DIGITAL WATERMARKING SCHEME FOR COPYRIGHT PROTECTION AND TAMPERING DETECTION

Ching-Sheng Hsu, Shu-Fen Tu

Department of Information Management
Ming Chuan University, Chinese Culture University
e-mails: cshsu@mail.mcu.edu.tw, dsf3@ulive.pccu.edu.tw
Taiwan

Abstract: Digital watermarking scheme is a common way for protecting copyright or verifying integrity for digital images. The working domain of the watermark is either spatial or frequency. Recently, some researches tried to work on another domain which is derived from matrix factorization. Depending on the purpose, either robust or fragile watermark is embedded into the host image. The aim of the proposed scheme in this paper is to embed robust and fragile watermark into the host image at the same time, so that the proposed scheme can possess dual functions: one is copyright protection, and the other is tampering detection. Finally, we provide some experiments to show the performance of the proposed scheme.

Key words: Digital watermark, QR decomposition, Robustness, Tampering detection.

1. INTRODUCTION

The development of the Internet has made the transmission of digital images faster and more convenient. However, some related issues have followed. For example, a person who is interested may tamper with the image content for deception purposes. Sometimes the created images may be stolen without the permission of the author. In view of this, the tamper detection side and copyright protection of digital images have become important issues. Digital watermarking is a commonly used method to protect digital images.

Digital watermark refers to a piece of information embedded in a digital image. A digital watermarking scheme consists of two phases: one is the embedding of the watermark; the other is the extraction of the watermark [8]. Digital watermarks can be divided into robust watermarks and fragile watermarks depending on the purpose of use. Robust watermarking is used to protect the copyright of digital images. Authors of digital images can choose an image representing a personal logo as a

watermark. When the copyright of a digital image is suspected of being violated, the author can extract the watermark from the image to prove his copyright. The basic requirement of the robust watermarking scheme is that after the partial image is edited, the extracted watermark must still have enough recognition to prove the copyright of the image. Fragile watermarking is used to protect the integrity of digital images. When digital image content is tampered with, the fragile watermark must be able to detect the part of the image that has been tampered with. Fragile watermarks are usually composed of features of the image so that the watermark is sufficiently sensitive to image tampering.

Generally speaking, when someone directly uses the image without the authorization of the original author, the person usually makes some modifications to the image. By doing so, the person hopes to evade the copyright controversy; furthermore, the person hopes the modification can destroy the robust watermark in the image. If the watermark is robust enough, the author can extract the robust watermark to prove the copyright. Let us think about a situation. Supposed that the person who steals the image is not for copyright infringement, and that his purpose of modifying the image is not to destroy the robust watermark in the image. but to intentionally add some fake content to the image and use it to frame the author. This person may add some fake content to the image to frame the original author. The author will argue that the content of these frauds is not from himself, but that someone has maliciously framed him. However, if the original author is asked to extract the robust watermark, the robust watermark will make the original author unable to prove his innocence. If the image contains a fragile watermark, the original author can prove his innocence by fragile watermarking. Because the fragile watermark can help the original author to point out that these images have been tampered with, the content of the fake is not the original. Therefore, a scheme with both robust and fragile watermarking is necessary. However, few studies have proposed a dual-purpose digital watermarking scheme. The purpose of this paper is to expand the previously proposed robust watermarking scheme into a dual-purpose watermarking scheme. The previous proposed watermarking method has a good robustness against cropping attacks. If we can combine the fragile watermark to the previous method, we can avoid the situation mentioned above that is maliciously framed by others.

The following sections are arranged as follows. The second section will discuss the relevant research. The third section explains in detail the practice of the proposed scheme. The fourth section is the experimental results and some discussions of this study. Finally, the fifth section is the conclusion of this study.

2. RELATED WORKS

2.1. QR decomposition

Because a digital image can be seen as a matrix, recently, some researchers employed matrix factorization, such as singular value decomposition (SVD) [1-2, 7] or QR decomposition [3, 5, 6, 9-11], to design their digital watermarking schemes. The main idea is to utilize matrix factorization to decompose the image and embed the watermark into one of the decomposed matrices. Because the computational complexity of SVD is higher than that of QR decomposition [10], QR decomposition is a more common method.

The QR decomposition can factor a $m \times n$ matrix A into a $m \times m$ orthogonal matrix Q and a $m \times n$ upper triangular matrix R, where

$$A = QR.$$

If the columns of A have correlation with each other, the absolute values of the elements of the first row of R are probably greater than those of the other rows. Since there usually exits correlations between adjacent pixels of an image, QR decomposition was common in researches related to image processing [3, 8-11]. There are some algorithms for computing the QR decomposition, such as Gram–Schmidt process, Householder transformations, or Givens rotations.

2.2. QR-based watermarking scheme

In our previous work, the host image is divided into 4×4 blocks, and each block is decomposed via QR factorization. The robust watermark is embedded in the first row of R matrix. In the proposed scheme, watermark bit 1 and bit 0 were represented using $f(x) \geq 0$ and $f(x) < 0$, respectively, and function f is defined as follows:

$$f(x) = sin\left(x \cdot K \cdot \frac{\pi}{180}\right), \quad (1)$$

where $K \geq 0$ is a real number. The wavelength λ of the function f is $(360/K)$, and thus, parameter K is a wavelength controller. For any element x_0 to embed the watermark bit, if the state of $f(x_0)$ is not match the watermark bit, x_0 is modified to x_0' such that $|f(x_0')| = 1$ and the difference between x_0 and x_0' is as small as possible. The R matrix with the robust watermark is merged with the original Q matrix, and the digital image is inverted. The main concern of a QR-based digital watermarking scheme is that the modification to the decomposed matrix may cause the pixel values out of range. Moreover, the allowable bounds vary from pixels to pixels. Nevertheless, few of papers gave a discussion about the allowable bounds of the modification. To cope with this issue, our previous work built several rules to define the allowable bound before modifying the elements of matrix R. Therefore, the watermark can be embedded to achieve the possible minimal modification within the allowable bounds.

3. THE PROPOSED SCHEME

3.1. Watermark embedding

Figure 1 illustrates the whole process of watermark embedding. Suppose that H is a gray-level image and RW is a binary image, which represents the robust watermark. At first, H is divided into nonoverlapping blocks of 4×4 sizes. Each block A is factorized to matrix Q and R via QR decomposition. Each four bits of RW are embedded into the four elements of the first row of R, respectively. Then, multiply Q and the modified R' to obtain the watermarked block A'. Repeat the above process to embed the watermark bits into the blocks. Because the focus of this study is to incorporate fragile watermarks into our previous methods, we will explain the method of embedding fragile watermark in detail. As for the method of hiding the robust watermark, readers can refer to our previous research for the method of embedding the robust watermark [5]. It should be noted that because the space of the lowest bit plane of the image is used to embed the fragile watermark, the lowest bits of each image block is not included in the QR decomposition operation.

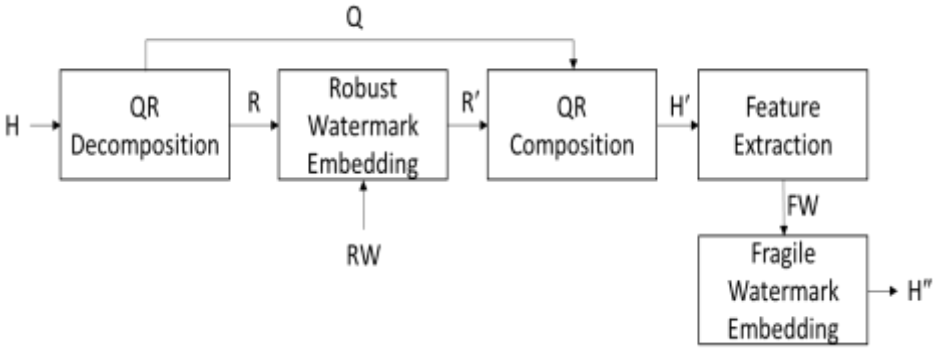

Fig. 1. The process of watermark embedding

Suppose that H' is the image with the robust watermark. At first, H' is divided into nonoverlapping blocks of 4×4 sizes. Supposed that p_i denotes the pixel value of a block B of H', where $i = 0..15$, and supposed that (x, y) denotes the coordinate of B in H'. For each block B, perform the following steps:

Step 1: $p_i' = p_i >> 2$, where $i = 0..15$.

Step 2: Calculate MD5(x, y, p_0', …, p_{15}') to generate 128-bit digest (d_0, d_1, …, d_{127}).

Step 3: $D_j = (d_{16\times j}, d_{16\times j}, \ldots, d_{16\times j+15})$, where $j = 0..7$. Perform XOR operation on D_0 to D_7 to get 16-bit authentication message $A = (a_0, a_1, \ldots, a_{15})$, which represents the fragile watermark FW.

Step 4: Set $p_i = (p_i >> 2) + a_i$ to watermark the block B with authentication message, where $i = 0..15$.

Repeat the above four steps to get the final watermarked image H''.

3.2. Watermark extraction

Since we hide two kinds of watermarks in the image, we can extract different watermarks for different purposes. When we suspect that the image is pirated, we can extract the robust watermark to prove the copyright. First, the image with doubts is first decomposed by QR decomposition. Next, the watermarking bits are extracted from each R matrix to form a watermarked image. The extracted watermark is compared with the original watermark image. If the similarity between the two reaches a certain level, we can claim our own copyright. Similarly, we focus on the extraction of fragile watermark in this section. For the extraction of robust watermarks, please refer to our previous research [5].

If we want to confirm the integrity of the image T, we can extract the fragile watermark to verify. The method of extracting the robust watermark is described as follows. First, split the image into blocks that do not overlap in size by 4×4. Supposed that p_i denotes the pixel value of a block B of H', where $i = 0..15$, and supposed that (x, y) denotes the coordinate of B in H'. Let EM denote a binary error map of the image, which visually displays the result of tampering detection. For each block B, perform the following steps:

Step 1: $p_i' = p_i >> 2$, where $i = 0..15$.

Step 2: Calculate MD5(x, y, p_0', …, p_{15}') to generate 128-bit digest (d_0, d_1, …, d_{127}).

Step 3: $D_j = (d_{16\times j}, d_{16\times j}, \ldots, d_{16\times j+15})$, where $j = 0..7$. Perform XOR operation on D_0 to D_7 to get 16-bit message $M = (m_0, m_1, \ldots, m_{15})$.

Step 4: Extract the authentication message $A = (p_0\%2, p_1\%2, \ldots, p_{15}\%2)$.

Step 5: If $A \neq M$, mark a black on EM; otherwise, mark a white on EM.

Repeat the above five steps to get the error map of the image T.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we will present two experimental results, one for the experimental results of copyright verification and the other for the experimental results of tamper detection. The experiment of copyright verification is mainly to show that the hiding of fragile watermark does not affect the robustness of the robust watermark. The experiment of tamper detection is mainly to show that the fragile watermark can successfully mark the modified part of the image.

4.1. Copyright verification

In general, a robust watermarking scheme is evaluated by its imperceptibility and robustness [8]. The imperceptibility means that the difference between the host image and the watermarked image should not be perceived by human eyes, and the robustness means that the extracted watermark should be like the original watermark. In this paper, we use *PSNR* to evaluate the imperceptibility of our scheme.

$$PSNR = 10 \times \log \frac{255^2}{MSE}, \tag{2}$$

where

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (p_{i,j} - p'_{i,j})^2 \cdot \quad (3)$$

The notation M and N in Eq. (3) denote the width and height of an image, respectively, and $p_{i,j}$ and $p'_{i,j}$ denote the pixel of the original and watermarked image, respectively. Generally, the difference between the watermarked image and the host image is visually imperceptible if *PSNR* is greater than 30. The robustness is evaluated by the indicator *NC* as follows.

$$NC = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} (\overline{w_{i,j} \oplus w'_{i,j}}) \quad (4)$$

The notation W and H in Eq.(4) denote the width and height of the watermark image, respectively, $w_{i,j}$ and $w'_{i,j}$ denote the bit of the original and extracted watermark, respectively, and '⊕' represents the logical XOR operation. The larger the *NC* is, the more robust the extracted watermark is.

Figure 2 is our experimental images. When the watermark bit is embedded in the R matrix, the K value in the Eq. (1) needs to be determined first. In our scheme, we set different K values for the four elements of the R matrix. The set of parameter K in the robust watermarking scheme is (25, 30, 40, 60). The PSNR of the watermarked image showed that our scheme has good imperceptibility. In this experiment, we simulated that image was cropped to destroy the embedded watermark. In order to show the robustness of the proposed scheme, we simulated different cropping ratios, as shown in figure 3. Figure 4 is the watermark images corresponding to each of the attacked images in figure 3. Figure 4 shows that even if the watermarked image is subjected to a cropping ratio of 60%, the NC value is still as high as 0.8 or more. Therefore, embedding the fragile watermark after hiding the robust watermark does not impact the effect of copyright verification.

(a) Host image (512 × 512 pixels) *(b) Watermarked image* PSNR = 42.2096 *(c) Watermark (128 × 128 pixels)*

Fig. 2. The experimental images

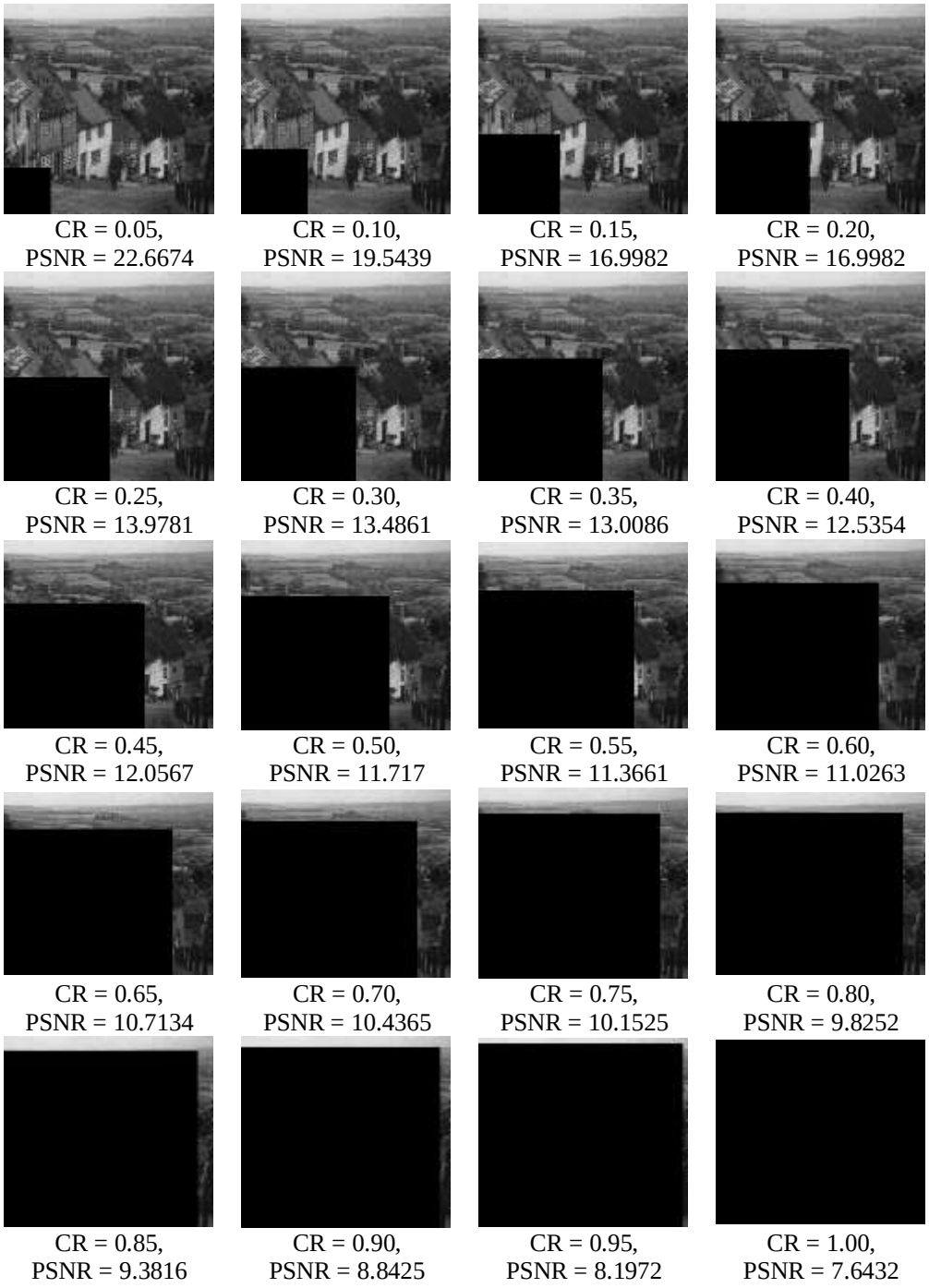

Fig. 3. The PSNR values for cropped images with different cropping ratio (CR) values

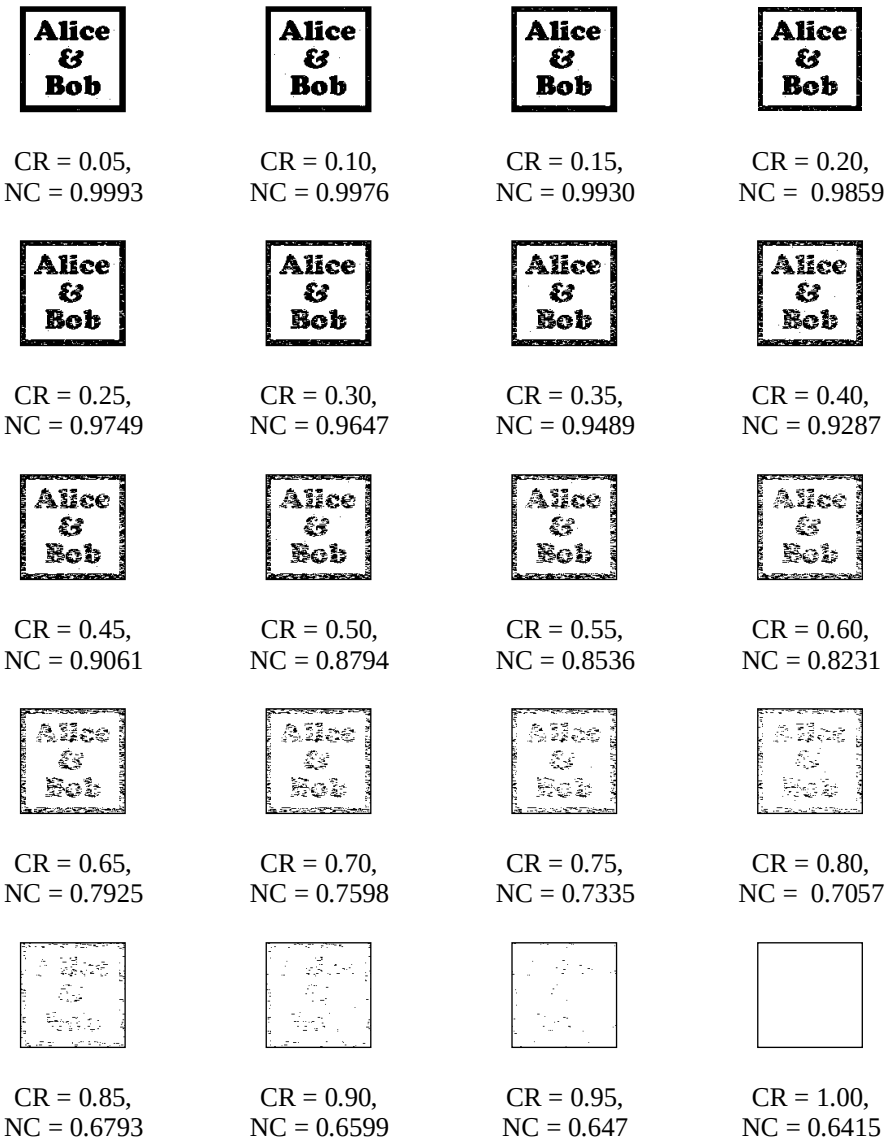

Fig. 4. The NC values of extracted watermarks for cropped images with different cropping ratio values

Figure 5 is a line graph of the correlation between CR and NC. It can be seen from this figure that NC did not drop sharply with the increase of CR value, which means that the method proposed in this study has good robustness.

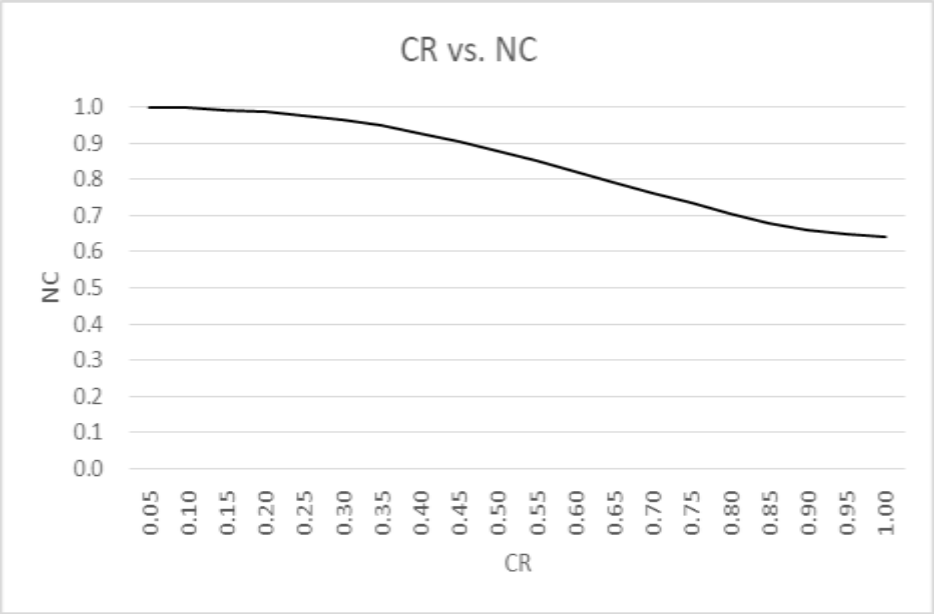

Fig. 5. The relationship between CR and NC

4.2. Tampering detection

This paper uses false negative rate (*FNR*) (Eq. (5)) and false positive rate (*FPR*) (Eq. (6)) to evaluate the efficiency of tampering detection and peak signal to noise ratio (*PSNR*) (Eq. (2)) to evaluate the quality of the watermarked image. The notations in Eq. (5) and Eq. (6) are explained as below [4, 8]:

FN: the number of false negative pixels, that is, the number of pixels that are tampered but are judged as untampered;

TP: the number of true positive pixels, that is, the number of pixels that are untampered but judged as tampered with;

FP: the number of true positive pixels, that is, the number of pixels that are tampered and judged as tampered;

TN: the number of true negative pixels, that is, the number of pixels that are untampered and judged as untampered.

$$FNR = FN / (FN + TP), \text{ and} \quad (5)$$

$$FPR = FP / (FP + TN). \quad (6)$$

We use figure 2 as experimental images under malicious cropping attacks. Using the proposed tampering detection scheme, we can successfully identify areas that have been tampered with. We visually display the detected area with the Error Map, as shown in figure 6.

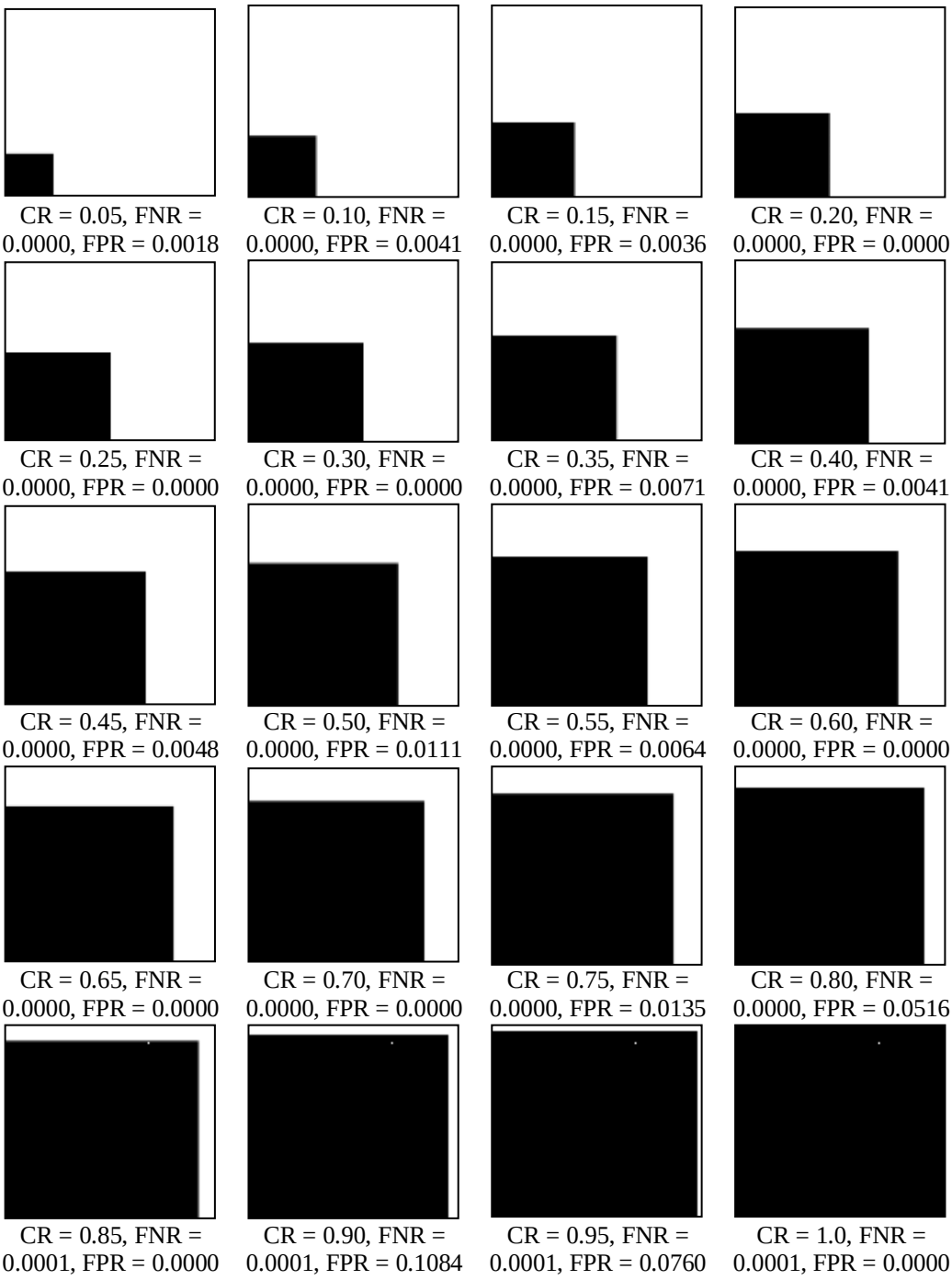

Fig. 6. The authentication result for cropping attacks

Since the cropping attack in figure 4 is less natural, we additionally simulate a malicious tampering that may occur. We add three objects to the watermarked image in figure 7(a): one is a puppy on the ground, one is an eagle in the sky, and the other is a chimney on the roof of a house. The image synthesized with the above three objects is shown in Figure 7(b). It can be observed from the Error Map of figure 7(c) that the objects can be detected. One can notice that the FNR of this experiment is 0, which means that the system can be successfully detected as long as it is tampered. The FPR of this experiment was 0.0088, indicating that the area that was not subjected to malicious tampering, about 0.88% of the pixels were misjudged as being tampered with. Although the place that has not been tampered with may be misjudged as being tampered with, the false positive rate is extremely low. Compared to false positive misjudgment, false negative is more serious because it means that the place that was maliciously tampered with may not be noticeable.

(a) Watermarked image
PSNR = 42.2096

(b) Watermarked image with malicious attack:
TR = 0.0389

(c) The authentication result:
FNR = 0.0000
FPR = 0.0088

Fig. 7. The authentication result for a malicious attack

4.3. Security analysis

In terms of security, this study uses a secret key to scramble the image block when embedding the robust watermark. Therefore, even if the attacker knows the algorithm of this study, but does not have this secret key, he cannot restore the order of the block and learn about our watermark image. Therefore, it is possible to prevent an attacker from modifying the embedded watermark content to claim the copyright of the image. In addition, this study also adds a secret key when generating the image authentication message. Therefore, the attacker cannot deduce the authentication message according to the algorithm of the study and deliberately change the content of the verification message to make the tampering detection invalid. In summary, this method meets the Kirchhoff's principle: all algorithms must be public, and only the key is secret. Kirchhoff's criterion is often used to evaluate the usefulness of cryptographic methods [9, 11]. If we need to pass the secret key over a public

network to a specific person, we can use asymmetric cryptography to protect the security of the secret key during transmission. We can encrypt the secret key with the public key of a specific person before transmitting the secret key. After receiving the specific key, that person can use his private key to decrypt the content of the secret key.

5. CONCLUSIONS

This study proposes a dual function watermarking scheme: one is copyright verification, and the other is tampering detection. Suppose we need to verify the copyright, we take the robust watermark representing the author's logo from the image. In order to avoid the verification of copyright, an attacker may try to destroy the image to remove the tough watermark. The experimental results of this study show that even if half of the image is cropped, the extracted tough watermark can clearly show the author's logo. Assuming we need to verify that the image has been tampered with, we can take out the fragile watermark. With the fragile watermark, we can not only know whether the image has been tampered with, but also the part that has been tampered with. In this research experiment, it can be clearly seen from the error map that this research method can mark the part that has been tampered with. In the future research, we will make the fragile watermark not only have the function of tamper detection, but also serve as the basis for enhancing the strength of the robust watermark.

REFERENCES

[1] I.A. Ansari, M. Pant, and C.W. Ahn, "SVD based fragile watermarking scheme for tamper localization and self-recovery," *International Journal of Machine Learning and Cybernetics*, 10.1007/s13042-015-0455-1, October 2015.

[2] S.C. Byun, S.K. Lee, A.H. Tewfik, and B.H. Ahn, "A SVD-Based Fragile Watermarking Scheme for Image Authentication," *Lecture Notes in Computer Science*, Vol. 2613, pp. 170-178, April 2003.

[3] J. Gao, L. Fan, and L. Xu, "Solving the face recognition problem using QR factorization," *WSEAS Transl. Mathematics*, Vol. 11, Iss. 8, pp. 712-721, August 2012.

[4] C.S. Hsu, S.F. Tu, "Probability-based Tampering Detection Scheme for Digital Images," *Opt. Commun.*, Vol. 283, No. 9, pp. 1737-1743, 2010.

[5] C.S. Hsu and S.F. Tu, "Digital watermarking scheme enhancing the robustness against cropping attack," *Proceedings of The 6th International Conference on Frontier Computing* (FC2017), pp. 143- 152, Osaka, Japan.

[6] H.F. Huang "A Fragile Watermarking Algorithm based on Chaos and QR Decomposition," *International Journal of Advancements in Computing Technology*, Vol. 5 Issue 4, pp. 117-124, February 2013.

[7] Q. Kang, K. Li, and H. Chen, "An SVD-based Fragile Watermarking Scheme With Grouped Blocks," *Proceedings of the 2nd International Conference on Information Technology and Electronic Commerce*, Dalian, China, December 2014.

[8] S. Katzenbeisser, F.A.P. Petitcolas, Information Hiding: Techniques for Steganography and Digital Watermarking, Artech House, Inc., MA, USA, 2000.

[9] O. Su, Y. Niu, H. Zou, Y. Zhao, and T. Yao, "A blind double color image watermarking algorithm based on QR decomposition," *Multimed. Tools Appl.*, Vol. 72, Iss. 1, pp 987-1009, September 2014a.

[10] Q. Su, Y. Niu, G. Wang, S. Jia, and J. Yue, "Color image blind watermarking scheme based on QR decomposition," *Signal Processing*, Vol. 94, pp. 219-235, 2014b.

[11] Q. Su, G. Wang, X. Zhang, G. Lv, and B. Chen, "An improved color image watermarking algorithm based on QR decomposition," Multimed. Tools Appl., vol. 76, no. 1, pp.707-729, 2017.

Information about the author(s):

Ching-Sheng Hsu got his BA degree from the Department of Information Management, National Cheng-Chi University, Taiwan, in 1994, MA degree from the Institute of Information Management, National Chi-Nan University, Taiwan, in 1998, and PhD degree from the Institute of Information Management, National Central University, Taiwan, in 2005. From 1998 to 1999, he was a software engineer at the Syscom Group Co, Taiwan, where his work focused on the Web-based stock trading systems. Currently, he is an associate professor of Department of Information Management, Ming Chuan University. His current research interests include blockchain technology, digital watermarking and information hiding, and intelligent computer-assisted learning and testing systems.

Shu-Fen Tu is a professor in Department of Information Management at Chinese Culture University in Taiwan. She received the BS degree in management information system from National Cheng-Chi University, Taiwan in 1996, the MS degree in information management from National Chi-Nan University, Taiwan in 1998, and the PhD degree from the Institute of Information Management, National Central University, Taiwan in 2005. From 1998 to 1999, she was a software engineer of the Syscom Group Co., Taiwan. Her current research interests include digital watermarking, secret sharing, and applications of blockchain.

Manuscript received on 2 January 2019