

# ENHANCING IMAGE WATERMARKING METHODS BY SECOND ORDER STATISTICS

††Jengan Tzeng, †Wen-Liang Hwang, and ‡I Liang Chern

†Institute of Information Science, Academia Sinica, Taiwan

‡Department of Mathematics, National Taiwan University, Taiwan

## ABSTRACT

A pirate attack on an image aims to create an invisible modification of it. We propose a watermarking strategy which applies the optimization techniques to the second order statistics of perceptually unaltered modifications of an image. The solution gives two orthogonal spaces. One of them characterizes most of the variations in the modification of the image. Our watermark is embedded in the other space that most potential pirate attacks do not touch. Thus, the embedded watermark is robust. We also show that our method is able to enhance many existing watermarking strategies. We also demonstrate the performance of our method.

## 1. INTRODUCTION

Digital signatures embedded in images, called “watermarks”, are important for protecting copyright ownership in many multimedia applications [3][4][6]. Usually, watermarking methods are classified into two types: visible and invisible watermarks. We will focus our discussion on invisible watermarks. Watermark *robustness* particularly refers to the ability to detect an embedded watermark in an image even when the image is modified by means of image operations. Despite much previous research, watermark robustness is still a worthwhile topic with plenty of unknown issues. A very nice survey of watermark attackings is given in [5].

We propose a watermarking method which uses optimization methods to embed invisible watermarks in images. A pirate attack on an image aims to create an invisible modification of it by means of image operations. Like many previous researches, we embed our watermark using features. The features can be obtained from DCT coefficients, wavelet coefficients, spatial patterns *etc.* By Monte Carlo simulation, we assume that the random variable  $\underline{\epsilon}$ , corresponding to the features of perceptually unaltered modifications of an image, is obtainable. Then, we can calculate from the simulated data the statistics of the random variable; thus, we can statistically characterize the pirate

attack on the image. By introducing an objective function of the first two statistics of the random variable, we optimize the function for the space used to embed our watermark. Our solution shows that the random variable  $\underline{\epsilon}$  can be partitioned into two spaces orthogonal to each other. One of them, called  $V$ , has most of the variations of perceptually unaltered modifications of the image. As a consequence, its complementary space, called  $W$ , has less chance of being modified by potential attackers. In other words, embedding a watermark in space  $W$  is more robust since doing so offers a higher probability of protecting the watermark against pirate attacks. To demonstrate its usefulness, we apply our method to Cox *et al.*'s method [1] by embedding a watermark in the appropriate subspace,  $W$ , of the DCT features.

In section 2, we will specify our problem in a very general setting. In section 3, we will present our watermarking method. Finally, we will give conclusions in the last section.

## 2. PROBLEM MODEL

Given a host image  $X$ , of size  $N = m \times n$  pixels, one can experiment on  $X$  by performing elementary image processing operations, such as translation, rotation, smoothing, compression *etc.*, or by combining elementary operations to obtain a perceptually unaltered image of  $X$ . We use the random variable  $\underline{X}$  for images obtained by means of such modifications of  $X$ . After performing enough experiments, we are able to obtain the statistics of the random variable  $\underline{X}$ .

Let us expand  $\underline{X}$  against the bases  $\{\Phi_{i,j} | i = 1, \dots, m, j = 1, \dots, n\}$ , whose dual bases are  $\{\tilde{\Phi}_{i,j} | i = 1, \dots, m, j = 1, \dots, n\}$ . We have

$$\underline{X} = \sum_{i,j} \langle \underline{X}, \Phi_{i,j} \rangle \tilde{\Phi}_{i,j}, \quad (1)$$

where  $\langle \underline{X}, \Phi_{i,j} \rangle$  is the inner product of  $\underline{X}$  and the basis function  $\Phi_{i,j}$ .

Let  $X^M = X + M$  be the watermarked image obtained by adding watermark  $M$  to image  $X$ . As be-

fore, we use  $\underline{X}^M$  to denote the invisible modifications of  $X^M$  by means of elementary image operations and their combinations:

$$\underline{X}^M = \sum_{i,j} \langle X, \Phi_{i,j} \rangle \tilde{\Phi}_{i,j} + \sum_{i,j} (\langle X^M - X, \Phi_{i,j} \rangle + \langle \underline{X}^M - X^M, \Phi_{i,j} \rangle) \tilde{\Phi}_{i,j}.$$

If we use inner product coefficients as our features; then,  $[\langle X, \Phi_{i,j} \rangle]$  represents the features of the host image and

$$\mathbf{m} = [\langle X^M - X, \Phi_{i,j} \rangle] \quad (2)$$

represents the watermark features added by copyright owners for identification of the ownership of  $X$ . The last term on the right,  $[\langle \underline{X}^M - X^M, \Phi_{i,j} \rangle]$ , has a special meaning; it represents the features, away from those of  $X^M$ , probably introduced by a pirate attack. We denote the perturbation from  $X^M$  as

$$\underline{\mathbf{e}}^M = [\langle \underline{X}^M - X^M, \Phi_{i,j} \rangle] \quad (3)$$

and the centered perturbation from  $X^M$  as

$$\underline{\mathbf{e}}^M = [\langle \underline{X}^M - X^M, \Phi_{i,j} \rangle - \langle E\{\underline{X}^M\} - X^M, \Phi_{i,j} \rangle]. \quad (4)$$

In order to identify the ownership of an image, one usually measures the similarity between the watermark descriptor  $\mathbf{m}$  provided by the owners and the watermark descriptor  $\mathbf{t}$  extracted from a test image. A frequently used measurement is the cosine angle of the two vectors:

$$\text{sim}(\mathbf{m}, \mathbf{t}) = \frac{|\mathbf{m}'\mathbf{t}|}{\|\mathbf{m}\| \|\mathbf{t}\|}.$$

We say that the two vectors are similar if their *sim* value is close to 1; or that the two are not similar if the value is close to 0. Let  $\mathbf{t}$  be extracted. We are concerned with the design of the watermark descriptor  $\mathbf{m}^*$  such that for most of  $\mathbf{t}$ , extracted from a perceptually unaltered modification of the watermarked image,  $\text{sim}(\mathbf{m}^*, \mathbf{t})$  will be as large as possible.

### 3. OUR WATERMARK DESIGN

When a reference image is provided, we use the following objective function to resolve the ownership:

$$\text{sim}(\mathbf{m}, O_l(\mathbf{m} + \underline{\mathbf{e}}^M)) = \frac{|\mathbf{m}'O_l(\mathbf{m} + \underline{\mathbf{e}}^M)|}{\|\mathbf{m}\| \|O_l(\mathbf{m} + \underline{\mathbf{e}}^M)\|}, \quad (5)$$

where  $O_l$  is a linear operator and  $\underline{\mathbf{e}}^M$  and  $\mathbf{m}$  were given in Eq.(3) and Eq.(2), respectively. We aim to find a

watermark feature  $\mathbf{m}^*$  and the linear operator  $O_l$  such that the results of Eq.(5) is large.

Suppose our feature space is  $R^N$  and our watermark feature  $\mathbf{m}^*$  lies in a subspace of it,  $\underline{\mathbf{e}}^M$  can be rewritten as

$$\underline{\mathbf{e}}^M = \underline{\alpha}\mathbf{m}^* + \underline{\mathbf{v}},$$

where  $\underline{\alpha}$  is a scalar random variable, obtained by projecting  $\underline{\mathbf{e}}^M$  onto  $\mathbf{m}^*$ , and  $\mathbf{m}^*$  and  $\underline{\mathbf{v}}$  are perpendicular to each other. Let  $W$  be the subspace of  $R^N$  such that the projection of most of realizations of  $\underline{\mathbf{v}}$  to  $W$  is small. If we let  $O_l$  to be the projection of a vector to the subspace  $W$ , denoted as  $P_W$ ; then, we have

$$\begin{aligned} \text{sim}(\mathbf{m}^*, P_W(\mathbf{m}^* + \underline{\mathbf{e}}^M)) &= \text{sim}(\mathbf{m}^*, (1 + \underline{\alpha})\mathbf{m}^* + P_W(\underline{\mathbf{v}})) \\ &\approx \text{sim}(\mathbf{m}^*, (1 + \underline{\alpha})\mathbf{m}^*) = 1. \end{aligned}$$

Hence, most of perceptually unaltered piracy attacks on  $X^M$  have a large *sim* value and are, thus, detectable. Since  $\underline{\mathbf{e}}^M$  is a random variable, we will resort to statistics to find the optimal watermark descriptor  $\mathbf{m}^*$  and the subspace  $W$ . We use the second order statistics to obtain our solution:

$$\min_{\mathbf{m}} E\{(\mathbf{m}'\underline{\mathbf{e}}^M)(\mathbf{m}'\underline{\mathbf{e}}^M)'\}, \quad (6)$$

where  $\underline{\mathbf{e}}^M$  is the centered perturbation of  $X^M$  given in Eq.(4). Figure 1 gives a simple schematic presentation of our motivation for using the second order statistics to find the subspace  $W$ . We see that most of the variations of  $\underline{\mathbf{e}}^M$  lie in subspace  $V$ . Thus, if we embed our  $\mathbf{m}^*$  to  $V$ 's complementary subspace  $W$ ; then, most of  $P_W(\underline{\mathbf{v}})$  will be small.

By simple calculation, we have

$$\begin{aligned} E\{(\mathbf{m}'\underline{\mathbf{e}}^M)(\mathbf{m}'\underline{\mathbf{e}}^M)'\} &= \mathbf{m}'U\Sigma U'\mathbf{m} \\ &= \sum_{i=1}^N \sigma_i^2 (\mathbf{m}'\mathbf{u}_i)^2, \quad (7) \end{aligned}$$

where  $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$ , in which the eigenvalues  $\sigma_i^2$  are arranged in decreasing order of magnitude and  $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N]$  is an orthonormal matrix containing the principle components of the covariance matrix of  $\underline{\mathbf{e}}^M$ . If we separate the eigenvalues into zero and non-zero components:

$$\begin{aligned} Z &= \{\sigma_i^2 | \sigma_i^2 = 0; i = d+1, \dots, N\}, \\ NZ &= \{\sigma_i^2 | \sigma_i^2 > 0; i = 1, \dots, d\}, \end{aligned}$$

and  $Z$  is not an empty set; then, we can obtain a simple optimization of Eq. (6) by assigning our watermark descriptor(feature)  $\mathbf{m}^*$  according to

$$\mathbf{m}^{*'}\mathbf{u}_k = \begin{cases} 0 & \text{if } \sigma_k^2 \in Z \\ \text{arbitrary number} & \text{if } \sigma_k^2 \in NZ \end{cases}$$

and  $\sum_k (\mathbf{m}^* \mathbf{u}_k)^2 = \|\mathbf{m}^*\|^2 = Cst$ .

The optimal watermark descriptor  $\mathbf{m}^*$  is characterized by the eigenvalues of the covariance of  $\mathbf{e}^M$ . If the variations are concentrated mostly in the subspace  $V$  spanned by a few eigenvectors and whose corresponding eigenvalues take a large proportion of the total variance of  $\mathbf{e}^M$ ; then, we expect that the image distortions due to pirate attacks, the intention of which is to make other perceptually unaltered versions of the host image, will tend to produce images with features in subspace  $V$ . Thus, by superimposing a watermark with features mainly on the complementary subspace of  $V$ , called  $W$ , we can obtain a robust watermark descriptor in the sense that this watermark descriptor has less chance of being erased by pirate attacks. In implementation, we obtain the subspaces  $V$  and  $W$  by thresholding the total variation  $\sum_i \sigma_i^2$  against a given number  $p \in (0, 1)$ : Let  $V$  be the space spanned by the eigenvectors; the square sum of their corresponding eigenvalues is more than  $p$  percentage of the total variation, and the space spanned by the rest of the eigenvectors is  $W$ . In Fig. 1, a schematic diagram of our watermarking strategy, the dimension of the features is reduced to 2 for demonstration purposes.

### 3.1. Watermark Encoding

The proposed watermarking strategy employs the following encoding algorithm. First, we choose bases  $\{\Phi_{i,j}\}$  for the host image  $X$ . Then, we find the perturbation from  $X$ , given as

$$\mathbf{e} = [\langle X - X, \Phi_{i,j} \rangle], \quad (8)$$

through simulations on  $X$  using elementary image operations and their combinations, such as translation, smoothing, rotation, compression, pasting *etc.* Then, we use singular value decomposition and obtain

$$Cov\{\mathbf{e} \ \mathbf{e}'\} = U \Sigma U',$$

where the matrix  $U = \{\mathbf{u}_1, \dots, \mathbf{u}_N\}$  is an unitary matrix,  $\Sigma = diag(\sigma_1^2, \dots, \sigma_N^2)$  is a diagonal matrix and the  $N$  is the image size. The eigenvalues  $\sigma_i^2$  are arranged in decreasing magnitude order. Now we choose a percentage  $p$  of the total variation  $\sum_i \sigma_i^2$  and separate the space of  $U$  into spaces  $V$  and  $W$  such that space  $V$  takes at least  $p$  percentage of the total variation and the rest is in space  $W$ . Spaces  $V$  and  $W$  are orthogonal to each other. Our watermark descriptor  $\mathbf{m}^*$  lies in space  $W$  such that

$$\mathbf{m}^* \mathbf{u}_k = \begin{cases} 0 & \text{if } \mathbf{u}_k \text{ in } V \\ \text{arbitrary number} & \text{if } \mathbf{u}_k \text{ in } W. \end{cases}$$

Let  $\mathbf{m}^* = [m_{i,j}^*]$ . Our watermark and watermarked image are, respectively,

$$M^* = \sum_{i,j} m_{i,j}^* \tilde{\Phi}_{i,j}$$

and

$$X^{M^*} = X + M^*.$$

### 3.2. Watermark Decoding

Given a test image  $T$ , to determine the ownership of the image, we first subtract the reference image, which in our implementation is  $X$ , from the test image  $T$  and represent the resultant image using the bases  $\{\Phi_{i,j}\}$ . The extracted feature  $\mathbf{t}$  is

$$\begin{aligned} \mathbf{t} &= [\langle T - X, \Phi_{i,j} \rangle] \\ &= \mathbf{m}^* + [\langle T - X^{M^*}, \Phi_{i,j} \rangle]. \end{aligned}$$

We then test  $sim(\mathbf{m}^*, \mathbf{t})$  against a threshold and verify the owner of the test image  $T$  when this value is greater than the threshold.

## 4. EXPERIMENTAL RESULTS

In this section, we will demonstrate our watermarking methods and compare our results with those obtained by J.Cox *et al.* [1]. We applied DCT to a Lena image of size 256 by 256. We then selected DCT coefficients from the upper left 32 by 32 corner, corresponding to the combinations of 32 low frequency bands horizontally and 32 low frequency bands vertically. We used the magnitude of these coefficients, so our vector space had dimension 1024. Then, we operated on the Lena image such that the resultant images and the original looked alike. Our operations included: blurring, compression with EZW or JPEG, small rotations (by  $\pm 1^\circ, \pm 2^\circ$ , and  $\pm 3^\circ$ ), small translations (by shifting 1 or 2 pixels either up, down, left and right), printing out and scanning in the printed-out image, and then grabbing the resulting image from the display, applying geometrical deformations. Totally, we obtained 125 such images. Most of our operations were carried out using the software program *xv* in the UNIX environment.

We then computed the covariance matrix from the collections of the features obtained from these images. Using PCA, we divided the feature space into subspace  $W$  and its complementary subspace  $V$ . The PCA results show that most of the eigenvalues are close to 0, and that the typical dimension of space  $W$  is about 950, given a total of dimension 1024. We studied the performance of our scheme

Table 1 lists the results obtained using both of our watermark methods and using Cox's method. Cox's method uses DCT coefficients as features. We also use the DCT coefficients as our features. However, we enhance the robustness of the Cox's method by selecting a subspace in which to embed the watermark feature. In Table 1, the first column lists the operations applied to the watermarked Lena image. The rest of the columns list the *sim* values obtained using the methods proposed by J. Cox *et al.* and using ours, respectively. One can see from the numbers that our methods are more effective than theirs and robustness of it is improved.

## 5. CONCLUSION

We have proposed an optimal solution for embedding a watermark. The proposed method can be used to enhance many existing watermarking strategies, as we have shown for the method proposed by Cox *et al.*. Our methods provide the best space for embedding a watermark if the first two ordered statistics are used. It may also be possible to find a solution using higher order statistics.

## 6. REFERENCES

- [1] I. J. Cox, J., Kilian, T. Leighton, and T. Shamon, Spread Spectrum Watermarking for Multimedia, *IEEE Trans. on Image Processing*, Vol. 6, pp. 1673-1687, 1997.
- [2] S. Craver, N. Memon, B. L. Yeo, and M. M. Yeung, Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications, *IEEE Journal on Selected Areas in Communications*, pp. 573-586, May 1998.
- [3] F. Hartung and M. Kutter, Multimedia Watermarking Techniques, *Proceedings of the IEEE*, Vol. 87, No. 7, July 1999.
- [4] C. T. Hsu and J. L. Wu, Hidden Digital Watermarks in Images, *IEEE Trans. on Image Processing*, Vol.8, No.1, January 1999.
- [5] M. Kutter, Digital Watermarking, *CVGIP 2000*, Invited Lecture, Taipei, Taiwan.
- [6] P. Moulin and J.A. O'Sullivan, Information-Theoretic Analysis of Watermarking, *IEEE ICASSP 2000*, Istanbul, Turkey, June 2000.
- [7] I. Pitas, A Method for Watermark Casting on Digital Images, *IEEE Trans. on Circuit and Systems for Video Technology*, Vol. 8, No. 6, October 1998.

Operations	J.Cox	Ours
Jpeg(65%)	0.8877	0.9414
Jpeg(50%)	0.8672	0.9411
Blur(3)	0.6319	0.9502
Blur(5)	0.2862	0.9482
Sharpen(75)	0.4437	0.9393
Scanning	0.0530	0.3459
StirMark	0.0877	0.8335
Rotation(0.5)	0.2992	0.9236
Rotation(-0.02)	0.4127	0.9406
Spread(5)	0.0427	0.5795
Translation	0.2790	0.9716
EZW(0.25bpp)	0.4697	0.7400

Table 1: Reference Image is the host Image. 1.Jpeg(65%) means Jpeg Compression with quality 65%. 2.Blur(3) means the Blur operation with parameter 3. 3.Sharpen(75) means the Sharpen operation with parameter 75. 4.Scanning: Grabbing an image from the screen. The image is obtained by first printed-out and then scanned-in. The attack involves several combined operations. 5. StirMark: geometrical deformation by StirMark. 5.Rotation(0.5) means rotating with angle 0.5. 6.Spread(5) means the Spread operation with parameter 5. 7.Translation means translation by one pixel. 8.EZW(0.25bpp) means compression of the Lena image using the EZW method at 32:1.

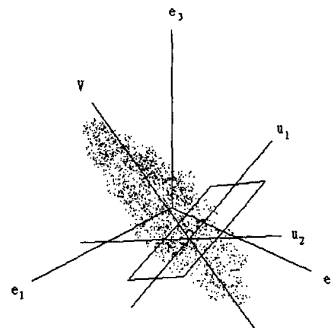


Figure 1: A simplified schematic diagram of our watermarking strategy.