# Semi-Fragile Watermarking for Telltale Tamper Proofing and Authenticating

Han Ho Ko and Sang Ju Park
Department of Electronic Engineering, Hongik University
72-1 Sangsu-dong, Mapo-ku, Seoul 121-791, Korea
Tel: +82-2-333-6232, Fax: +82-2-320-1119
E-mail: sve97@hotmail.com, sjpark@hongik.edu

**Abstract:** Extreme development in digital multimedia has raised anxiety in the minds of copyrighted content owners. This has resulted in the creation of several watermarking techniques.

This paper, proposes a method of embedding a perceptually transparent digital signal, named semi-fragile watermark in the wavelet domain, utilizing the characteristics of the human visual system. So as to detect attacks inflicted on the content and use an algorithm to specify the character of the attack.

## 1. Introduction

The massive distribution and development of digital multimedia and with the aid of image processing software, which is easily available in the modern day market, make editing and inappropriate distribution of digital content a problem. Consequently providers of intellectual property are naturally concerned with intellectual rights. Therefore watermarking and cryptographic systems, which have been developed for the issues mentioned above, have recently come to their attention [7].

Watermark is more recommended over cryptography for the reason that cryptography does not offer protection after the decryption process. Two types of watermarks are presently under consideration, namely robust watermarking and fragile watermarking. Researches are mostly focused on the development of robust watermarking, designed for the copyright protection of multimedia content. Such methods embed a perceptually transparent digital signal in the original signal without degrading the quality of the content, so as to withstand any illegal attacks to remove the watermark [4]. And the other type, which is equally important but still underdeveloped, is addressed as fragile watermarking. This technique also embeds an imperceptible watermark in the host content but the object is not to withstand the attack but to detect and localize the alteration, which has been inflicted on the watermarked content. Unlike the applications of the robust watermarking, the fragile watermarking is primarily used for tamper proofing and authenticating the content in question.

Most multimedia content in digital format is stored in compressed form, to facilitate the matters concerning storage space and transmission. Naturally fragile watermarking utilizing the hash function [8] is inappropriate, because of the properties of the hash function, which states that two different inputs must not produce the same output and this could result in a severe problem when it comes to lossy compression. So recently a new type of watermark has been proposed and is being developed, titled semi-fragile watermarking, which has the combined characteristics of both the robust and fragile watermarking [3]. Like robust watermarking, semi- fragile watermarking should be able to tolerate unintentional attacks such as lossy compression, cropping and rotation but must detect any malicious modifications, such as replacing or adding of features. However the primary application of the semi-fragile watermarking is tamper proofing and authenticating, so the features of the semi-fragile marking system generally resemble those of the fragile watermarking.

## 2. Basis Techniques

### 2.1 Telltale Tamper Proofing

The core requirement of a semi-fragile watermark is that it must be able to determine the authenticity of the content. So to speak, it must decide with objective assurance that, one content is equal to or similar in the sense that the change is perceptually unnoticeable.

The expression tamper proofing describes that the watermark must detect any malicious changes. This is easily achieved by using the hashed digest of the original signal to decide the authenticity of the content [8]. But the disadvantage of the hash function other than the one mentioned previously is that it cannot localize the attack, only detect it. The more advanced tamper proofing method is one that enables the watermark to localize the attack, where there are several methods. For example row-column hash function (RCHF) technique, block-base hash function (BBHF) technique which utilizes the hash function [8]. Other methods use the characters of the DCT transform and they are the method using the block correlation detector [3] or the method of embedding a spread-spectrum watermark [5]. But the phrase telltale tamper proofing aims at not only localizing the attack, but also characterizing it. In order to do this effectively the watermark is embedded in the wavelet domain. Unlike the more widely used discrete cosine transform (DCT), it produces information of both spatial localization regions and frequency region information due to the hierarchical decomposition formula of the wavelet transform. Another advantage that could be achieved by using the wavelet transform is that it makes the employment of the human visual system (HVS) simpler.

### 2.2 Human Visual System

It is a necessity to take into account the visual effect of embedding a watermark into a host image in order to create a more effective watermark. Lewis and Knowles applied the assumption of the human visual system to design an extra efficient algorithm for image compression in [1]. The assumptions are that the human sight is less sensitive to the high frequency band areas and diagonal noise patterns and also that the human vision take little notice of the noise in the texture areas with high concentration of high frequency components. Utilizing the assumptions mentioned above, by embedding a watermark with extra weight added to the portion that the human eyesight is less sensitive to generate a more effective watermark. Figure 1 illustrates the pre-calculated weight values of Barbara, utilizing HVS. The darker region is considered visually less sensitive or has relatively greater weight values

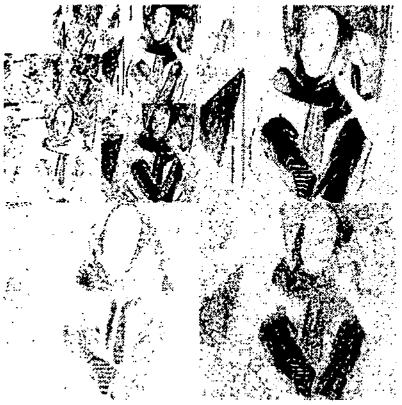

Figure 1. Weight Map

## 3. The Proposed

### 3.1 Embedding Process

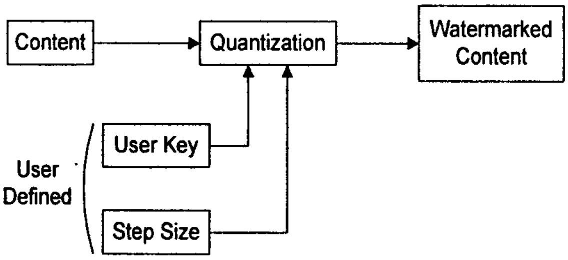

Figure 2. Watermark Embedding Process

The embedding process is initiated when a user key generates a Pseudo random pattern, $w(i)$, which will work as the watermark and an appropriate quantization step size, $\Delta$, is chosen. So as not to degrade the perceptual quality of the image. Both of these elements will work as a secret key to prevent illegal extraction of the watermark.

Then the watermark is embedded during the quantization procedure as denoted in Figure 2. The elementary unit of this watermarking system is a block size of $2\times2$ pixels, embedded in the zigzag order similar to the scanning order of the embedded zerotree wavelet algorithm (EZW) [6], excluding the LL subband. In which could result in a severe distortion in the reconstructed image if tampered with.

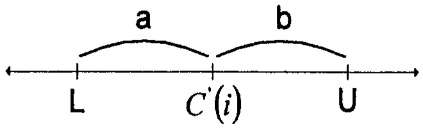

Figure 3. Embedding Method

L and U in Figure 3 represent the lower and upper bound of quantization bin. $C'(i)$ is placed in the middle of the bin. Table 1 display the algorithm of the proposed method where $C(i)$ is the wavelet transformed coefficient of the host image and $C^*(i)$ is the watermarked coefficient. The term $\omega$ denotes the pre-calculated weights generated using the HVS characteristics.

| $\omega(i)$ | $C(i)$ | $C^*(i)$ |
|---|---|---|
| 1 | $C(i)\in a$ | $C'(i)+\omega\times\Delta/2$ |
| 0 | $C(i)\in a$ | $C(i)$ |
| 1 | $C(i)\in b$ | |
| 0 | $C(i)\in b$ | $C'(i)-\omega\times\Delta/2$ |

Table 1. Embedding Method

### 3.2 Detection Process

The detection process is begun by extracting the watermark. This procedure is simplified in Figure 4, which illustrates that the watermark is extracted by inserting the user key and the quantization step size of the possible tampered watermarked image into the detector.

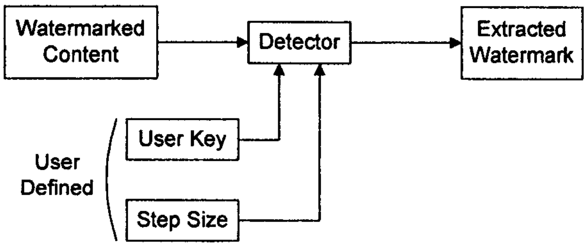

Figure 4. Watermark Extraction

The extracted watermark then undergoes an exclusive-OR operation as indicated in algorithm (1), which is the

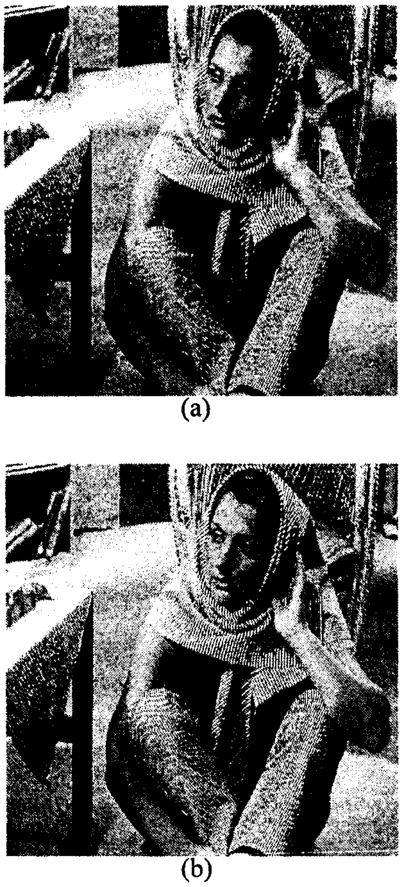

Figure 5. Original and Watermarked Image: (a) Original Image and (b) Watermarked Image with $\Delta = 5$ (PSNR=35.9238dB)

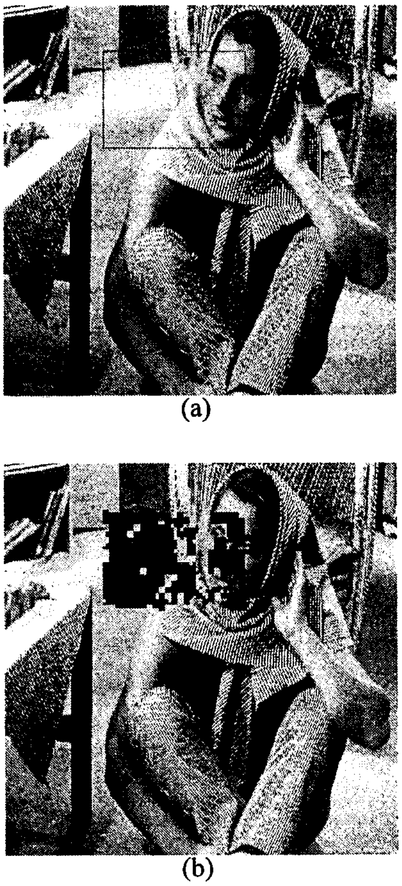

Figure 6. (a) Tampered Image the Portion in the rectangle has been blurred (b) After Detection Process ($\alpha = 0.5$, $\beta = 0.3$ ,$\gamma = 0.2$, $\delta = 0.6551$)

fundamental operation for the proposed system, with $w^*(i)$ being the extracted watermark and $l$ as the resolution level. The denominator 12 represents the 4 pixels of the 3 different subbands in the same resolution level that of the same spatial location, which is the basic unit of the proposed detection process. We determine whether a block has been altered in response to the value of A. The threshold T is defined to match the sensitivity and the application of the image in question. If A is large then the threshold the image is considered authentic, but if not altered.

$$A_l(w,w^*)=\frac{1}{12}\sum_{i=1}^{12} w(i)\oplus w^*(i) \qquad (1)$$

And by taking advantage of the wavelet transform decomposition, precedes a more detail analysis of the image. The method above is similar to those of the previous techniques, but what makes the purposed unique is the algorithm which is indicated in (2), where $\delta$ denotes the combined value of the numerous resolution levels and $\alpha,\beta,\gamma$ notes the different weight for each of that resolution level. As the same with algorithm (1) the judgment is made by comparing $\delta$ with the threshold, authentic if larger and altered if not.

$$\delta=\alpha A_l(w,w^*)+\beta A_{l-1}(w,w^*)+\gamma A_{l-2}(w,w^*) \qquad (2)$$

By changing the three parameters the detecting process can determine the characteristics of the attack inflicted on the image. So as to speak by modifying the three parameters we can verify which frequency components of the watermarked image has been altered and so the terms of the telltale tamper proof watermark can be better approved. For a more precise analysis of the altered image the number of the parameters can be increased.

## 4. Results

The simulation was conducted on $512\times512$ size 'Barbara' image and we observed how it reacts to familiar intentional and unintentional attacks such as blurring and JPEG and SPHIT compression algorithms. Figure 5 illustrates the original and the watermarked image. Despite the rather low PSNR value hardly any visual difference was noticeable on paper all even on the computer screen, due to the usage of HVS characteristics. Table 2 denotes the suitable thresholds for the different $\alpha,\beta,\gamma$ parameters.

| $\alpha$ | 0.1 | 0.2 | 0.3 | 0.5 | 0.7 |
|---|---|---|---|---|---|
| $\beta$ | 0.2 | 0.3 | 0.4 | 0.3 | 0.2 |
| $\gamma$ | 0.7 | 0.5 | 0.3 | 0.2 | 0.1 |
| $\delta$ | 0.7634 | 0.6965 | 0.6557 | 0.6551 | 0.7376 |

Table 2. Thresholds for Different Parameters

By using the thresholds in Table 2 a simulation is performed. Figure 6 demonstrates the detection after a blurring attack.

| Q<br>Factor | bpp | 0.3<br>0.4<br>0.3 | 0.5<br>0.3<br>0.2 | 0.7<br>0.2<br>0.1 | $\alpha=\beta=\gamma$ |
|---|---|---|---|---|---|
| 1 | 4.9996 | 0.7430 | 0.7168 | 0.7401 | 0.7358 |
| 5 | 3.2532 | 0.7810 | 0.7685 | 0.8015 | 0.7898 |
| 10 | 2.5339 | 0.8287 | 0.8073 | 0.7860 | 0.8204 |
| 15 | 1.8936 | 0.9077 | 0.8589 | 0.8456 | 0.8997 |
| 20 | 1.6375 | 0.9483 | 0.9184 | 0.9365 | 0.9459 |
| 25 | 1.4238 | 0.9693 | 0.9376 | 0.9058 | 0.9698 |
| 30 | 1.2988 | 0.9667 | 0.9500 | 0.9334 | 0.9715 |
| 35 | 1.1864 | 0.9866 | 0.9702 | 0.9667 | 0.9817 |
| 40 | 1.0928 | X | 0.9976 | 0.9837 | X |
| 45 | 1.0202 | X | 0.9579 | 0.9296 | X |
| 50 | 0.9602 | X | 0.9931 | 0.9759 | X |
| 55 | 0.9011 | X | 0.9728 | 0.9322 | X |
| 60 | 0.8328 | X | 0.9998 | 0.9999 | X |
| 65 | 0.7677 | X | 0.9919 | 0.9573 | X |

Table 3. Performance for JPEG

| bpp | 0.3<br>0.4<br>0.3 | 0.5<br>0.3<br>0.2 | 0.7<br>0.2<br>0.1 | $\alpha=\beta=\gamma$ |
|---|---|---|---|---|
| 5.0 | 0.7015 | 0.6973 | 0.7557 | 0.6935 |
| 4.7 | 0.7253 | 0.6640 | 0.7383 | 0.6990 |
| 4.5 | 0.6814 | 0.6606 | 0.7391 | 0.6688 |
| 4.3 | 0.6714 | 0.6674 | 0.7393 | 0.6671 |
| 4.0 | 0.6978 | 0.6706 | 0.6804 | 0.6972 |
| 3.7 | 0.7357 | 0.7719 | 0.8080 | 0.7493 |
| 3.5 | 0.8134 | 0.8256 | 0.8378 | 0.8148 |
| 3.3 | 0.8306 | 0.8245 | 0.8300 | 0.8336 |
| 3.0 | 0.888 | 0.9000 | 0.9574 | 0.8874 |
| 2.7 | 0.9869 | 0.9565 | 0.9260 | 0.9861 |
| 2.5 | X | 0.9774 | 0.9948 | X |
| 2.3 | X | 0.9998 | 0.9997 | X |
| 2.0 | X | 0.9984 | 0.9979 | X |

Table 3. Performance for SPHIT

How the proposed method reacts to JPEG and SPHIT compression formula is displayed in Table 3 and 4. As it exhibits the proposed is more robust to significantly higher compression rates than the previous techniques for JPEG and slightly higher rates for SPHIT. The X denotes that there is not suitable threshold for that specific stage.

## 5. Conclusion and Future Studies

Semi-fragile watermarking has proven superior to fragile watermarking because of the fact that it has the ability to adapt to the growing needs of multimedia signals. Furthermore telltale tamper proofing has open a new way of protecting copyrighted information, which could not have been possible by preceding methods.

We proposed a semi-fragile watermarking method for images, which utilizes human visual nature and also is more robust to lossy compression algorithms than the existing systems. More over it has the ability to characterize the nature of the attack inflicted on the image, with more precision than the existing. Future studies will involve characterization of geometrical tampering and a mathematical model for the parameters in the proposed method.

## References

[1] AS. Lewis, and G. Knowles, "Image compression using the 2-D wavelet transform" IEEE Trans. Image Processing 1, pp. 244-250, April 1992.

[2] D. Kundur, and D. Hatzinakos, "Digital Watermarking for Telltale Tamper Proofing and Authentication," Proceedings of the IEEE, vol. 87, no. 7, pp.1167-1180, July 1999.

[3] E. T. Lin, C. I. Podilchuk, and E. J. Delp, "Detection of Image Alterations Using Semi-Fragile Watermarks," Proceedings of the SPIE International Conference on Security and Watermarking of Multimedia Contents II, vol. 3971, San Jose, CA, January 23-28,2000.

[4] G. C. Langelaar, I. Setyawan, and R. L. Langendijk, "Watermarking Digital Image and Video Data," IEEE Signal Processing Magazine, vol. 17, no. 5, pp.20-46, September 2000.

[5] J. Fridrich, "Image watermarking for tamper detection," Proceeding of the IEEE International Conference on Image Processing, vol. 2, pp.404-408, Chicago, Illinois, October 1998.

[6] J. M. Shapiro, "Embedding Image Coding Using Zerotrees of Wavelet Coefficients," IEEE Transactions on Signal Processing, vol.41, no. 12, pp. 3445-3462, December 1993.

[7] R. B. Wolfgang, C. I. Podilchuk and E. J. Delp, "Perceptual Watermarks for Digital Images and Video," Proceeding of the IEEE, vol. 87, no. 7, pp.1108-1126, July 1999.

[8] R. B. Wolfgang, and E. J. Delp, "Fragile Watermarking Using the VW2D Watermark," Proceeding of the IS&T/SPIE Conference on Security and Watermarking of Multimedia Contents, pp.204-213, San Jose, CA, January 1999.