A MULTISCALE FRAGILE WATERMARK BASED ON THE GAUSSIAN MIXTURE MODEL 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
hyuan, xzhang@ee.ryerson.ca

ABSTRACT

The wavelet coefficients in 2-D discrete wavelet transform (DWT) subspaces have a peaky, heavy-tailed marginal distribution that can be well described by a Gaussian mixture statistical model. In this paper, a multiscale implementation of fragile watermarks based on the Gaussian mixture model is presented. The presented new method can embed a message bit stream, such as personal signatures or copyright logos, into a host image. With the embedded message bits spreading over the whole image area, the new method can detect and localize any image tampering since it will inevitably destroy a certain message bits. Compared with some other fragile watermark techniques, the statistical model based method modifies only a very small amount of image data to embed watermarks and the modification is hardly perceived by human vision because it occurs at texture edges. Besides, the multiscale implementation of fragile watermarks based on the presented method can help distinguish some normal image operations such as compression from malicious attacks, which is meaningful in terms of semi-fragile watermarking applications.

1. INTRODUCTION

The digital images and related technologies have been very popular since 1990s. 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. The digitized images have certain advantages of easy operation and reproduction. However, these advantages may also facilitate copyright violation and content tampering. Therefore, the authentication techniques are required in applications where verification of integrity and authenticity of an image is essential [1].

Digital 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. Authentication watermarks, also known as fragile watermarks, can be used to 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 [2, 3]. In the compressed domain, a wavelet-based fragile watermarking method that allows spatial and frequency localization of image tampering is proposed in [4]. These methods need to change a large amount of image data to embed watermarks, which is not quite efficient and may reduce the quality of the watermarked image. In our previous work [5], we proposed a statistical method that modified only a very small amount of image data and had an imperceptible alteration of the host image. However, it is a type of one bit watermark approach and cannot be used to localize the image tampering.

In this paper, a novel fragile watermarking method that embeds watermarks at multiple wavelet scales is presented. The presented method has four attractive features: 1) Can embed personal signatures or logos into the host image; 2) Can detect and localize any slight image tampering; 3) Can distinguish some normal image operations from malicious attacks in case there are any unauthorized changes made on the watermarked image; 4) Embed watermarks through modifying only a few image data. Numerical examples are given to demonstrate the effectiveness of the new method.

2. THE MULTISCALE FRAGILE WATERMARKING METHOD

The presented fragile watermarking method has three steps. First, a Gaussian mixture statistical model in the wavelet domain is employed and a related EM (Expectation Maximization) algorithm is used to find the

model parameters. Second, the watermark information for image authentication is embedded into the statistical model by forming a special relationship among model parameters. Third, personal messages are embedded at multiple wavelet scales.

2.1. The Gaussian mixture statistical model in the wavelet domain

The presented fragile watermarking method is based on a statistical modeling of images in the wavelet domain. The obtained model parameters are used to guide the modification of selected wavelet coefficients in order to construct a special relationship for image authentication purposes. The statistical modeling includes two steps.

In the first step, a Gaussian mixture statistical model is developed. The wavelet coefficients are found to have a peaky, heavy-tailed marginal distribution [6], which can be expressed by using a two component Gaussian mixture:

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

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

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 σ_s^2, capturing the peakiness around zero, while the component $g(w_i, 0, \sigma_l^2)$ corresponding to the large state has a relatively large variance σ_l^2, capturing the heavy tails. Note w_i, $i=1,...,K$, represents the wavelet coefficient.

In the second step, an EM algorithm similar to [5] can be applied on the Gaussian mixture model to obtain the model parameters $[p_s, p_l, \sigma_s^2, \sigma_l^2]$.

2.2. Embedding information into the statistical model

As known, the 2-D wavelet transform decomposes an image into three wavelet subspaces (horizontal, vertical and diagonal) at each scale. If the Gaussian mixture model and the EM algorithm are applied to these three subspaces, three different sets of model parameters will be obtained. We can modify the large coefficients of a single wavelet subspace so that its variance parameter σ_l^2 of the large coefficients has the same value as that of another wavelet subspace $\sigma_l'^2$. This specially formed relationship can be used as watermark for image authentication purposes. Any image operations or malicious attacks will break this relationship and be detected. The large coefficients usually represent image edges in the space domain which, when modified, are generally difficult to be detected by human vision. Moreover, the quantity of large coefficients in a wavelet subspace is small. The changes made on them will then be imperceptible.

In the new approach, all large coefficients are modified by a same amount Δw in order to reach the target parameter value $\sigma_l'^2$. The relationship between σ_l^2, $\sigma_l'^2$ and Δw can be expressed as follows:

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

where P is the number of large coefficients and K is the total number of coefficients in the wavelet subspace. Once the large coefficients are modified and the special relationship among model parameters of different wavelet subspaces is formed, the watermarked image is formed and can be used for authentication purposes.

2.3. Multiscale embedding of personal messages and the benefits

We consider the following three aspects very attractive for a fragile watermarking system: 1) Can embed some personal messages into the host image. 2) Can localize the image tampering if there is any. 3) Can distinguish some normal image operations from malicious attacks. Integrated with some coding techniques and implemented at multiple wavelet scales, our proposed method is able to achieve the above objectives.

To embed personal messages into the host image, we need to embed a message bit stream instead of a single bit. Instead of using the whole wavelet subspace to embed the watermark, each wavelet space is divided into blocks to embed a bit stream. Every three wavelet blocks obtained at the same position from the wavelet subspaces (Horizontal, Vertical, Diagonal) can form a special relationship to encode two message bits. An example is shown in Table 1.

Formed relationship	Coded bits
$\sigma_{l,V}^2 = \sigma_{l,H}^2$	00
$\sigma_{l,V}^2 = \sigma_{l,D}^2$	01
$\sigma_{l,H}^2 = \sigma_{l,D}^2$	10
$\sigma_{l,V}^2 = \sigma_{l,H}^2 = \sigma_{l,D}^2$	11

Table 1. Code map for message bits embedding

The parameters $\sigma_{l,H}^2, \sigma_{l,V}^2, \sigma_{l,D}^2$ represent the variances of the large coefficients of the three wavelet blocks obtained from horizontal subspace, vertical subspace and diagonal subspace respectively. Various parameter equity relationships among $\sigma_{l,H}^2$, $\sigma_{l,V}^2$ and $\sigma_{l,D}^2$ can be formed in the way shown in Table 1 to encode different two bits into these three wavelet blocks. Since there are N^2 groups of such wavelet blocks, at most $2N^2$ bits at a single wavelet

scale can be embedded. Any unauthorized changes made in a specific area of the watermarked image will destroy the corresponding relationship and message bits, therefore the tampering can be detected and localized.

The new method can be used to embed message bits into multiple wavelet scales so that the watermark embeddability can be further enhanced. Furthermore, it can help us to distinguish some normal image operations such as image compression from malicious attacks. Therefore we may determine the source of tampering. As will be shown in the next section, the compression has a gradually decreased impact on wavelet coefficients and fragile watermarks as the wavelet scale increases. Other malicious attacks do not have this characteristic.

For the proposed method, the distortion is imperceptible because it employs a statistical approach to embed watermarks, which modifies only a few image data at both high and low scale levels.

3. EXPERIMENTAL RESULTS

The 512×512 image of peppers is used to demonstrate the effectiveness of the new fragile watermarking method. In the experiment, our lab logo "CASPAL" is embedded into the peppers 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. According to Table 1, at least 15 blocks are needed at each wavelet subspace. To facilitate the operation, each wavelet subspace is divided into 16 blocks so that 32 message bits (representing the logo) can be embedded. Fig. 1 shows the wavelet subspaces with 32 message bits embedded into 16 divided wavelet blocks. Every three wavelet blocks obtained at the same position from the wavelet subspaces are used to embed two message bits. For example, the three shaded wavelet blocks embed message bits "00", which are the initial bits of the letter "C", using the relationship shown in Table 1. Fig. 2 shows the peppers image after embedding the logo. There is no perceptible distortion on the watermarked image compared to the original one.

To test the sensitivity of the watermark detection, first we perform a single pixel tampering experiment by 20 times, in each of which a randomly selected pixel is modified by a small amount and its impact to the embedded watermark is recorded. Table 2 shows the mean value of relative parameter differences deviated from the constructed parameter equity relationship as in Table 1, which can represent the sensitivity of the watermark detection.

As can be seen, no matter how slight the tampering is, it will be detected because it destroys the formed parameter equity relationship by a noticeable amount. The location of the tampering can be determined since it only destroys the message bits at positions of the tampering.

Number of trials	20
Average parameter difference off balance	0.11%
Corresponding message bits destroyed	Yes

Table 2. Average parameter difference caused by single pixel modification

In the present experiment, the total number of modified coefficients in Fig.2 is 680 (out of 512×512). Compared with some conventional fragile watermarking methods [2-4] that modify half of the image pixels, the new method modifies much fewer image data. In our experiment, the PSNR (peak-signal-to-noise-ratio) is 52.12 db, indicating that the new method makes an imperceptible alteration of the host image to embed watermarks.

	Noise variance				
Scale Level		0.0001	0.0002	0.0005	0.0010
	1	1.25%	2.37%	6.06%	19.42%
	2	4.65%	8.59%	10.27%	17.49%

Table 3. Parameter difference caused by noise

Table 3 lists the relative parameter differences caused by the additive noise with different variances using the new watermark embedding method at two scales. As can be seen, no matter how slight the tampering is, the previously formed parameter equity relationship will be broken and the tampering will be detected. Moreover, the parameter differences tend to become larger with the increase of the extent of tampering.

	Compression ratio				
Scale Level		60%	38%	25%	15%
	1	1.47%	2.36%	2.86%	4.53%
	2	0.84%	1.27%	1.76%	2.75%
	3	0.21%	0.31%	1.18%	1.99%
	4	0.09%	0.23%	0.54%	0.81%

Table 4. Parameter difference caused by compression

	Malicious attacks		
Scale Level		Gaussian white noise	Content change
	1	3.06%	1.65%
	2	7.27%	0.34%
	3	6.13%	0.98%
	4	14.75%	1.13%

Table 5. Parameter difference caused by some malicious attacks

By embedding the fragile watermarks at multiple wavelet scales, we can distinguish some normal image operations such as image compression from malicious attacks. Table 4 shows the impact of the JPEG compression on the watermarked image that has

watermarks embedded at scales 1 to 4. The numbers in Table 4 represent relative parameter differences. It can be seen that at the same compression level, the relative parameter difference decreases as the wavelet scale increases. On the other hand, some malicious attacks, including additive Gaussian white noise and deliberate slight change of image contents, are simulated. The simulation results are shown in Table 5. The parameter changes due to these malicious attacks do not have the same characteristic as the image compression. Therefore we may use this feature to distinguish image compression from these malicious attacks.

4. CONCLUSIONS

In this paper, a new fragile watermarking method is presented. A Gaussian mixture statistical model is used to embed watermarks, resulting in a very small amount of modified image data and an imperceptible alteration of the original image. Integrated with coding techniques, the new method can embed personal signatures or logos into the host image so that any unauthorized changes will be detected and localized. A multiscale implementation of the approach can enhance the robustness of tampering detection and help distinguish some normal image operations, such as compression, from malicious attacks.

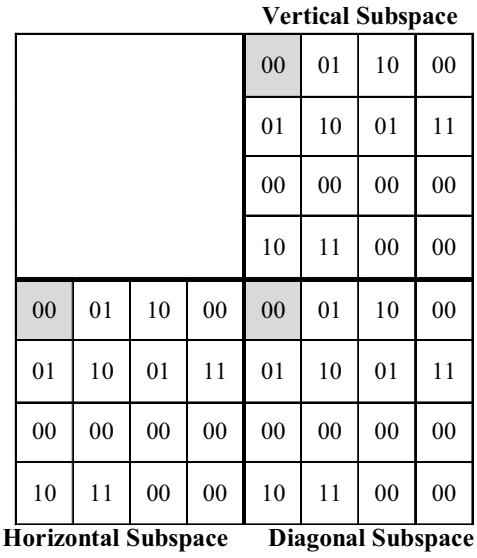

Fig. 1. Message bits embedded into the wavelet blocks

5. REFERENCES

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

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

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

[4] A. Paquet and R. Ward, "Wavelet-based digital watermarking for image authentication," IEEE CCECE02 Proceedings, vol. 2, pp. 879-884.

[5] H. Yuan and X.-P. Zhang, "Fragile watermark based on the Gaussian mixture model in the wavelet domain for image authentication," Proc. IEEE ICIP, Barcelona, Spain, Sep. 2003.

[6] 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.

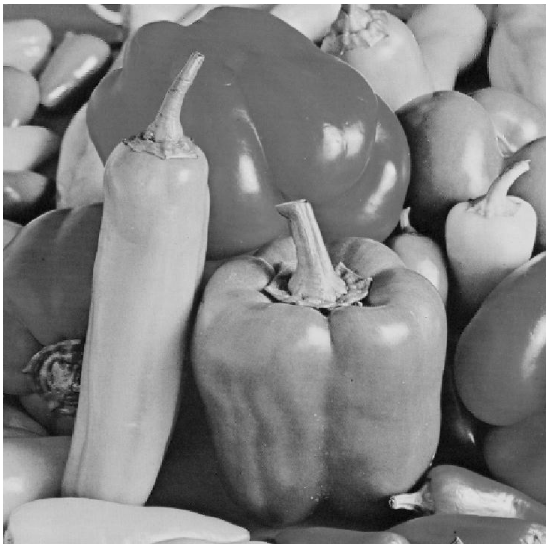

Fig. 2. The watermarked peppers image with the lab logo "CASPAL" embedded.