AN SVD-BASED WATERMARKING METHOD FOR IMAGE CONTENT AUTHENTICATION WITH IMPROVED SECURITY

Seungjae Lee[1]*, Dalwon Jang, and Chang D. Yoo*

[1]Digital Contents Research Division, ETRI, Daejon, Korea
Dept. of EECS, Div. of EE, KAIST, Daejon, Korea
[1]SeungjaeLee@kaist.ac.kr, dal1@kaist.ac.kr, and cdyoo@ee.kaist.ac.kr

ABSTRACT

For image content authentication, a secure watermarking method using quantization-based embedding on the largest singular value (SV) is proposed. The block-wise quantization-based embedding can be vulnerable to vector quantization (VQ) attack and attacks associated with histogram analysis. To overcome these security problems, the proposed method places interdependency among image blocks and dithers the quantized value. By adjusting the threshold of the detector, a trade-off between the robustness to JPEG compression and the probability of miss detection can be made. The proposed method can detect a tampered area with high sensitivity. This is confirmed by experimental results and security analysis.

1. INTRODUCTION

The broader availability of the Internet and various image processing tools opens up, to a greater degree, the possibility of someone downloading an image from the Internet, distorting it, and then distributing it without the permission of the rightful owner. For reason such as this and many more, image authentication has become an active research area.

Depending on the degree of allowable modification, image authentication can be classified into complete authentication and content authentication. The former does not allow any, and the latter does so long as the content is not altered. The latter is implemented by either digital signature based method [1], [2] or watermarking based method [1], [3]-[6].

Previous methods for image content authentication have been proposed in various domains: spatial domain [2], [3], discrete cosine transform (DCT) domain [1], [4], discrete wavelet transform (DWT) domain [5], etc. Wu and Liu [4] proposed the DCT based authentication method using lookup table (LUT). Lin and Chang [1] proposed the self authentication and recovery image (SARI) watermarking system. By using two invariant properties in the DCT domain, SARI not only authenticates but also recovers the modified blocks. Fridrich [3] used quantized projections of image blocks onto smoothed random bases. Kunder and Hatzinakos [5] proposed the DWT-based watermarking algorithm. A watermark is embedded via odd-even quantization of the four-level wavelet coefficients. Xie et. al [2] proposed the approximate image authentication codes which use the most significant bits of an image block as the digital signature. In the method proposed by Sun et. al [6], singular value decomposition(SVD) is performed in the spatial domain, and watermark is embedded by quantizing the largest SV of an image block. Their watermark is robust against JPEG compression and can indicate tampered areas; however, their method is vulnerable to VQ attack and histogram analysis attack, an attack associated with quantization-based embedding.

Although the proposed method is based on the quantization-based embedding, it is free from security problems mentioned above. By placing dependency among randomly chosen blocks and dithering, the proposed method is secure against VQ attack and histogram analysis attack. With improved security, the proposed method satisfies the general requirements for image content authentication system: it can tell the authenticity of an image, even with benign degradation such as JPEG compression, and locate tampered area. The embedded watermark only slightly degrades the image quality.

This paper is organized as follows. Section 2 presents possible attacks on previous content authentication methods. Section 3 explains the proposed method. Section 4 and 5 give the analysis of concerning security and experimental results. Section 6 summarizes.

2. POSSIBLE ATTACKS ON PREVIOUS METHODS

2.1. VQ Attack

Various block-wise independent authentication methods [6], [8] satisfying the locality property have been proposed. However, these are generally vulnerable to VQ attack. If an attacker can form an image database by gathering a number of images of the same size generated by same key, he or she can create a counterfeiting image by replacing an unwatermarked image block by a similar watermarked block obtained from the image database. Although increasing the embedding block size may be a solution to overcome VQ attack, it pays the price for poor locality. Using interdependency among image blocks, the proposed method is robust against VQ attack and does not sacrifice locality.

2.2. Histogram Analysis Attack

Histogram analysis attack is a statistical method for breaking the security of an authentication method. Histogram analysis can reveal vital information that may be used in an attack. For example, an attacker may find out LUT by gathering image data generated from the same LUT [8], and then he or she can modify the image content. In [6], by estimating a quantization step size, the image content can be modified without being detected. Fig. 1 shows an example of such a modified image that is considered authentic.

0-7803-8874-7/05/$20.00 ©2005 IEEE

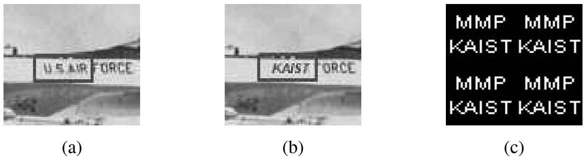

Fig. 1. An example of image content modification using histogram analysis. The original image reads 'U.S AIR', but the modified image reads 'KAIST'. (a) original image (b) modified image (c) authentication result

3. PROPOSED METHOD

For watermark embedding, the proposed method uses quantization-based embedding on the largest SV. To overcome security problems mentioned in Section 2, the proposed method introduces procedures shown in Fig. 2(a). The watermark extraction and authentication of the proposed method is shown in Fig. 2(b). The detailed explanation of each block is given below. The extraction and authentication part is given in Section 3.6.

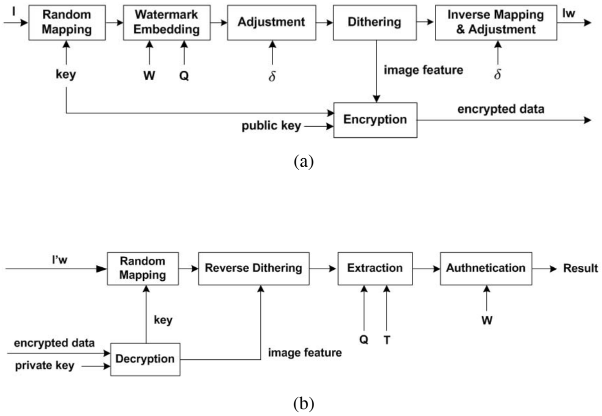

Fig. 2. Proposed image authentication method. W, Q, T and δ represent watermark, quantization step size, threshold, and intended adjustment value, respectively. (a) watermark embedding process (b) watermark extraction and authentication process

3.1. Random Mapping

Previous block-wise independent methods [6], [8] have good localization but are weak against VQ attack. Random Mapping (RM) divides the image that is to be authenticated into 4×4 blocks and randomizes their order in a secret way whose information is locked in a secret key. The embedding procedure which follows is performed in this randomized order.

3.2. Watermark Embedding

Watermark embedding is performed by obtaining the largest SV of an 8×8 image block (neighboring 4 blocks of randomized 4×4 blocks), and then quantizing it using the quantization index modulation (QIM) method [9]. The largest SV of an 8×8 image block is modified with quantization step size Q as follows:

- If watermark bit is 0, the remainder of the largest SV divided by Q is modified to Q/4.
- If watermark bit is 1, the remainder of the largest SV divided by Q is modified to 3Q/4.

3.3. Adjustment of Quantized Value

The pixels values obtained after embedding have floating point values, and rounding these values to the nearest integer will change the quantized largest SV. This can lead to a detection error.

In the adjustment procedure, pixels of an 8×8 image block are modified until the largest SV is within δ of the intended value so that detection error is prevented. If the relationship between the variation of the largest SV and pixel values can be obtained, a fast and effective adjustment is possible.

Consider the SVDs of two M×M image blocks **A** and **B** that can be represented by

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{S}\mathbf{V}^T, \\ \mathbf{B} &= \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^T \end{aligned} \qquad (1)$$

where **U**, $\mathbf{U}_1$, **V** and $\mathbf{V}_1$ are M×M orthogonal matrices, **S** and $\mathbf{S}_1$ are diagonal matrices with the diagonal elements representing the SVs.

For a given small ϵ, $\mathbf{U}_1$ and $\mathbf{V}_1^T$ can be approximated as **U** and **V** respectively when **A** and **B** satisfy the following conditions

$$\max_{i,j}(|a_{ij} - b_{ij}|) \leq \epsilon, \qquad (2)$$

$$\sum_{i=1}^{M}\sum_{j=1}^{M} |a_{ij} - b_{ij}| \leq M^2\epsilon \qquad (3)$$

where a_{ij} and b_{ij} are *ith* row and *jth* column elements of **A** and **B**.

From this, **E**=**B**-**A** $\approx$ **U**($\mathbf{S}_1$-**S**)$\mathbf{V}^T$ and therefore, $\mathbf{S}_1$-**S** $\approx \mathbf{U}^T$ **EV**. Assuming i, jth element of **E**, $|e_{ij}| = 0, 1$, the difference between the largest SVs of **A** and **B** can be controlled by the image error block **E**. We have found that the difference of the largest SV is proportional to the number of 1 or -1 in **E**. The appropriate number of pixels (K) which will be used to adjust the largest SV is determined by the following:

$$K = \left\lfloor \frac{M^2|\Delta\sigma_{(i,j)}|}{T_{(i,j)}} \right\rfloor \qquad (4)$$

where $T_{(i,j)}$ is the largest SV variation by adding value 1 to all pixels in block(i,j), and $\Delta\sigma_{(i,j)}$ is the variation required so that the largest SV obtained by rounding the pixel value is within δ of the intended value in block (i,j).

3.4. Dithering of Quantized Value

Block-wise quantization-based method can be vulnerable to histogram analysis attack as mentioned in Section 2.2. An attacker with the knowledge of the secret mapping key can perform a histogram analysis to figure out the quantization step size, and with it he or she can distort the image without being detected.

By adding image dependent uniformly distributed random noise in the range (-Q/2,Q/2], the proposed method dithers the quantized value, and this procedure can strengthen the security of the proposed method. After dithering, an attacker can not estimate the quantization step size by histogram analysis.

Image feature bits to be used for generating random noise are extracted as follows:

1. After watermark embedding and adjusting the largest SV, image blocks are divided into two disjoint sets **P** and **Q**

$$\begin{aligned}\mathbf{P} &= \{P_1, P_2, P_3, \cdots, P_{N/2}\}, &(5)\\ \mathbf{Q} &= \{Q_1, Q_2, Q_3, \cdots, Q_{N/2}\} &(6)\end{aligned}$$

   where P_i, Q_i and N are *ith* elements of sets P and Q, and the number of 8×8 blocks, respectively.

2. Generate image feature bit from the following equation

$$B(P_i, Q_i) = \begin{cases} 1 & if \quad SV(P_i) \geq SV(Q_i), \\ 0 & if \quad SV(P_i) < SV(Q_i) \end{cases} \tag{7}$$

   where $SV(P_i)$ and $SV(Q_i)$ are the largest SVs.

After generating feature bits, random noises are generated by using them as a key to the random function and are added to the quantized values of the largest SVs to be robust against histogram analysis attack. To prevent any change in feature value due to dithering for the following three cases, $|SV(P_i) - SV(Q_i)|$ must be more than 3Q/2.

- case 1: $|SV(P_i) - SV(Q_i)| \simeq 0$
- case 2: $|SV(P_i) - SV(Q_i)| \simeq Q/2$
- case 3: $|SV(P_i) - SV(Q_i)| \simeq Q$

If the distortion is such that the change in the largest SV is less than 3Q/4, then, undistorted image feature bits can be extracted. However, when image feature bits are distorted, transmission by separate channel must be considered. We have found that JPEG compression up to quality factor 50 does not distort the image feature bits. When the image feature bits and the secret mapping key are transmitted, the public key algorithm is used [10].

3.5. Adjustment of Dithered Value and Reverse Random Mapping

Adjustment due to rounding effect of the pixel value after dithering is performed. After this, the randomly ordered 4×4 blocks are returned to their original to generate watermarked image.

3.6. Watermark Extraction and Image Authentication

Watermark extraction and image authentication is preformed as shown in Fig. 2(b). For extraction, the watermarked image is reordered by RM and dithering is subtracted from the watermarked image by using information obtained from the extracted image feature or from the transmitted data. The watermark is extracted by the following:

$$W(i, j) = \begin{cases} 0 & if \quad z \in (Q/4 - T, Q/4 + T), \\ 1 & if \quad z \in (3Q/4 - T, 3Q/4 + T) \end{cases} \tag{8}$$

where z, Q and T are the remainder of the largest SV of an image block divided by Q, quantization step size, and threshold, respectively.

After extracting the watermark, the result of authentication is performed by comparing the original watermark and the extracted watermark. The extracted watermark varies with T, and the authentication result change accordingly. By adjusting the threshold, a trade-off between the robustness to JPEG compression and the the probability of miss detection can be made.

For given Q and T, the probability of miss detection is $2T/Q$, and the robustness of JPEG compression can be calculated by the watermark bit error rate. Fig. 3 shows that by using smaller threshold, a tampered area can be detected more precisely, and by using larger threshold, the robustness of JPEG compression is improved. The proposed method can use various thresholds that are between δ, value used in adjusting the quantized value, and Q/4 to authenticate an image and locate the tampered area.

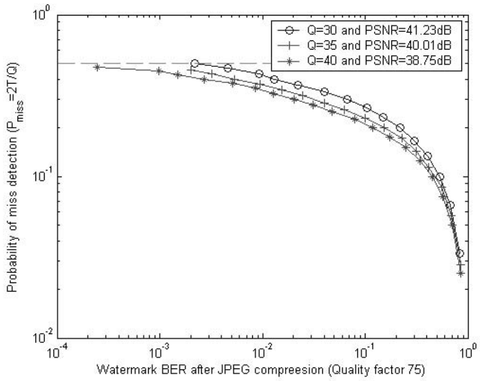

Fig. 3. The relation of the robustness of JPEG compression and the probability of miss detection in F-16 image

4. SECURITY ANALYSIS

By embedding each bit of watermark into 8×8 block composed of four random 4×4 blocks, interdependency among these four blocks is introduced, and this procedure improves the robustness to VQ attack.

If random order is somehow revealed, VQ attack and attacks associated with histogram analysis is possible. If an attacker can gather many images that are of the same size as the image considered and also watermarked with the same secret mapping key, he or she can perform histogram analysis on the largest SV and figure out the composition of each 8×8 block. Although this is a laborious task, it can threaten the security of the watermarked image. However, by dithering the largest SV, the proposed method is safe from histogram analysis attack.

The probability of miss detection depends on T and Q as mentioned in Section 3.6, and its value is 2T/Q. For a given Q, by adjusting T, the proposed method can detect a tampered area with high sensitivity; moreover, the probability of miss detection in neighboring 8×8 block is $(T/Q)^4$. For example, when Q is 40 and T is 0.5, the probability of miss detection in 8×8 block is 0.025^4. Smaller threshold leads to smaller probability of miss detection.

5. EXPERIMENTAL RESULTS

The proposed method was tested by number of different images. For example, results using a 512 by 512 gray scale 'F-16' image are presented. The watermark image is a 64×64 binary logo shown in Fig. 1(c). Using the proposed method, the watermark was embedded into the test image with Q=30 and δ=0.2. The image feature bits were transmitted by separate channel using public key algorithm [10]. In the experiments, four cases were considered: no modification, content modification, JPEG compression, content modification, and content modification after JPEG compression.

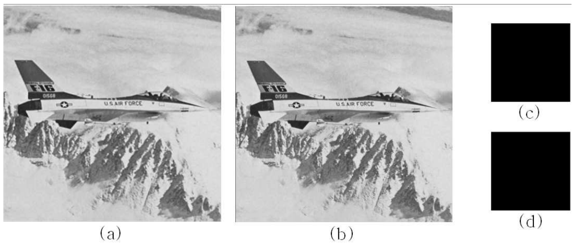

Fig. 4. Result of no modification. (a) watermarked image with Q=30 (PSNR=41.23dB). (b) no modification (c) authentication result with T=7.5 (d) authentication result with T=0.5

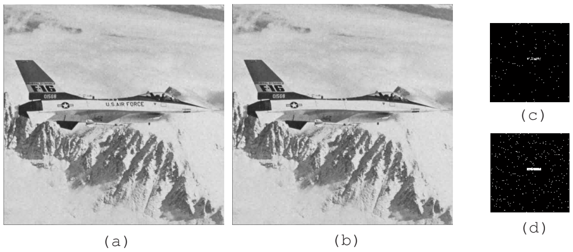

Fig. 5. Result of content modification. (a) watermarked image with Q=30 (PSNR=41.23dB) (b) modified image (c) authentication result with T=7.5 (d) authentication result with T=0.5

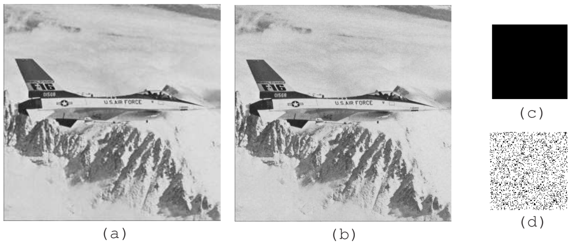

Fig. 6. Result of JPEG compression. (a) watermarked image with Q=30 (PSNR=41.23dB) (b) compressed image with quality factor 80. (c) authentication result with T=7.5 (d) authentication result with T=0.5

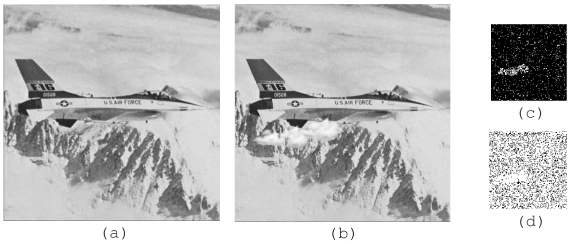

Fig. 7. Result of content modification after JPEG compression. (a) watermarked image with Q=30 (PSNR=41.23dB). (b) content modified and compressed image with quality factor 80. (c) authentication result with T=7.5 (d) authentication result with T=0.5

Result with no modification is shown in Fig. 4. Both authentication results generate no error. Result to content modification is shown in Fig. 5. The letters on the side of the ’F-16’, ‘U. S. AIR FORCE’ have been removed, and the proposed method is able to indicate the exact location of manipulation. In Fig. 6, JPEG compression result is shown. Result with threshold T=7.5 produces no error, but result with threshold T=0.5 generates errors in entire area. Generally, after JPEG compression, authentication errors are not localized in specific area but spread over the entire area. Fig. 7 shows the result of content modification after JPEG compression. Although two manipulations are processed, the proposed method can indicate the tampered area.

6. CONCLUSIONS

In this work, an SVD-based image content authentication method with improved security is proposed. By embedding watermark into randomly ordered block, adjusting and dithering the quantized largest SV of an image block, the proposed method is robust against VQ attack and is safe from histogram analysis attack. With smaller threshold, the probability of miss detection and the sensitivity of the localization property can be improved. Experimental results and security analysis support the improvement of security.

7. ACKNOWLEDGMENTS

This work was supported by the Ministry of information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program & grant No. R01-2003-000-10829-0 from the Basic Research Program of the Korea Science and Engineering Foundation.

8. REFERENCES

[1] C.-Y. Lin and S.-F. Chang, “Semi-fragile watermarking for authenticating JPEG visual content,” in *Proc. SPIE Int. Conf. Security and Watermarking of Multimedia Contents II*, vol.3971, pp. 140-151, 2000.

[2] L. Xie, G. R. Arce, A. Basch, and E. B. Basch, “Image enhancement toward soft image authentication,” in *Proc. IEEE Int. Conf. Mutimedia and Expo (ICME)*, vol.1, pp.497-500, 2000.

[3] J. Fridrich, “Image watermarking for tamper detection,” in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol.2, pp. 404-408, 1998.

[4] M. Wu and B. Liu, “Watermarking for image authenticaion,” in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol.2, pp. 437-441, 2000.

[5] D. Kundur and D. Hatzinakos, “Digital watermarking for telltale tamper proofing and authentication,” *Proc. IEEE*, vol.87, pp. 1167-1180, July 1999.

[6] R. Sun, H. Sun, and T. Yao, “A SVD-and quantization based semi-fragile watermarking for image authentication,” in *Proc. Int. Conf. Signal Processing (ICSP*), vol. 2, pp. 26-30, 2002.

[7] M. Holliman and N. Memon, “ounterfeiting attacks on oblivious block-wise independent invisible watermarking schemes,” *IEEE Trans. Image Processing*, vol.9, pp. 432-441, Mar. 2000.

[8] M. Yeung and F. Mintzer, ”An invisible watermarking technique for image veirification,” in *Proc. Int. Conf. Image Processing (ICIP)*, vol.1, pp.680-683, 1997.

[9] B. Chen and G.W. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423-1443, May 2001.

[10] *Handbook of Applied Cryptography*, CRC, Boca Raton, FL, 1997.