following questions in the single feature authentication scheme.

1) To obtain small false negative probability of A_1, the false positive probability of A_1 should be increased.
2) To obtain small false positive probability of A_2, the false negative probability of A_2 should be increased.

As can be seen, there are contradictions between the two false probabilities. To compromise between the false negative probability and the false positive probability, we use the multiple-features scheme. It is noted that the false negative probability becomes bigger although the intersection of A_1 and A_2 decreases the false positive probability. Additionally, the false positive probability becomes bigger although the union of A_1 and A_2 decreases the false negative probability. Consequently, the verified authentication set V of the proposed authentication scheme satisfies the following condition.

$$A_1 \cap A_2 < V < A_1 \cup A_2$$

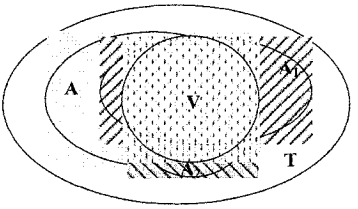

Figure 4. The proposed scheme's authentication set V. A is the ellipse region of horizontal line, A_1 is the ellipse region of backlash, A_2 is the ellipse region of oblique line, V is the circular region of vertical line, and T is the blank ellipse region.

The authentication process is described as follows. First, the local watermark $W^{L'}$ and global watermark $W^{G'}$ are generated from the received image to form a new watermark W'. Then, the watermark $\widehat{W}$ that is composed of $\widehat{W}^L$ and $\widehat{W}^G$ is extracted from the received image. Finally $\widehat{W}$ is compared with W' using XOR operation to implement the authentication, which produces the comparison results $W^{LL'}$ and $W^{GG'}$.

For the local watermark and the global watermark complement each other, the verification result depends on the two watermarks. When satisfying one of the following four conditions, the 8×8 sub-image is marked by 0, i.e. the gray region in Figure 5, which represents the tampered region. Here, $W^{LL'}(i, j)$ corresponds to the sub-image with size of 8×8 at position $(8i, 8j)$, and a gray part represents a sub-image with size of 8×8. Thus, the gray part in Figure 5(a) is marked when Condition (1) is satisfied, the gray part in Figure 5(b) is marked when Condition (2) is satisfied, and the gray part in Figure 5(c) is marked when one of Condition (3) and (4) is satisfied.

Condition (1): $W^{LL'}(i, j) = 1$ and $W^{GG'}(i, j+1) = 1$;
Condition (2): $W^{LL'}(i, j) = 1$ and $W^{GG'}(i+1, j) = 1$;
Condition (3): $W^{LL'}(i, j) = 1$ and $W^{GG'}(i+1, j+1) = 1$;
Condition (4): $W^{LL'}(i+1, j) = 1$ and $W^{GG'}(i, j+1) = 1$.

Figure 5. The marked sub-images. Here, a gray part represents a sub-image with size of 8×8.

6 Security Analysis

It is noted that a good authentication algorithm requires that the embedding process should not perturb the extraction of the feature watermarks, i.e. the exact extraction of the feature watermarks from the original watermarked image should be guaranteed. This property is satisfied in this paper, as shown in Eq. (8) [6,8], which makes the feature watermarks survive the embedding process.

$$((M\times\Delta_1)/\Delta_2)\times\Delta_1=M \quad (8)$$

Here, M is a multiple of the quantization step Δ_1 and $\Delta_2<\Delta_1$. The security represents not only that the watermark to be embedded is exclusive, but also that the watermark extracted from the received image can not be forged. The former emphasizes on the security of the algorithm to extract the feature watermarks, while the latter takes the embedding algorithm's security into consideration. In this paper, the local watermark and the global watermark are generated from the chaotic map controlled by the secret key. If the attacker has not the secret key, the watermark can not be forged. Additionally, the DM-quantization is adopted to embed watermarks, which uses the dither vector and makes it more difficult to extract the watermark from the watermarked image or forge it.

7 Experimental Results

First, we authenticate the modified watermarked images without attacks and check the location of the modified positions. Then, we authenticate the modified watermarked images after JPEG compression with a quality factor of 70 and check the location of the modified positions. Experimental results are shown in Figure 6.

To evaluate the performances of the proposed watermarking scheme, it is compared with the popular Chang's algorithm [6]. P_f denotes the probability that an image block gives an indication of modification but not malicious attacks. The comparative results are listed in Table 1. By observing the comparative results of signal processing attacks, our scheme obviously excels Chang's scheme.

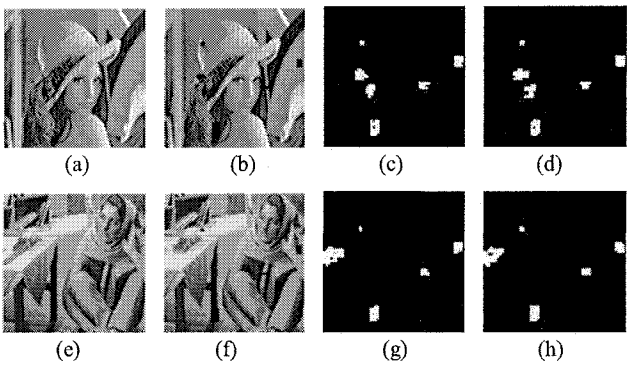

Figure 6. Experimental results. (a), (e): two original images, (b), (f): two modified watermarked image, (c), (g): two authentication results with no attacks, (d), (h): two authentication results after JPEG compression with a quality factor of 70.

Table 1. Comparisons of Two Schemes (Pf %)

Algorithms	No attack	JPEG (QF=70)	Gaussian noise(σ^2=30)	Lowpass filter	Salt-pepper	Sharpen	Histogram equalization
Our	0	0	0	8.5	34.6	58.3	57.4
Chang's	0.2	3.1	12.3	80.4	48.3	90.2	73.8

8 Conclusions

In this paper, a new semi-fragile authentication scheme with two novel feature watermarks is proposed. The local watermark and the global watermark are generated from the content of the image. The DM-quantization embedding rule guarantees the security of the watermarking scheme. The authentication process can prove that the modified content is exactly located. Simultaneously, experimental results show that the verification results of the proposed scheme are valid and satisfactory. In future work, our scheme will be further improved, including more simple and efficient extraction method of the feature watermark and the watermark embedding rule that is robust against more common signal processing attacks.

Acknowledgement

This work was supported by the National Natural Science Foundation of China through the grant number 60374066 and 60574082, and France Telecom R&D Beijing.

References

1. Lin,P.-L., Huang, P.-W., Peng, A.-W.: A Fragile Watermarking Scheme for Image Authentication with Localization and Recovery. Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering, Washington, DC, USA, (2004) 146-153
2. Kundur, D., Hatzinakos, D.: Digital watermarking for telltale tamper-proofing and authentication. Proceedings of the IEEE Special Issue on Identification and Protection of Multimedia Information, Vol. 87, No. 7, (999) 1167-1180
3. Lin, E. T. Podilchuk, C. I., Delp, E. J.: Detection of Image Alterations Using Semi-fragile Watermarks. Proceedings of SPIE in Security and Watermarking of Multimedia Contents II, CA USA, Vol. 3971, (2000) 152-163
4. Maeno, K., Sun, Q., Chang, S.-F., Suto, M.: New Semi-Fragile Image Authentication Watermarking Techniques Using Random Bias and Non-Uniform Quantization. SPIE in Security and Watermarking of Multimedia Contents IV, CA USA, 2002, 4657: 659-670
5. Queluz, M. P.: Spatial Watermark for Image Content Authentication. Journal of Electronic Imaging, Vol. 11, No. 2, (2002) 275-285
6. Lin, C.-Y., Chang, S.-F.: Semi-Fragile Watermarking for Authenticating JPEG Visual Content. Proceedings of SPIE in Security and Watermarking of Multimedia Contents II, CA USA, Vol. 3971, (2000) 140-151
7. Queluz, M.P.: Content-Based Integrity Protection of Digital Images. Part of the IS&T/SPIE Conference on Security and Watermarking of Multimedia Contents, CA USA, Vol. 3657, (1999) 85-93
8. Lin, C.-Y., Chang, S.-F.: A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation. IEEE Transactions on Circuits and Systems of Video Technology, Vol. 11, No. 2, (2001) 153-168
9. Hu, Y.-P., Han, D.-Z.: Using Two Semi-Fragile Watermark for Image Authentication. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou China, Vol. 9, (2005) 5484-5489
10. Chen, B., Wornell, G. W.: Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. IEEE Transaction on Information Theory, Vol. 47, No. 4, (2001) 1423-1443