# A SECRET KEY BASED MULTISCALE FRAGILE WATERMARK IN THE WAVELET DOMAIN

*Hua Yuan and Xiao-Ping Zhang*

Department of Electrical & Computer Engineering
Ryerson University, 350 Victoria Street,
Toronto, Ontario, CANADA, M5B 2K3
xzhang@ee.ryerson.ca

## ABSTRACT

The distribution of the wavelet coefficients in 2-D discrete wavelet transform (DWT) subspaces can be well described by a Gaussian mixture statistical model. In this paper, a secret key based fragile watermarking scheme is presented based on this statistical model. The Gaussian statistical model parameters are obtained by an expectation maximization (EM) algorithm and modified in a way to form special relationships for image authentication. The secret key is designed to securely embed a message bit stream, such as personal signatures or copyright logos, into a host image. Because of the secret embedding key, the new method is robust to most image tampering, even when the attackers are fully aware of the watermark embedding algorithms. Besides, the secret embedding key can be encrypted and embedded as a robust watermark into the same host image of the fragile watermarks for the benefit that the decoding of fragile watermarks only requires a single encryption key other than the image itself. The new method also has the advantage of changing only a few image data for watermark embedding and being able to distinguish some normal image operations such as compression from malicious to achieve a semi-fragile application.

## 1. INTRODUCTION

With the advancement of the multimedia storage, transmission technologies and the development of the World Wide Web, people are now able to handle an increasing amount of digital information over the Internet. Digital multimedia is ubiquitous today. However, “seeing is no longer believing” [1]. Multimedia is easily reproduced and modified without any trace of manipulations. Therefore, the authentication techniques are required in applications where verification of integrity and authenticity of an image is essential [1, 2].

Fragile watermarking provides a possible solution to the above problem, since it makes possible to identify the author of an image by embedding some secret information in it and detect any unauthorized alterations in an image. The fragile watermarks can be embedded in either the space domain or the compressed domain of an image. With the focus in the space domain, several fragile watermarking methods that utilize the least significant bit (LSB) of the image data were developed [3, 4]. In the compressed domain, a wavelet-based fragile watermarking method that uses an optimal quantization step to detect image tampering is presented in [5]. The nature of multiresolution discrete wavelet decomposition allows the method to have spatial and frequency localization of image tampering. Other researchers noted the constraints of a single fragile watermark and developed a hybrid authentication watermark consisting of a fragile watermark and a robust watermark [6]. With the focus on the semi-fragile watermarking application, a technique [7] that accepts some acceptable operations such as JPEG compression and reasonable brightness adjustment on the watermarked image and rejects the manipulations due to malicious attacks is developed. Meanwhile, a generic content-based approach is proposed in the wavelet domain to extract invariant features from image contents and then sign and embed them back into the images as watermarks for the semi-fragile purposes [8]. In our previous work [9], we proposed a statistical model for fragile watermarking and implemented the watermarks at multiple wavelet scales to achieve a semi-fragile capability.

In this paper, a secret key based fragile watermarking method is presented. The secret key is used in watermark embedding and also required during the decoding process. The method is generally invulnerable to watermark counterfeiting because the embedding key can be any random combination of information bits and are kept unknown to potential attackers. When the embedding key is encrypted and embedded into the host image as a robust

1-4244-0367-7/06/$20.00 ©2006 IEEE

watermark, it is supposed to survive most unauthorized image tampering and help decode the embedded fragile watermarks that can detect and localize these image tampering.

## 2. THE SECRET KEY BASED FRAGILE WATERMARKING METHOD

The presented fragile watermarking method utilizes Gaussian mixture model parameters to form some special relationships for image integrity protection. Once these special relationships are formed, the unauthorized image tampering or attacks will break the relationships hence be detected and localized. Due to the security concerns, the formed authentication relationships are implemented based on a secret key and a secret code map so that it is generally impossible for an attacker to reproduce such relationships through image tampering.

Fig. 1. has an overview of the watermark embedding process. Authentication messages are initially translated into some binary bit streams. Then the wavelet subspaces at multiple scales are divided into a number of wavelet blocks depending on how many message bits being embedded and how many wavelet scales these bits will spread into. The binary bit streams are finally embedded into the wavelet blocks by forming some special relationships specified by the code map. The whole watermarking embedding process is secure and robust to malicious attacks because it is performed on a private key basis that guides a secret mapping between embedded bits and their corresponding wavelet blocks.

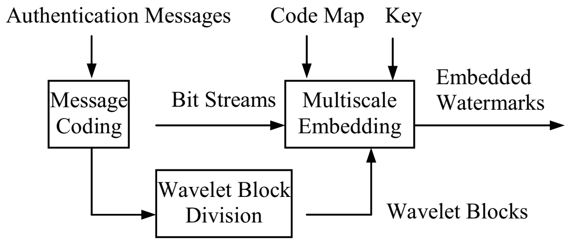

Fig. 1. Multiscale embedding of authentication messages

### 2.1. Special relationships formed by the Gaussian mixture model

The image authentication is achieved by the special relationships formed using the Gaussian mixture model in the wavelet domain [9]. The wavelet coefficients are found to have a peaky, heavy-tailed marginal distribution [10], which can be expressed by using a two component Gaussian mixture:

$$p(w_i) = p_s \cdot g(w_i, 0, \sigma_s^2) + p_l \cdot g(w_i, 0, \sigma_l^2), \qquad (1)$$

$$p_s + p_l = 1, \qquad (2)$$

where the class of small coefficients is represented by subscript "*s*" and the class of large coefficients by subscript "*l*". The *a priori* probabilities of the two classes are represented by $p_s$ and $p_l$, respectively. The Gaussian component $g(w_i, 0, \sigma_s^2)$ corresponding to the small coefficients has a relatively small variance $\sigma_s^2$ and the component $g(w_i, 0, \sigma_l^2)$ corresponding to the large state has a relatively large variance $\sigma_l^2$. All the model parameters $\left[p_s, p_l, \sigma_s^2, \sigma_l^2\right]$ can be obtained through an EM algorithm.

As known, the 2-D wavelet transform decomposes an image into three wavelet subspaces (horizontal, vertical and diagonal) at each scale. Suppose we take any two different wavelet blocks from the subspaces and apply the Gaussian mixture model to them, two different sets of model parameters are obtained. We can modify each large coefficient $w_i$ of one wavelet block by a certain amount $\Delta w$ so that its variance parameter $\sigma_l^2$ of the large coefficients will have the same value as that of the other block $\sigma_l'^2$ after modification, which can be formulated as:

$$\sum_{i=1}^{P}\left[(w_i + \Delta w)^2 - w_i^2\right] = K(\sigma_l'^2 - \sigma_l^2), \qquad (3)$$

where $P$ is the number of modified coefficients and $K$ is the total number of coefficients in the wavelet block. Once the modification is complete, a special parameter equity relationship between two wavelet blocks is formed and can be used for image authentication purposes. In case there is an image tampering or attack, it will break this relationship and hence be detected.

### 2.2. Secret key and code map based multiscale fragile watermarking

The formed special relationships can be used to implement and verify the existence of fragile watermarks. In order to embed fragile watermarks with sensitive information (such as a personal signature or logo) into the host image, a technique that involves proper wavelet block division to accommodate the embedded information bits is required, as shown in Fig. 1. Moreover, for secure embedding of the watermark information, a secret embedding key and a code map, which are unknown to potential attackers, are developed for watermarking.

The wavelet subspaces are divided into a number of blocks depending on how many information bits are to be embedded. Using the developed parameter equity relationship, each group of three wavelet blocks is able to embed two bits of watermark information. For security concerns, a secret embedding key is introduced to map between the embedded bits of a message bit stream and the selected group of three wavelet blocks. An example is shown in Fig. 2, in which the group with light shade is embedded with information bits 00 and the group with heave shade is embedded with another two bits information 01 from the message bit stream. The secret key can be encoded into a binary bit stream that controls the wavelet block selection for each group to embed two bits information. To unambiguously define the three wavelet block positions in a group, a straightforward way is to use the wavelet subspace index and the wavelet block coordinates within that subspace, which can be described as a format W-A-B, where W is the wavelet subspace index, with a value ranging from 0 to 2 to represent the three different wavelet subspaces (Vertical, Horizontal and Diagonal), and A and B are the row and column indexes of the wavelet block within that wavelet subspace, respectively, taking a value ranging from 0 to $2^n$, depending on how many wavelet blocks are needed and divided. Symbols W, A and B are all binary encoded in real implementation. Applying the above rule, the three light shaded blocks in Fig. 2 can be coded as 00-00-00, 01-01-01 and 10-00-10, respectively. Therefore the group that contains these three blocks to embed bits 00 is coded as 00-00-00-01-01-01-10-00-10. Once the key is generated, it is also possible to encrypt and embed the key into the host image as a robust watermark so that the decoder does not need any information other than a pre chosen encryption key and the image itself for watermark extraction.

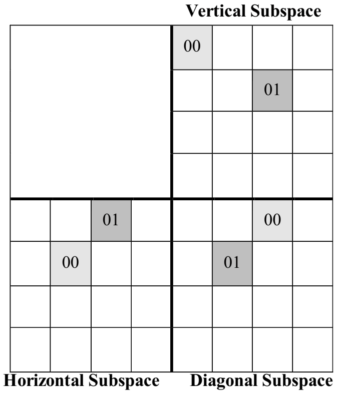

Fig. 2. Wavelet block selection for embedded message bits

| Formed relationship | Coded bits |
|---|---|
| $\sigma_1^2 = \sigma_2^2$ | 00 |
| $\sigma_1^2 = \sigma_3^2$ | 01 |
| $\sigma_2^2 = \sigma_3^2$ | 10 |
| $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ | 11 |

Table 1. Code map for message bits embedding

Since there exist multiple parameter equity relationships among three wavelet blocks, the code map is introduced to relate the actual embedded bits to the utilized relationships. Just like the secret key, the code map is unknown to potential attackers for security and is required at the decoding end for watermark extraction. An example of the code map is shown in Table 1. The parameters, $\sigma_1^2$, $\sigma_2^2$, $\sigma_3^2$, represent the variance parameters of three wavelet blocks in the same group. The coded bits for each relationship are interchangeable in Table 1 so that the same code map is required at the decoding end for proper watermark extraction. Without the knowledge of the code map, the watermarks cannot be properly resolved, either by the author who implement the watermarks or by any potential attackers.

The new fragile watermarking method can be applied to multiple wavelet scales to enhance the embeddability rate. Furthermore, by implementing the watermarks into multiple wavelet scales, we are able to distinguish some normal image operations such as image compression from malicious attacks, which is a desired property in semi-fragile applications. As will be shown in the simulations, the compression has a gradually decreased impact on wavelet coefficients and fragile watermarks when the wavelet scale increases. Other malicious attacks do not possess this characteristic.

## 3. EXPERIMENTAL RESULTS

The 512×512 Lena image is used to demonstrate the effectiveness of the new fragile watermarking method. In the experiment, our lab logo "CASPAL" is embedded into the Lena image. Since a 5 bits stream is used to encode the alphabet (00001 for A, 00010 for B, so on…), the total number of bits required to embed the logo is 30. Therefore we divide the wavelet subspaces at each scale into 16 (4×4) blocks so that they can accommodate 32 bits of information. Fig. 3 shows the wavelet subspaces with 32 message bits embedded into 16 divided wavelet blocks based on the secret key and code map. Every three wavelet blocks selected by the secret key compose a group to embed two message bits. For example, the three light shaded wavelet blocks embed message bits "00", which are the initial bits of the letter "C", using the relationship shown in the code map table (Table 1).

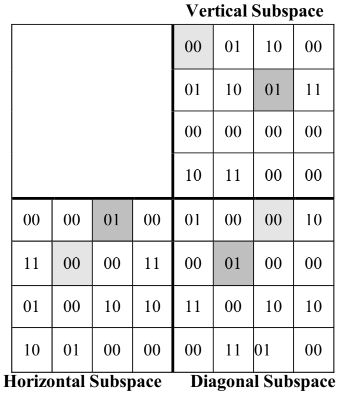

Fig. 3. "CASPAL" embedded into wavelet blocks using the secret embedding key and code map

Fig. 4 shows the Lena image after embedding the logo. There is no perceptible distortion on the watermarked image compared to the original one.

Fig. 4. Lena image with "CASPAL" embedded using the secret key and code map.

To demonstrate the semi-fragile functionality of the proposed watermarking method, some malicious attacks like the Gaussian white noise and deliberate image tampering are imposed on the watermarked Lena image. The severity or the extent of tampering can be recorded using the relative parameter differences to indicate how much the constructed parameter equity relationship is deviated because of the tampering. TABLE 2 lists the relative parameter differences caused by these malicious attacks at different wavelet scales, while Table 3 displays the JPEG compression impact on the parameter differences at these scale levels. It can be observed that at the same compression level, the relative parameter difference decreases as the wavelet scale increases. The parameter changes due to the malicious attacks do not possess this kind of characteristic. Therefore we are able to distinguish image JPEG compression from the malicious attacks by implementing the fragile watermarks at multiple wavelet scales.

| | | Malicious attacks | |
|---|---|---|---|
| | | Gaussian white noise | Content change |
| Scale Level | 1 | 8.35% | 4.78% |
| | 2 | 4.62% | 5.67% |
| | 3 | 5.52% | 1.73% |
| | 4 | 7.56% | 2.76% |

Table 2. Relationship between malicious attacks and parameter difference

| | | Compression ratio | | | |
|---|---|---|---|---|---|
| | | 60% | 38% | 25% | 15% |
| Scale Level | 1 | 1.68% | 2.76% | 3.23% | 4.11% |
| | 2 | 0.75% | 1.23% | 1.62% | 2.52% |
| | 3 | 0.30% | 0.57% | 0.96% | 1.65% |
| | 4 | 0.17% | 0.34% | 0.60% | 1.04% |

Table 3. Relationship between JPEG compression and parameter difference

## 4. CONCLUSIONS

In this paper, a multiscale fragile watermarking method is presented. The new method securely embeds fragile watermarks into a host image depending on a secret embedding key and a code map. The secret configurable algorithm parameters including the initial model parameters and convergence criteria used in EM algorithm provide additional security. All these secret key, code map parameters can also be encrypted and embedded into the same host image using some robust watermarking approaches. The new method is invulnerable to watermark counterfeiting because the secret key and the code map are kept unknown to potential attackers and are required at the decoding end for watermark extraction. A semi-fragile approach is achieved by applying the new method to multiple wavelet scales so that it can distinguish some normal image operations, such as compression, from malicious attacks.

## 5. REFERENCES

[1] B. B. Zhu, M. D. Swanson and A. H. Tewfik, “When seeing isn’t believing,” IEEE Signal Processing Magazine, vol. 21, no. 2, pp. 40-49, Mar. 2004.

[2] M. Celik, G. Sharma, E. Saber and A. M. Tekalp, “A hierarchical image authentication watermark with improved localization and security," Proc. ICIP, Thessaloniki, Greece, Oct. 2001.

[3] S. Walton, “Information authentication for a slippery new age,” Dr. Dobbs Journal, vol. 20, no. 4, pp. 18-26, Apr. 1995.

[4] P.W. Wong, “A public key watermark for image verification and authentication,” Proc. IEEE ICIP, Chicago, USA, Oct. 4-7, 1998, pp. 425–429.

[5] A. Paquet and R. Ward, “Wavelet-based digital watermarking for image authentication,” Proc. CCECE02, vol. 2, pp. 879-884.

[6] J. Fridrich, “A hybrid watermark for tamper detection in digital images,” Proc. Int. Symposium on Signal Proc. and Applications, pp. 301-304, Aug. 1999.

[7] C. Y. Lin and S. F. Chang, “Semi-fragile watermarking for authenticating JPEG visual content,” Proc. SPIE Security and Watermarking of Multimedia Contents II, pp. 140-151, Jan. 2000.

[8] Q. B. Sun and S. F. Chang, “Semi-fragile image authentication using generic wavelet domain features and ECC,” Proc. ICIP, vol. 2, Sep. 2002.

[9] H. Yuan and X.-P. Zhang, “A multiscale fragile watermark based on the Gaussian mixture model in the wavelet domain,” Proc. ICASSP, Montreal, Canada, May 17-21, 2004.

[10] J. Romberg, H. Choi and R. Baraniuk, “Bayesian tree-structured image modeling using wavelet-domain hidden Markov models,” IEEE Trans. Image Proc., Vol.10, No.7, July 2001.