Individual Restoration of Tampered Pixels for Statistical Fragile Watermarking

Maki YOSHIDA†, Member, Kazuya OHKITA†‡, Nonmember, and Toru FUJIWARA††, Fellow

SUMMARY An important issue of fragile watermarking for image is to locate and restore the tampered pixels individually and accurately. This issue is resolved for concentrated tampering. In contrast, for diverse tampering, only localization is realized. This paper presents a restoration method for the most accurate scheme tolerant against diverse tampering. We analyze the error probability and experimentally confirm that the proposed method accurately restores the tampered pixels. We also show two variations based on the fact that the authentication data used for deriving the watermark is a maximum length sequence code.

key words: watermarking, fragile, tampering, detection, restoration, maximum length sequence code

1. Introduction

Malicious tampering of the digital image is becoming more and more pervasive. Fragile watermarking is a method to locate the tampered areas by using embedded data called watermark. Although there are many schemes that have been presented previously, say [1]–[5], [7], [8], [11]–[16], only the schemes in [14], [16] realize pixel-wise localization with high accuracy. The scheme in [16] is further capable of restoring the tampered pixels to the watermarked ones. The pixel-wise localization is important to find the detailed pattern of the tampering. In addition, the restoration of the watermarked image rather than the original one has a great significance for the contents owners in terms of contents protection (keeping the original image secret even from a third party who are given the watermarked one with the secret key).

The scheme in [16] uses a hierarchical mechanism in which an image is divided into small blocks and an embedded watermark is derived from both blocks and pixels. The hierarchical scheme works well as long as the proportion of both tampered pixels and tampered blocks is small (e.g., the proportion of tampered pixels is about 5%). In other words, the hierarchical scheme in [16] is not tolerant against the diverse tampering where a small number of tampered pixels are isolated and scattered over the entire image. On the other hand, the scheme in [14], which introduces a statistical mechanism into fragile watermarking, uses two different distributions corresponding to tampered and original pixels to decide whether each pixel is tampered or not. The error probability is significantly small against any tampering pattern when the proportion of the tampered pixels is small (e.g., 1%). Thus, the next issue to be addressed is to enable the scheme in [14] to recover the tampered pixels.

The objective of this paper is the development of accurate restoration of the tampered pixels for the statistical scheme in [14]. We show that the statistical mechanism for localization also works for restoration. We also present two variations of the statistical scheme based on the fact that the “tailer-made” authentication data used for deriving the watermark in [14] is a maximum length sequence code. These two variations have different parameters from the original one. As a result, one is more tolerant against more extensive tampering while the other is more accurate for a slight modification. Thus, we can use them as the situation demands.

The proposed restoration method uses exhaustive attempts for each pixel. Specifically, for each pixel, it finds the pattern of 5 MSBs most consisting with the embedded watermark. The original pattern only obeys the corresponding distribution whereas the other patterns, considered as tampered, obey the different distribution. Thus, the original pattern should be distinguished from the others as long as localization is accurate. We analyze the error probability of this restoration and experimentally confirm that the tampered image is exactly recovered if the proportion of the tampered pixels is small in so far as the localization is accurate. The computational time is about 9 seconds for an image sized 512 × 512 pixels in a standard environment.

The rest of this paper is organized as follows. In Sect. 2, we recall the statistical scheme in [14] and the statistical mechanism for localization. In Sect. 3, we propose a restoration method based on the localization mechanism. In Sect. 4, we show two variations of the statistical scheme and compare the performance with the original one. Concluding remarks are given in Sect. 5.

2. Statistical Scheme in [14]

The statistical scheme in [14] assumes that an attacker alters the gray values of some pixels without changing the image size. It aims to check whether the 5 most significant bits (MSBs) of the gray value of each pixel is altered or not by using the watermark embedded into the 3 least significant
bits (LSBs) of the all pixels.

In the following, let \( N \) be the number of pixels contained in an image, and \( p_i \in [0, 255] \) with \( 1 \leq i \leq N \) their gray values. For non-negative integers \( p \) and \( u \) with \( 1 \leq u \leq \lceil \log_2 p \rceil \), let \( B(p, u) \) denote the \( u \)-th bit in the binary representation of \( p \), where \( B(p, \lceil \log_2 p \rceil) \) is the MSB and \( B(p, 1) \) is the LSB.

2.1 Embedding Algorithm

Given a host image and a secret key, the embedding algorithm first generates a number of authentication bits from 5 MSBs of the gray value of each host pixel, and then embeds a folded version of the authentication data and some additional test data into the remaining 3 LSBs of the entire image as follows.

1. For the gray value \( p_i \) of each pixel, generate 31 authentication bits \( b_{i, \ell} \) with \( 1 \leq \ell \leq 31 \) as follows.

\[
b_{i, \ell} = \sum_{u=1}^{5} B(p_i, u + 3) \cdot B(\ell, u) \mod 2 \tag{1}
\]

2. Pseudo-randomly divide \( 31 \cdot N \) authentication bits into \( 31 \cdot N/11 \) subsets according to the secret key such that each subset contains 11 bits of different 11 pixels (i.e., 31 authentication bits are distributed to 31 distinct subsets). Then, calculate XOR of the 11 authentication bits in each subset. The \( 31 \cdot N/11 \) folding XORs are called the sum-bits.

3. Pseudo-randomly generate \( 2 \cdot N/11 \) bits, called the test-bits, according to the secret key. The watermark bits are made up of the sum-bits and test-bits.

4. Permute the \( 3 \cdot N \) watermark bits in a pseudo-random way determined by the secret key, and replace the 3 LSBs of all pixels with them.

In [14], it is shown that PSNR of a watermarked image is approximately 37.9 dB under the assumption that the original distribution of the 3 LSBs is uniform. This PSNR value is higher than the general expectation for acceptable images and the value of IHC evaluation criteria ver.3, which are 35 dB [10] and 30 dB [18], respectively.

2.2 Effect of Tampering

In [14], it is proved that any alteration on 5 MSBs of each pixel will result in the change of 16 authentication bits. The 16 changed authentication bits correspond to 16 subsets. For a subset, its sum-bit will be changed if the number of changed authentication bits is odd. Thus, the corresponding 16 sum-bits can be changed. For a pixel with 5MSBs un-tampered (resp. a pixel with at least one MSB altered), let \( k_U \) (resp. \( k_T \)) denote the number of sum-bits of a pixel that do not equal their corresponding LSBs. The peak of the distribution of \( k_U \) is near zero whereas that of \( k_T \) is at 16. This difference enables the accurate identification of the pixels that have at least one MSB being altered.

We overview the effect of tampering to the watermarked image analyzed in [14]. Denote the ratio between the number of pixels with at least one MSB altered and the image size \( N \) as \( r_M \), and denote the ratio between the number of LSBs that have been changed and the number of all LSBs \( 3 \cdot N \) as \( r_L \).

The probability of a sum-bit being changed is

\[
e_U = \sum_{1 \leq v \leq 11} \binom{10}{v} \cdot r_M^v \cdot (1 - r_M)^{10-v},
\]

The probability of a sum-bit being different from the LSB at the corresponding position is

\[
e_T = \sum_{0 \leq v \leq 10} \binom{10}{v} \cdot r_M^v \cdot (1 - r_M)^{10-v},
\]

and \( E_\delta = e_\delta \cdot (1 - r_L) + (1 - e_\delta) \cdot r_L \) with \( \delta \in \{U, T\} \). The distributions of \( k_U \) and \( k_T \) are given by

\[
P(k_U = k) = \binom{31}{k} \cdot E_U^k \cdot (1 - E_U)^{31-k},
\]

\[
P(k_T = k) = \sum_{v=\max(0, k-16)}^{\min(15, k)} \binom{15}{v} \cdot E_U^v \cdot (1 - E_U)^{15-v} \cdot E_T^{k-v} \cdot (1 - E_T)^{16-k+v},
\]

for \( 0 \leq k \leq 31 \).

2.3 Localization Algorithm

In the localization algorithm, the test data is used to estimate the modification strength \( r_M \) and \( r_L \) while the authentication bits are used to reveal the trace of any alteration. From the estimation and the trace of alteration, the localization algorithm statistically judges whether 5 MSBs of each pixel are altered or not.

Specifically, the localization algorithm takes as input a tampered image and the corresponding secret key, and localizes pixels with tampered 5 MSBs as follows.

1. For a given image, calculate the \( 31 \cdot N/11 \) sum-bits according to its 5 MSBs in the way as given in the embedding algorithm, and generate the same \( 2 \cdot N/11 \) test-bits.
According to the secret key.

2. Compare the calculated sum-bits and the generated test-bits with the LSBs at their corresponding positions. Regard the ratio between the number of different test-bits and \(2 \cdot N/11\) as an estimate of \(r_L\), and the ratio between the number of different sum-bits and \((31 \cdot N/11)\) as an estimate of \(E\). According to Eqs. (2) and (3), obtain an estimate of \(r_M\) numerically. With the estimates of \(r_L\) and \(r_M\), obtain the distributions of \(k_U\) and \(k_T\) from Eqs. (4) and (5), respectively.

3. For each pixel, examine its 31 corresponding sum-bits, and count the number of sum-bits being different from their corresponding LSBs, denoted \(k\). If

\[
(1 - r_M) \cdot P(k_U = k) < r_M \cdot P(k_T = k),
\]

then this pixel is judged as a tampered pixel, indicating that there is alteration on its 5 MSBs.

2.4 Accuracy of Localization

Equation (6) is a MAP criterion that minimizes the total number of false decisions [14]. There are two types of false decisions. Let \(N_{\text{MSB,fp}}\) be the number of pixels falsely judged as tampered (called false positive) and \(P_{\text{MSB,fn}}\) the number of pixels falsely judged as untampered (called false negative). Denote the total number of false decisions as \(N_{\text{total}}\) (i.e., \(N_{\text{total}} = N_{\text{MSB,fp}} + N_{\text{MSB,fn}}\)). Let \(\theta\) be a threshold of \(k\) at which the two curves \((1 - r_M) \cdot P(k_U = k)\) and \(r_M \cdot P(k_T = k)\) intersect. If \(r_L\) and \(r_M\) are correctly estimated, the expectations of \(N_{\text{MSB,fp}}\) and \(N_{\text{MSB,fn}}\) are given by

\[
E(N_{\text{MSB,fp}}) = N \cdot (1 - r_M) \cdot \sum_{v=0}^{31} P(k_U = v),
\]

\[
E(N_{\text{MSB,fn}}) = N \cdot r_M \cdot \sum_{v=0}^{31} P(k_T = v).
\]

3. Proposed Restoration Method

In this section, we present a restoration algorithm, analyze the number of pixels falsely restored, and experimentally confirm that the tampered pixels are exactly restored.

3.1 Restoration Algorithm

The proposed algorithm aims to restore the 5 MSBs of each pixel with “tampered” judgement. The number of possible patterns of 5 MSBs is 32. We find the pattern for which the sum-bits are most consistent with the LSBs at the corresponding positions. The unmatched number \(k\) for the original pattern only obeys the distribution of \(k_U\). From the difference between the distributions of \(k_U\) and \(k_T\), the pattern with the smallest \(k\) is the original pattern with high probability. In this way, the statistical mechanism for localization also works for restoration.

The proposed algorithm restores the 5 MSBs of each pixel with “tampered” judgement as follows.

1. For each pattern of the 5 MSBs, calculate the sum-bits according to the pattern, count the number of sum-bits being different from the corresponding LSBs, denoted \(k\) same as in the localization algorithm.

2. Replace the 5 MSBs to the pattern with the smallest \(k\).

3.2 Accuracy of Restoration

If there are other patterns whose \(k\) is equal or smaller than that of the original pattern, restoration will be failed. So, the probability of failure is

\[
p_R = \sum_{i=0}^{31} P(k_U = i) \cdot \left[1 - \left(\sum_{j=1}^{31} P(k_T = j)\right)^{31}\right].
\]

Denoting the number of the tampered pixels labeled “tampered” but being unable to be restored as \(N_R\), its average is

\[
E(N_R) = p_R \cdot (N \cdot r_M - E(N_{\text{MSB,fn}})).
\]

If \(N_R = 0\), the original 5 MSBs of all localized pixels can be obtained.

3.3 Experimental Results

We first confirm that the proposed algorithm completely restores a tampered image for diverse and concentrated tampering with a small number of the tampered pixels by using test image Lena sized 512 × 512 as the host image. Figure 1 shows Lena and the watermarked image. Figure 2 shows a concentrated tampering case, in which a flower was planted on the girl’s hat with the number of the tampered pixels \(N_T = 2461\) and the ratio of the tampered pixels \(r_M = 0.9\%\). Figure 3 also shows a diverse tampering case, in which a small black box is scattered over the entire image with \(N_T = 2000\) and \(r_M = 0.75\%\). In both cases, the tampered pixels are completely restored.

We also confirm the accuracy for the different numbers of the tampered pixels \(N_T\) for for Baboon, Peppers, and Airplane (see Fig. 4) in addition to Lena where the same diverse tampering patterns are used (a horizontal-striped box sized \(N_T\) pixels is placed in the central part as shown in Fig. 5).
Table 1 lists the actual values of $N_R$. The results are similar among these images. Specifically, $N_R$ is enough small for small $N_T$ (i.e., small $r_M$ and $r_L$ like 1.5% and 0.60%, respectively), and increases with the value of $N_T$ (i.e., those of $r_M$ and $r_L$). That is, the accuracy depends on $r_M$ and $r_L$ regardless of images.

In all cases, the computational time is about 9 seconds on the following condition: CPU is Intel(R) Core(TM)2 Quad CPU Q6000 2.30GHz 2.39GHz, the memory is 2GB, and OS is Windows XP Professional Service Pack 3, the C compiler is gcc 4.3.2.

### 4. Two Variations of the Statistical Scheme

In this section, we first show that the authentication data in the statistical scheme is the $(31, 5, 16)$-maximum length sequence (MLS) code, which is the dual of the Hamming code. We then present two variations with $(15, 4, 8)$ and $(63, 6, 32)$ MLS codes. We also compares the accuracy among these three variations.

#### 4.1 Useful Property of MLS Codes

From Eq. (1), it can be easily seen that the vector of 31 authentication bits $(b_{i,1}, \ldots, b_{i,31})$ is obtained by multiplying the vector of 5MSBs $(B(p_i, 4), \ldots, B(p_i, 8))$ to the matrix whose $t$-th column is the binary representation of $t$. That is,

$$(b_{i,1}, \ldots, b_{i,31}) = (B(p_i, 4), \ldots, B(p_i, 8)) \begin{bmatrix} 1 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$ 

This matrix is the generator matrix of the $(31, 5, 16)$-MLS code, which encodes 5 information bits to the corresponding codeword of length 31.

The common property of the MLS codes useful for the statistical scheme is that the distance of every two codewords is the same large number (in the above case, the distance is 16).

To keep the quality of watermarked image and the accuracy of localization and restoration high, we use $(2^4 - 1, 4, 2^3)$ and $(2^6 - 1, 6, 2^5)$-MLS codes where the 4 and 3 LSBs are respectively replaced with the watermark bits.
4.2 Proposed Variations with \((2^m - 1, m, 2^{m-1})\)-MLS Code

Variation for \(m = 4\): We straightforwardly use the \((15, 4, 8)\)-MLS code. The \(15\) authentication bits are derived from the \(4\) MSBs of each pixel value, and the \(15 \cdot N\) authentication bits are folded into \(3 \cdot N\) sum-bits (each subset contains \(5\) authentication bits). The watermark bits consist of the \(3 \cdot N\) sum-bits and pseudo-randomly generated \(N\) test bits, and are embedded into the remaining \(4\) LSBs of all pixels.

Under the assumption that the original distribution of the \(4\) LSBs is uniform, the average energy of distortion caused by watermarking on each pixel is

\[
E_D = \frac{1}{256} \cdot \sum_{i=0}^{15} \sum_{v=0}^{15} (u - v)^2.
\]

So, PSNR is approximately

\[
\text{PSNR} \approx 10 \cdot \log_{10} \left( \frac{255^2}{E_D} \right) = 31.8 \text{(dB)}.
\]

Variation for \(m = 6\): The straightforward usage of the \((63, 6, 32)\)-MLS code is to derive the \(63\) authentication bits from the \(6\) MSBs and use the remaining \(2\) LSBs of all pixels for the embedding space (each subset contains \(32\) authentication bits). However, this usage makes the number of false decisions extremely large because the folded number of authentication bits is large.

To solve this problem, we derive the \(63\) authentication bits from the 5MSBs and use the remaining \(2\) LSBs of all pixels for the embedding space. Specifically, we add the bit “0” to the 5 MSBs of each pixel and use the resulting 6 bits for the \((63, 6, 32)\)-MLS code to derive the \(63\) authentication bits. The \(63 \cdot N\) authentication bits are folded into \(63 \cdot N/22\) sum-bits (each subset contains \(22\) authentication bits), and the watermark bits consist of these sum-bits and pseudo-randomly generated \(3 \cdot N/22\) test bits. The watermark bits are embedded into the \(3\) LSBs of all pixels.

In this variation, PSNR of a watermarked image is the same as the original statistical scheme (approximately 37.9 dB).

4.3 Effects of Tampering

We analyze the effects of tampering to the watermarked image for \((2^m - 1, m, 2^{m-1})\)-MLS code with \(m = 4, 6\) similar to \(m = 5\) (the original statistical scheme).

Let \(\sigma^{(m)}\) denote the number of the authentication bits contained in each subset, (i.e., \(\sigma^{(4)} = 5\) and \(\sigma^{(6)} = 22\)). Let \(r^{(m)} = \frac{2^m - 1}{2^{m-1}} \cdot r_M\), denoting the probability that an authentication bit is changed.

The probability of a sum-bit being changed is

\[
e^{(m)} = \sum_{1 \leq v < \sigma^{(m)}} \left( r^{(m)} \right)^v \cdot \left( 1 - r^{(m)} \right)^{\sigma^{(m)} - v}.
\]

Similar to \(e_U^{(m)}\) and \(e_T^{(m)}\), it follows that

\[
e_U^{(m)} = \sum_{1 \leq v < \sigma^{(m)}} \left( r^{(m)} \right)^v \frac{1}{\sigma^{(m)} - v} \left( 1 - r^{(m)} \right)^{\sigma^{(m)} - v},
\]

\[
e_T^{(m)} = \sum_{0 \leq v < \sigma^{(m)}} \left( r^{(m)} \right)^v \frac{1}{\sigma^{(m)} - v} \left( 1 - r^{(m)} \right)^{\sigma^{(m)} - v}.
\]

The expectations of \(N_{fp}^{(m)}\), \(N_{fn}^{(m)}\), and \(N_{total}^{(m)}\) are given by

\[
E(N_{fp}^{(m)}) = N \cdot (1 - r_M) \cdot \sum_{v=0}^{2^{m-1}-1} P(k^{(m)} = v),
\]

\[
E(N_{fn}^{(m)}) = N \cdot r_M \cdot \sum_{v=0}^{[\sigma^{(m)}]} P(k^{(m)} = v),
\]

\[
E(N_{total}^{(m)}) = E(N_{fp}^{(m)}) + E(N_{fn}^{(m)}).
\]

4.4 Comparison

Using the image Lena as the host, we compare the quality of the watermarked images and the number of false decisions in localization. The similar results are provided for the experimental values and those in restoration, and omitted here.

As shown in Fig. 6, the image quality for \((15, 4, 8)\)-MLS code is more rapid starting from a smaller number whereas the growth rate of the image quality in localization. The similar results are provided for the experimental values and those in restoration, and omitted here.
note that the curves given by Eq. (7) for any MLS code with the same parameter \((2^m - 1, m_2^m - 1)\) is the same because we use the common property that the distance of every two codewords is the \(2^m - 1\). In other words, to improve the accuracy, modification of watermark generation is more important than selection of the used MLS code like the variation for \(m = 6\).

The accuracy of localization depends on the difference between the distributions of \(k_T^{m}(i)\) and \(k_U^{m}(i)\) corresponding to tampered and untampered pixels. For a large number of tampered pixels, the effects of other tampered pixels becomes large. Thus, the number of folded authentication bits should be small. Thus, for a large number of tampered pixels \(N_T > 5139\), the \((15, 4, 8)\) MLS code is better than the \((31, 5, 16)\) MLS code. In contrast, for a small number of tampered pixels, the effects of other tampered pixels is small and a larger distance of the codewords makes the two distribution significantly different. Thus, for a small number of tampered pixels \(N_T < 4202\), the use of the \((63, 6, 32)\) code is more effective. Taking for instance of the situation where slightly modifying direction of eyes of a person in pictures will spoil the evidence that she/he looks at something/someone, false decision is critical and the \((63, 6, 32)\) code works better.

5. Conclusion

We have proposed a restoration method for the statistical fragile watermarking scheme in [14]. We have confirmed that the proposed method exactly restore the tampered pixels regardless of the tampering pattern (concentrated or diverse) if the proportion of the tampered pixels is small. We have also presented two variations of the statistical scheme which use the \((15, 4, 8)\) and \((63, 6, 15)\) MLS codes instead of the \((31, 5, 16)\) code. The suitable situations of these variations are different. Thus, we can use them as the situation demands.

References

Maki Yoshida received the M.E. and Ph.D. degrees in Information and Computer Sciences from Osaka University in 1998 and 2001, respectively. In 2001–2013, she was an Assistant Professor at Osaka University. In 2013, she joined National Institute of Information and Communications Technology (NICT). Her current research interests include cryptography, digital watermarking, and formal verification of cryptographic protocols.

Kazuya Ohkita received the M.E. degree in Information and Computer Sciences from Osaka University in 2010. His research interest is digital watermarking.

Toru Fujiwara received the B.E., M.E., and Ph.D. degrees in Information and Computer Sciences from Osaka University in 1981, 1983, and 1986, respectively. In 1986, he joined the faculty of Osaka University. Since 1997, he has been a Professor at Osaka University. He is currently with the Graduate School of Information Science and Technology. His current research interests include coding theory and information security.