SECURE, BLIND IMAGE WATERMARKING TECHNIQUE FOR COPYRIGHT PROTECTION

Faisal Alurki and Russell Mersereau
School of Electrical and Computer Engineering, Georgia Institute of Technology,
Atlanta, GA 30332-0250, USA.
Tel: 404-894-2977; fax: 404-894-8363
e-mail: faisal,rmm@ece.gatech.edu

ABSTRACT

We present a new oblivious digital watermarking method for copyright protection of still images. The technique is based on decorrelating the image samples then embedding the watermark by taking block DCT of the decorrelated image. Watermark insertion is obtained by amplitude modulating the DC component of the DCT blocks with the watermarking bits. The technique shows satisfactory robustness to Stirmark benchmark test and other class of geometric deformations. Image decorelation is achieved using a key to increase the security of the watermarking system. The watermark is a readable sequence of binary digits channel encoded to increase its robustness against various types of attacks.

1 introduction

A digital watermark is a short sequence of information containing an owner identity or copyright information embedded in a way that is difficult to erase. Digital watermarking has been proposed as a possible solution for resolving ownership and for copyright protection of multimedia data such as still images [1, 2]. In recent years the field of digital watermarking have captured a lot of attention from researchers. Several techniques have been proposed for copyright protection of still images [3, 4, 5]. These techniques demonstrate high degree of robustness against several types of attacks simulated by the latest Stirmark benchmark. In this paper we present a new blind digital watermarking algorithm. The technique is based on permuting the location of the image samples, then embedding the watermark in the transform domain of the permuted image. Watermark insertion is achieved by first taking a block DCT of the permuted image then, amplitude modulating the DC component of the DCT blocks. Embedding the watermark in the DC value of a block of an image transform increases the watermark robustness to image processing operations, this is because the DC values of an image transform tend to preserve their values under image processing operations and they do not change their values drastically. The watermark is embedded by amplitude modulating the DC value in each block. After the watermark is embedded we take the inverse block transform and use the same key to descramble, to get the watermarked image. The permutation operation increases the security of the watermark from being extracted assuming the embedding algorithm is known, i.e., without knowing the key used to permute the pixels locations, an attacker cannot remove the embedded watermark. Furthermore, the system implementation complexity is low and watermark embedding and extraction is simple.

2 System Analysis

Let $x(n_1, n_2)$ be some natural image to be watermarked, the image gets scrambled by a key K to get the decorrelated image $\overset{*}{x}\ (n_1, n_2)$. The permutation process is described in details in [6] . Next, we take a block DCT of $\overset{*}{x}\ (n_1, n_2)$, the blocks are chosen to be of middle size, typically 32×32 or 64×64. The randomness introduced by the scrambling of pixel locations results in spectrum equalization or spectrum whitening of the host image. Furthermore, the scrambling operation results in a signal with similar statistical properties for each block in the transform domain. The statistical properties become close in similarities as the block size increases. For example, by selecting a blocks of size 32×32 or 64×64 the DC components of the blocks will have similar values. Figure 1 shows the magnitude of the DC values for a 64×64 DCT blocks of the randomly scrambled image $\overset{*}{x}\ (n_1, n_2)$. The figure indicates that the DC values are very close in magnitude. For simplicity, we replace the DC values for all blocks by a constant value A close to their values. Typically A is chosen to be the mean value of all DC values rounded to the nearest integer multiple of the block size. Next, we embed the watermarking bits, which are independent of the host image, by amplitude modulating the new DC component A of the DCT blocks as follows

$$\tilde{A}_i = A(1 + \alpha d_i), \tag{1}$$

where $\tilde{A}_i$ is the modulated DC coefficient of the i^{th} block α is a scaling factor controlling the strength of the watermark, basically α operates as a modulation index and

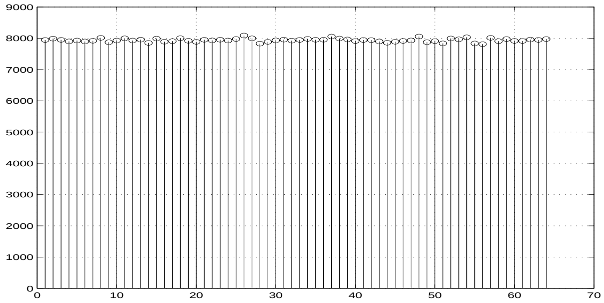

Figure 1: DC components of 64 × 64 block coefficients of scrambled image.

d_i is the i^{th} watermarking bit mapped to polar format, i.e., ±1. The coefficient $\tilde{A}_i$ replaces the original DC coefficient in the corresponding block. For each block, the watermark signal is viewed as a two-dimensional impulse signal with amplitude $A\alpha d_i$, with the impulse located at the origin, i.e., zero frequency. The inverse DCT gives the watermark signal $w(n_1, n_2)$ in the spatial domain. $w(n_1, n_2)$ has constant amplitude equal to $\frac{A\alpha d}{L}$, i.e., each pixel in the block will be perturbed by $\frac{A\alpha d}{L}$, where L is the block size. After embedding the watermark by amplitude modulating the DC component of the DCT blocks, we take the inverse DCT and unscramble the pixels locations using the same key to get the watermarked image $\tilde{x}(n_1, n_2)$.

The watermarked image is expected to undergoes different types of attacks. We model this process as a signal travelling through a noisy channel which adds noise to the watermarked image. Let $y(n_1, n_2)$ be the received watermarked image, which consist of the watermarked image $\tilde{x}(n_1, n_2)$ plus some noise denoted by $N(n_1, n_2)$, i.e.,

$$y(n_1, n_2) = \tilde{x}(n_1, n_2) + N(n_1, n_2). \qquad (2)$$

To recover the watermark the same key K is used to permute the received image. The permutation process decorelates the samples of the watermarked image and the samples of the additive noise. Therefore, the noise samples can be thought of as a sequence of uncorrelated random variables. The DCT of the received image is given by $Y(k_1, k_2) = \tilde{X}(k_1, k_2) + \mathcal{N}(k_1, k_2)$, after taking the DCT of the received image we extract the watermarked coefficients. Let $V(Y_i)$ be a vector which represent the received watermarking coefficients, which are modeled as independent and identically distributed random variables. We model each received watermarking coefficient Y_i as a random signal which consist of three parts. The first part is deterministic and it is equal to the DC component A. The second part is random and it is introduced by the watermarking sequence d which has amplitude equal to $A\alpha d_i$. The last part is $\mathcal{N}_i$, which is introduced by the processing noise. Since the random signal $\mathcal{N}_i$ is the sum of uncorrelated identically distributed random variables, by the central limit theorem [7] this signal follows Gaussian distribution with zero mean and variance $\sigma^2_{\mathcal{N}}$ which represent the noise power. Therefore, we can model the i^{th} received watermarked coefficient as a random variable consisting of a deterministic signal perturbed by an additive random signal which has two components, one is discrete due to the embedded watermark and the other is continuous Gaussian noise introduced by processing operations.

To evaluate the system performance, we determine the bit error rate (BER) due to image processing operations. To do that we first try to develop an expression for computing the BER. The expected value of any received watermarking coefficient is given by

$$E\{Y_i\} = E\{A + A\alpha d_i + \mathcal{N}_i\} = E\{A\} + E\{A\alpha d_i\} + E\{\mathcal{N}_i\}. \qquad (3)$$

Assuming the Gaussian noise is zero mean and that $E\{d\} = 0$ for i.i.d watermarking bits, it follows that $E\{Y_i\} = A$, which is the original value of the DC coefficient before being perturbed by the additive noise. Since A is much larger than the magnitude of the any watermarking bit and larger than the magnitude of the noise, we remove this average value to improve the detection of the watermark and minimize the BER. Since the technique is blind, the detector does not know the value of A and an estimate of its value must be performed first. This is obtained by computing the average of the DC values of all the received watermarking coefficients. This seems to give a good estimate for the actual value of A. Note that when using this approximation we are making the assumption that the received image after the decorrelation process can be modeled an ergodic stochastic process [7]. In that case, we make the assumption that time averages equal ensemble averages, i.e.,

$$E\{Y_i\} \approx \frac{1}{n}\sum_{i=1}^{n} Y_i = A. \qquad (4)$$

After removing the average value, the remaining signal $\hat{Y}_i$ consists of two independent random components, i.e., $\hat{Y}_i = Y_i - A = A\alpha d_i + \mathcal{N}_i$. These signals are independent. Therefore, their equivalent probability density function is given by the convolution of the two densities [7]. Assuming $d_i = \pm 1$ equally likely the pdf of the watermark has the following distribution

$$f_{\mathbf{d}}(d) = \frac{1}{2}\delta(x - \alpha A) + \frac{1}{2}\delta(x + \alpha A). \qquad (5)$$

Therefore, when we convolve this pdf with the Gaussian noise, the result is the sum of two Gaussian pdf's one with mean αA and the other with mean $-\alpha A$, i.e.,

$$\mathbf{f}_{Y_i}(\hat{y}) \approx \frac{1}{\sqrt{2\pi}\sigma_{\hat{y}}} e^{\frac{-(\hat{y}-\alpha A)^2}{2\sigma_{\hat{y}}{}^2}} + \frac{1}{\sqrt{2\pi}\sigma_{\hat{y}}} e^{\frac{-(y+\alpha A)^2}{2\sigma_{\hat{y}}{}^2}}, \qquad (6)$$

where $\sigma^2_{\hat{y}}$ is the variance which represent the noise power per block. The above analysis indicates that an estimate to the probability of bit error of the proposed watermarking system is reduced into a simple binary detection theory problem. In particular, we can model the watermarking system as a baseband binary digital communication system in which we model the received signal as a Gaussian random variable with mean αA and variance $\sigma_{\hat{y}}^2$. An expression for the probability of error for this standard communication theory problem can be shown to be [8]

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{(\alpha A)^2}{\sigma_{\hat{y}}^2}}\right). \tag{7}$$

From this equation it is clear that in order to minimize the probability of error, we need to maximize the energy in the watermarking bits. The energy in each watermarking bit is maximized either by increasing A or increasing the modulation index α. Increasing A, implies increasing the block size and that decreases the watermark payload. Given some fixed block size, the only way to increase the watermark power is by increasing the modulation index. However, increasing the modulation index is constrained by the perceptual quality of the watermarked image. Hence, given some perceptual level of distortion we increase the modulation index until we reach that level, i.e., we maximize α until $\delta(\tilde{x}, x) \leq \gamma$ where δ is some distortion measure and γ is the maximum allowable distortion. For example in [9] it was suggested that the peak signal to noise ratio (PSNR) between the original and the watermarked image due to watermark embedding should not exceed 38 dB.

Watermark extraction is done by first prefiltering the received image [5] to improve watermark detection, then the resulting image is decorrelated with the same key. Next, we Take block DCT of the permuted image and extract the DC value of each block. A vector $\mathbf{Y}$ is formed from those DC values. Each component Y_i of $\mathbf{Y}$ is compared with some threshold τ. The value of the extracted bits are set according to the following

$$Y_i = \begin{cases} 1 & \text{if} \quad Y_i \geq \tau \\ 0 & \text{if} \quad Y_i < \tau \end{cases} \qquad \text{where} \quad \tau = \frac{\sum_{i=1}^{k} Y_i}{k} \tag{8}$$

where k represent the number of DC blocks.

3 EXAMPLE AND RESULTS

To test the robustness of the technique to different attacks, we used several test images for evaluation, in this paper we show the result for three images: LENNA, BABOON, CAMERAMAN. All images are gray level of size 512×512. A DCT block of size 32×32 were used. The value of α varies between $0.03 - 0.04$, depending on the image texture. The watermark message size is 75 bits. The watermark is encoded using a rate $\frac{1}{3}$ convolutional encoder, to enhance the watermark robustness against image attacks. Furthermore, the encoded message was passed through an interleaver. The purpose of the interleaver is to reduce the probability of having a burst error. The average PSNR between the original and the watermarked images is about 38 dB. Figure 2 shows an example of the original LENNA image and the resulting watermarked image. To test the robustness of the proposed watermarking scheme, we ran two different types of experiments. The first test was the latest Stirmark benchmark test [10, 9]. The output results are shown in Table I, a score of "1" indicate a full recovery of the watermark for all levels of attacks "0" is given otherwise. For geometric attacks such as rotation, linear transformation and shearing we used template technique similar to the one suggested in [11] to resynchronize the image and recover the watermark. For high compression ratios such as below %20 quality factor we were not able to recover the watermark. Watermark robustness can be improved by increasing the modulation index, however, image perceptual quality will be effected by increasing the value of the modulation index. currently testing the robustness of the technique against other types of benchmarks such as the checkmark proposed in http://watermarking.unige.ch/Checkmark. Another experiment was conducted using the software Photoshop where the watermarked image was subjected to different types of geometric deformation. Figures 3 to 5 show examples of some geometric deformations which were applied to the watermarked image For all of these attacks a full recovery of the watermark was obtained despite the severity of the deformation. Recovery was obtained directly without the need to resort to the synchronization template or any searching techniques.

Figure 2: (a)Original Image. (b) Watermarked image.

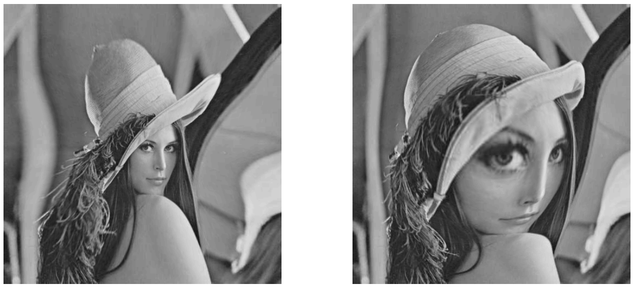

Figure 3: Watermarked image after pinching and spherizeing attacks.

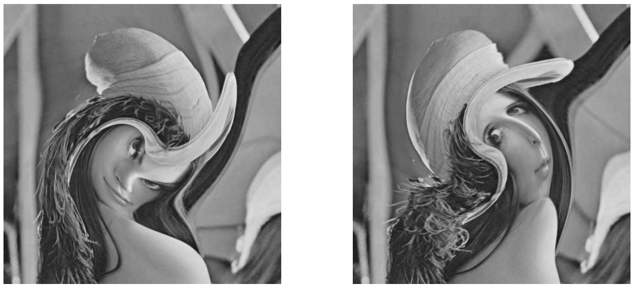

Figure 4: Watermarked image after twirling attacks.

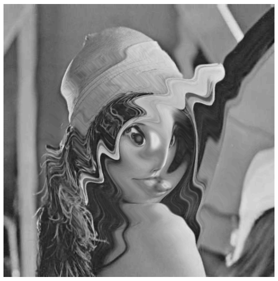

Figure 5: Watermarked image after zigzag attack.

4 Conclusion

In this paper, we presented a new secure, oblivious digital watermarking technique based on amplitude modulating the DC component of DCT blocks of a decorrelated image. The watermark is a sequence of binary digits channel encoded to improve its robustness. We have used the latest Stirmark benchmark test as our criterion for examining the robustness of the watermark. Furthermore, we tested the robustness of the technique against some geometric deformations. In all these test the technique indicates robustness against most of the attacks. Future work will concentrate on improving the over all robustness, in particular to attacks such as random bending and testing the technique against other attacks and apply the technique for colored images.

Robustness Test	Lenna	Baboon	Camera
columns & rows removal	1	1	1
rotation	1	1	1
JPEG compression	1	1	1
JPEG below 20% Q factor	0	1	0
cropping	1	1	1
rotation-scale	1	1	1
FMLR	1	1	1
filtering	1	1	1
Sharpening-3-3	1	1	1
scaling	1	1	1
aspect ratio	1	1	1
shearing	1	1	1
linear transformation	1	1	1
Stirmark random bend	0	1	0

Table 1: Scores after applying Stirmark benchmark test.

References

[1] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, “Multimedia data embedding and watermarking technologies,” *Proc. of IEEE*, vol. 86, no. 6, pp. 1064–1087, June 1998.

[2] F. Hartung and M. Kutter, “Multimedia watermarking techniques,” *Proc. Of The IEEE*, vol. 87, no. 7, pp. 1079–1107., July 1999.

[3] Ping. W. Wong and Edward J. Delp, *Security and Watermarking of Multimedia Contents II, Volume 3971*, Proceeding of SPIE, Bellingham, Washington, 2000.

[4] Ping. W. Wong and Edward J. Delp, *Security and Watermarking of Multimedia Contents III, Volume 4314*, Proceeding of SPIE, Bellingham, Washington, 2001.

[5] G. Langelaar, I. Setyawan, and R. Lagendijk, “Watermarking digital image and video data,” *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 20–46, Sept 2000.

[6] F. Alturki and R. Mersereau, “Secure image transform domain technique for steganographic applications.,” *Proc. of IS & T/SPIE: Electronic Imaging III International Conference on Security and Watermarking of Multimedia contents ,San Jose, California*, January 2001.

[7] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering,* 2^{ed} *edition,* Addison Wesley, Massachusetts, U. S. A, 1994.

[8] S. Haykin, *Communication Systems,* 3^{rd} *edition*, John Wiley, New York, 1994.

[9] M. Kutter and F. Petitcolas, “A fair benchmark for image watermarking systems,” *in (Wong and Delp, eds), Proceeding of the SPIE, Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 226–239, Jan 1999.

[10] F. Petitcolas, R. Anderson, and M. Kuhn, “Attacks on copyright marking systems,” *in (Aucsmith, eds) Information Hiding: Second International Workshop, Lecture Notes in Copmuter Science,volume 1525*, pp. 218–238., April 1998.

[11] M. Kutter, *Digital Image Watermarking: Hiding Information in Images*, Ph.D. thesis, Ecole Polytechnique Federale De Lausanne, 1999.