AN ASYMMETRIC WATERMARKING METHOD FOR COPYRIGHT PROTECTION UTILIZING DUAL BASES

†*Jengnan Tzeng,* ‡*Wen-Liang Hwang, and* †*I-Liang Chern*

†Department of Mathematics, National Taiwan University
‡Institute of Information Science, Academia Sinica, Taiwan

ABSTRACT

We present an asymmetric watermarking method for copyright protection that uses different matrix operations to embed and extract a watermark. It allows for the public release of all information, except the secret key. We investigate the conditions for a high detection probability, a low false positive probability, and the possibility of unauthorized users successfully hacking into our system. The robustness of our method is demonstrated by the simulation of various attacks.

1. INTRODUCTION

Digital security information embedded in content, called watermarking, has many applications, including authentication, copyright protection, copy protection, fingerprinting, and broadcasting channel tracking [4, 10, 14].

Notable security problems of the symmetric watermarking approach(i.e., one secret key for encoding and decoding) stem from the need to make the secret key available to owners and recipients, as well as from the need to identify which secret key is associated with which image in a large image database. Another problem is that the watermark is present as evidence of ownership, so it provides an attacker with the knowledge to remove the watermark [2]. The solution to the problem is a watermarking system that satisfies Kirckhoffs'principle [9], which states that a security system must assume that an adversary knows everything about the algorithm, except the secret keys.

Asymmetric watermarking is another approach that satisfies Kerckhoffs' principle. This system uses two sets of keys: one for embedding, and one for detecting. The latter is made public, so anyone has access to it and is permitted to use it to verify whether an image is watermarked or not. Some interesting asymmetric schemes have been proposed for watermarking [8, 11, 12, 5, 6, 13, 7]. Hartung and Girod [8] proposed the first asymmetric watermarking method. Furon and Duhamel [7] provided a useful survey of various methods, as well as an in-depth discussion of asymmetric watermarking.

In our previous study of symmetric watermarking, we proposed a robust subspace watermarking method. Based on that method, we have modified the detection approach so that the new method retains the robustness property and becomes an asymmetric watermarking method.

Section 2 provides a summary of our previous subspace symmetric watermarking method. Section 3 illustrates how we have extended it to develop our asymmetric method. In Section 4, we describe an attack scenario called projection attack and show how to avoid it by a specially designed detection matrix. The simulation results of various attacks are demonstrated in Section 5. Finally, in Section 6, we present our conclusions.

2. THE SYMMETRIC SUBSPACE WATERMARKING METHOD

In [15], we proposed a subspace symmetric watermarking method for copyright protection. The method, which models watermarking as a communication with side information [4], makes the keys heavily dependent on the original image and on potential modifications of the watermarked image. The robustness of the approach lies in hiding a watermark in the subspace that is least susceptible to potential modifications. The distribution of the features of forged images is derived by principal component analysis of the simulation of images attacks. One of the subspaces of the feature space is called the watermark space $\mathcal{W}$, in which the watermark is hidden. The orthogonal complement of $\mathcal{W}$ is denoted as $\mathcal{V}$, representing the subspace that is most susceptible to modifications of the image. This approach allows a copyright owner to custom-select the watermark space that is most resistant to possible attacks.

Let ϕ_o be the feature of the original image. Watermark w is embedded into $\mathcal{W}$ by:

$$\phi_w = \phi_o + Gw,$$

where G is a secret matrix whose columns are a basis of $\mathcal{W}$, and ϕ_w is the feature of the watermarked image. Because watermark w is in $\mathcal{W}$, the watermark is robust against

0-7803-9134-9/05/$20.00

possible attacks. A pirate can simulate attacks on our watermarked image and obtain a good approximation of space $\mathcal{W}$, but he cannot detect the secret matrix G from the space.

Our symmetric method uses the key G to embed, and its inverse G^T to extract, watermark w. By choosing G such that $G^T\phi_o = 0$, the method does not need a reference image to detect a watermark. The key is content-dependent; therefore, when the number of watermarked images is large, there are problems that copyright owners need to manage so that the correct key of a watermarked image can be located. It is also necessary to secretly communicate the key to another party. In an asymmetric watermarking method, a verifi erdoes not need exclusive permission to access a published key database, which reduces the key management effort. Also, anyone can prove copyright of a watermarked image without secret key communication.

3. THE ASYMMETRIC WATERMARKING METHOD

Following the previous symmetric watermarking method, we divide our feature space into subspaces $\mathcal{W}$ and $\mathcal{V}$. The difference from our symmetric method is that we further divide $\mathcal{W}$ into two orthogonal subspaces $\mathcal{G}$ and $\mathcal{H}$. Let G and H denote the secret matrices whose columns form a basis of subspaces $\mathcal{G}$ and $\mathcal{H}$ respectively. We use matrix G to embed our watermark w into subspace $\mathcal{G}$ and detect w by using the published keys (D, w), where matrix D is a weighted mixture of the matrices H and G.

3.1. Encoding and Decoding

Embedding our watermark w into subspace $\mathcal{G}$ is achieved by the function

$$\phi_w \quad = \quad \phi_o + Gw. \tag{1}$$

We require the watermark strength $||w||$ to be as large as possible, in order to obtain a high signal-to-noise ratio (SNR) of our watermark signal w to the original image feature ϕ_o. However, $||w||$ should not be so large that the perceptual quality of the watermarked image is degraded. Finally, a feature reconstruction function is applied to ϕ_w to obtain a watermarked image X_w.

Our detection is a hard decision function δ with a threshold ϵ. The decision function applies the detection matrix D to the extracted feature ϕ_e and then uses the sim function to measure the similarity between $D\phi_e$ and w. Our detector is:

$$\delta(\phi_e) = \begin{cases} 1 & \text{if } |sim(w, D\phi_e)| \geq \epsilon, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where

$$sim(w, D\phi_e) = \frac{w^T D\phi_e}{||w||||D\phi_e||}. \tag{3}$$

We give the matrix D the form:

$$D = G^T + BH^T, \tag{4}$$

where B is a matrix; H is a matrix, whose columns are a basis of $\mathcal{H}$; and $H^T G = 0$.

We have shown that a pirate who has (D, w) and the algorithm of our watermarking method does not have the knowledge to obtain Gw. In the next section, we propose that the design of B is important to the security issue.

4. PROJECTION ATTACK AND SPECIAL DESIGN OF THE DECTION MATRIX D

We evaluate the security threats of malicious attacks on our watermarking system. One type of effi cientattack is called projection attack, which tries to fi ndthe feature $\tilde{\phi}$ that satisfi es

$$\min_{\phi} \|\phi - \phi_w\|^2,$$

with the constraint $w^T D\phi = 0$ [1]. This means $\tilde{\phi}$ is the feature without a watermark that is closest to ϕ_w. As a projection attack is extremely effective in removing a watermark, we pay particular attention to it.

We claim that if the detection matrix is derived such that

$$D\phi = 0,$$

then our asymmetric watermarking method is totally threatened by the projection attack. Therefore, we need to design the detection matrix D in such a way that our asymmetric watermarking method can resist a projection attack.

The following theorem shows that we can construct a special matrix D so that the projection attack yields σ_o. Because ψ_o is the perceptually robust feature of the original image, there is a high probability that the image reconstructed from $\sigma_o \in \mathcal{V}$ will be perceptually distorted.

Theorem Given G, H and $\phi_o = \psi_o + \sigma_o$i, where ψ_o is a component of ϕ_o in $\mathcal{W}$. We defi ne $m_o = D\phi_o$, $\psi_w = \psi_o + Gw$ and the coeffi cientsof s, t, such that s, t satisfy $\psi_o = Gs + Ht$,

(a) To avoid a projection attack, the detection matrix D (defi nedin Equation 4) must be chosen such that $D\phi_o \neq 0$.

(b) If D is chosen such that

$$D^T w = \lambda\psi_w, \tag{5}$$

where $\lambda \neq 0$, then applying the projection attack to ϕ_w (defi nedin Equation 1) obtains σ_o.

(c) If D and w satisfy Equation 5, then

$$w = \frac{\lambda}{1-\lambda}s, \tag{6}$$

[1] The real constraint is $|sim(W, D\phi)| < \epsilon$, where ϵ is the threshold. We use a simplifi ed constraint so that the attack can be analyzed.

where $\lambda \neq 1$.
(d) If B is constructed as

$$B = \frac{(1-\lambda)}{\|s\|^2} st^T + \sum_{i,j} c_{i,j} u_i v_j^t, \quad (7)$$

where $u_i \perp s$, $v_j \perp t$, $c_{i,j}$ is a real number, and w satisfies Equation 6, then D satisfies Equation 5.
(e) If w and D are chosen according to Equations 6 and 7 respectively, then w is parallel to m_o.

(f) If B satisfies Equation 7, and $\lambda = 1 + \frac{\|s\|^2}{\|t\|^2}$, then $D\phi_o = 0$.

Figure 1 shows that the projection attack on our watermarked image makes the image perceptually unacceptable. Thus, our watermarking method is secure under projection attack.

Fig. 1. The image obtained from the feature extracted by applying a projection attack to the watermarked image. The PSNR of the noise image, obtained by subtracting the bottom image from the top image, is 17 dB.

5. SIMULATION RESULTS AND ROC CURVE

We now demonstrate the resistance of our asymmetric watermarking method to the following attacks.
Spreading Noise into a Watermark Space. The simulation results in [15] indicate that a pirate can simulate attacks on our watermarked image and obtain a good approximation of space $\mathcal{W}$, but he cannot obtain the secret matrix G from the space. In this scenario, we evaluate the efficiency of an attack on our watermark space $\mathcal{W}$ by jamming it with random noise. We embed 64 random noises that have various levels of energy into the watermark space of a watermarked Lena image. Performance results for this attack are shown in Figure 3. We plot the mean, obtained by averaging the detection values of the 64 random noise attacks on the $\mathcal{W}$ space, versus the SNR that is measured by $20 \log_{10} \frac{||w||}{||n||}$, where n is our random noise. One can observe from the figure that even at a very low SNR, the detection value is still quite high compared to our threshold. This proves that our method is robust against this type of attack.
Blind Attacks Blind attacks are carried out with the intention of removing a watermark when the attacker doesn't know the watermarking method. For each of the 61 images in our database, we produce 32 watermarked images and performe an average of 100 attacks on each image. These attacks include: shifting, blurring, JPEG compression, sharpening, rotation, stirmarking, and combinations of these attacks. The means of the $|sim|$ values are larger than 0.9 and most of the standard deviations of the $|sim|$ values are smaller than 0.1.
ROC Curve As proposed in [3], we model the detection probability and false positive probability as Gaussian distributions. We compute the mean and the standard deviation of the Gaussian distribution of the false positive probability from the detected values of the un-watermarked images. In the same way, we compute the parameters of the detection probability for the watermarked images. From the Gaussian distributions, we can draw the ROC curve of our empirical data. Figure 2 shows the ROC curves of different c values obtained in this manner. We choose $c = 0.1$ and use 0.5 as our threshold. This corresponds to the false positive probability below 10^{-5} in our simulation. For numerical precision, the figureshows only the parts of the curves whose false positive probability is above 10^{-15}. The intersections of the curves and the axis of false positive probability (x-axis) are 0.34, 0.50, and 0.56 for curves of $c = 0.15$, $c = 0.1$, and $c = 0.05$, respectively. From the distribution of the ROC curves, it is clear that our asymmetric watermarking method has the same robustness as our symmetric watermarking method.

6. CONCLUSION

To resolve the weaknesses of current symmetric watermarking methods, we have designed an asymmetrical watermarking method for copyright protection that satisfies the zero knowledge principle. All of our watermarking operations, except the secret matrices G and H, have been released and are publicly available. Our asymmetric design is robust because it enhances the watermark space concept of our previous symmetric watermarking method. As our watermark is heavily dependent on the original image, it cannot be removed without the watermarked image being perceptually distorted. Our method is secure, since we embed secret information Gw within a subspace of $\mathcal{W}$, and provide the public with a key ($D = G^T + BH^T$) to detect Gw. However, because the secret basis of $\mathcal{G}$ is hidden from the public, estimating Gw is extremely difficult.

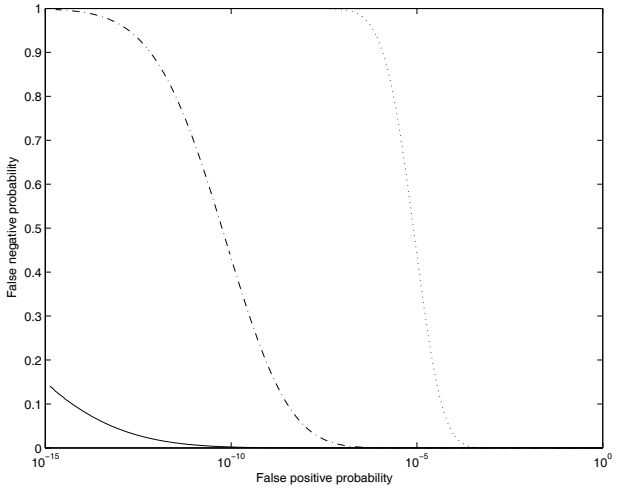

Fig. 2. ROC curves of our empirical data. The curves plot the false positive probability in the logarithmic (base 10) scale against the false negative probability, which is defined as one minus the detection probability, as a function of the threshold and c value. The solid curve corresponds to $c = 0.15$, the dash-dot curve to $c = 0.1$, and the dotted curve to $c = 0.05$.

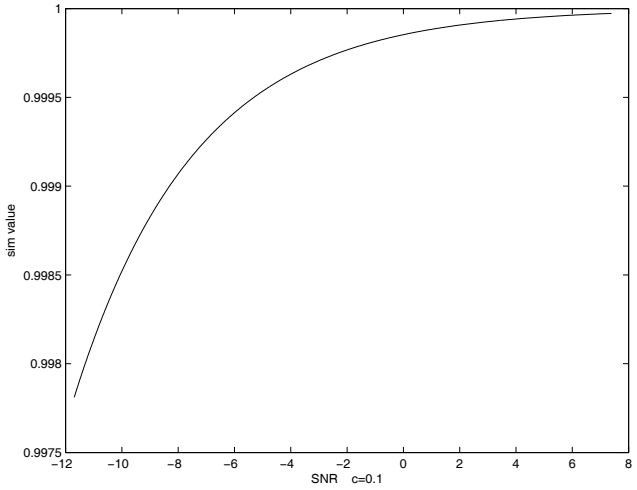

Fig. 3. A white noise spreading attack on $\mathcal{W}$.

7. REFERENCES

[1] M. Barni, F. Bartolini, T. Furon, "A general framework for robust watermarking security," *Signal Processing*, 83, pp. 2069-2084, 2003.

[2] S. Craver, "Zero knowledge watermark detection," *Proceedings of the Third International Workshop on Information Hiding*, vol. 1768 of Lecture Notes in Computer Science, pp. 101-116, Spring 2000.

[3] I. Cox, M. Miller, and J. Bloom, "Digital Watermarking," *Morgan Kaufmann Publishers*, pp. 173-177, 2002.

[4] I. Cox, M. Miller, and A. Mckellips, "Watermarking as communication with side information," *Proc. IEEE*, vol. 87, pp. 1127-1141, July 1999.

[5] J. Eggers, J. Su, and B. Girod, "Public key watermarking by eigenvectors of linear transforms," *Proc. Eur. Signal Process. Conf.*, Tampere, Finland, Sept. 2000.

[6] T. Furon, I. Venturini, and P. Duhamel, "An unified approach of asymmetric watermarking schemes," *Proceeding of SPIE: Security and Watermarking of Multimedia Contents III*, P. W. Wong and E. Delp, Eds, San Jose, U.S.A., Jan. 2000.

[7] T. Furon and P. Duhamel, "An asymmetric watermarking method," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 981-995, April 2003.

[8] F. Hartung and B. Girod, "Fast public-key watermarking of compressed video," *Proc. IEEE Int. Conf.Image Process.*, Oct. 1997.

[9] A. Kerckhoffs, "La cryptographie militaire," *J. Sci. Militaires*, vol. 9, pp. 5-38, Jan. 1883.

[10] P. Moulin and A. Ivanovic, "The zero-rate spread-spectrum watermarking game," *IEEE Trans. on Signal Processing*, vol.51, no. 4, pp. 1098-1117, April 2003.

[11] R. Van Schyndel, A. Tirkel, and I. Svalbe, "Key independent watermark detection," *Proc. Int. Conf. Multimedia Comput. Syst.*, vol. 1, Florence, Italy, June 1999.

[12] J. Smith and C. Dodge, "Developments in steganography," *Proc. of Third Int. Workshop on Information Hidding,* A. Pfitzmann, Ed. Dresde, Germany, pp. 77-87, Sept. 1999.

[13] J. Stern and J. P. Tillich, "Automatic detection of a watermarked document using a private key," *Proc. 4th Int. Workshop Information hiding,* vol. 2137, I. S. Moskowitz, Ed., Pittsburg, PA, Apr. 2001.

[14] W. Trapper, M. Wu, Z. Wang, K. J. R. Liu, "Anti-Collusion fingerprintng for multimedia," *IEEE Trans. on Signal Processing*, vol. 51, no.l 4 pp. 1069-1087, April 2003. Special Issue on Signal Processing for Data Hiding in Digital Media & Secure Content Delivery.

[15] J. Tzeng, W. L. Hwang, and I. Liang Chern, "Enhancing image watermarking methods with/without reference images by optimization on second order statistics," *IEEE Trans. on Image Processing*, vol. 11, pp. 771–782, July 2002.