A NEW METHODOLOGY FOR DETECTION OF WATERMARKS IN GEOMETRICALLY DISTORTED IMAGES

Ibrahim B. Ozer, Mahalingam Ramkumar, Ali N. Akansu

Department of Electrical and Computer Engineering
New Jersey Institute of Technology
New Jersey Center for Multimedia Research
University Heights, Newark, NJ, 07102.

ABSTRACT

The goal of this study is to investigate the performance of a previously proposed watermarking scheme [1] for watermarked images distorted by different geometrical attacks such as StirMark [2] and swirl. Although the geometrical distortions are not visually noticeable, the watermark cannot be detected due to the high mean square error between the original and attacked images. In order to extract the watermark from distorted images, the effects of these attacks must be minimized. This work focuses on the recovery of the attacked image using reliable distortion estimation methods enabling watermark detection after different geometrical attacks. The performance test is done for several images attacked by StirMark.

1. INTRODUCTION

Digital watermarking is a means of protecting multimedia data from intellectual piracy. It is achieved by imperceptibly modifying the original data to insert a "signature". The signature is extracted when necessary to show proof of ownership. Let I be the original (cover) image. A watermark embedding function $\mathcal{E}$ inserts a watermark S in the image I to generate the watermarked image $\hat{I} = \mathcal{E}(I, S)$. The existence of the watermark S in an image $\tilde{I}$ is checked by a detector function $\mathcal{D}$. Watermark detectors can be broadly classified into two categories. *Cover image escrow* detectors need the original image I to check for the presence of the signature S in $\tilde{I}$. On the other hand, *oblivious detection* methods *do not* require the original image. We shall term the output of the detector function,

$$s_d = \begin{cases} \mathcal{D}(\tilde{I}, S, I) & \text{cover image escrow} \\ \mathcal{D}(\tilde{I}, S) & \text{oblivious detection} \end{cases} \quad (1)$$

as the *detection statistic.* The detection statistic is an indication of the *degree of certainty* with which the signature S is detected in the image $\tilde{I}$.

Typically, the signature S takes the form of a Gaussian or binary pseudo random sequence $\mathbf{s}$ (say of length N) generated from a *key* $\mathcal{K}$. The watermark embedding and detection operations can therefore be written as

$$\hat{I} = \mathcal{E}(I, \mathbf{s}) \quad \tilde{\mathbf{s}} = \mathcal{D}(\tilde{I}, < I >) \quad s_d = \frac{\mathbf{s}^T \tilde{\mathbf{s}}}{|\mathbf{s}||\tilde{\mathbf{s}}|} \quad (2)$$

In other words, the detection statistic is a measure of (normalized) *inner product* of the embedded and the detected signature sequence. The inner product of randomly generated signature sequences will also be random. More specifically, for large N, the distribution of the inner product will be Gaussian $\mathcal{N}[0, \frac{1}{N}]$. If the creator (or pirate) has *absolutely no freedom* in choosing the signature, and if the detection statistic s_d obtained is say 6 times the standard deviation (if $s_d = 6\frac{1}{\sqrt{N}}$), then we could say that the signature is detected with a probability of error of less than $Q(6) \approx 1 \times 10^{-9}$. This is due to the fact that on an average only 1 out of 1×10^9 signatures chosen randomly can yield such a high correlation.

2. PROTOCOL FOR WATERMARKING

One way to survive geometric attacks like StirMark would be to cause the watermarking method to *introduce* geometric distortions [3]. Let $\mathcal{G}(I)$ be a function of some geometric features of the image I. The watermark is can be introduced by *specifying* $\mathcal{G}(\hat{I})$. However, we cannot expect such methods to be robust to compression. Just as small geometric distortions can modify the MSE significantly, small changes in MSE (such as those that might be introduced due to lossy compression) can change $\mathcal{G}(I)$ significantly. In this light it is not surprising that the watermarking method proposed by Rongen et. al [3] is robust to StirMark, but not robust to compression. However, with some aids to undo such geometric distortion, "conventional" watermarking methods can be used effectively.

Watermarking also needs a *restrictive protocol* to be able to resolve ownership unambiguously. Such a protocol, was proposed in Ref. [1], and is briefly described below. The main aim of the *restrictions* placed by the protocol is to make it extremely difficult for a pirate to engineer a successful attack. Most attacks on watermarks, are performed by introducing a large distortion in the MSE sense without affecting the visual fidelity. *Counterfeit* attacks, which are directed at creating ambiguities in resolving ownership use the *inadequacies of watermarking protocols*, like "freedom" available in choice of the signature, and watermarking algorithms. It was shown in Ref. [1] that such attacks too, additionally, rely on methods to create a "fake" original which is far away in the MSE sense to the actual original, but yet is visually close.

Therefore, a good protocol needs to recognize such at-

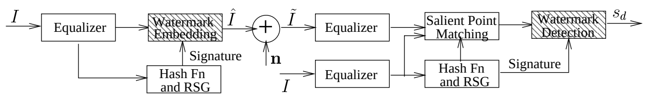

Figure 1: Watermark Embedding and Detection Protocol

tacks, and possibly undo them in order to extract the watermark with a reasonably high degree of certainty. Ref. [1] identified pixel scaling, histogram modification, and imperceptible geometric distortion (such as those introduced by StirMark [2]) as the principal ways in which such fake originals can be created. The watermark detection algorithm therefore needs *regulated* algorithms for undoing or minimizing the effects of such attacks. The block diagram of the proposed protocol is shown in Figure 1. The overall protocol consists of

1. A prescribed algorithm for equalizing histogram. The signature is added to the original image after equalizing its histogram. The histogram of the image in question is equalized (using the same equalizer) before performing detection of the signature.
2. *A prescribed algorithm for determining* salient *points and re-warping the image if necessary.* In this paper, we propose such an algorithm.
3. A prescribed algorithm for determining scale factors of pixel values and re-scaling.
4. A fixed hash function $\mathcal{H}$ operates on the histogram equalized original image I to produce the seed $\mathcal{H}_I$.
5. The seed $\mathcal{H}_I$ is input to a *fixed random sequence generator* $\mathcal{G}$ to generate the signature sequence $\mathbf{s}$. $\mathbf{s}_N^d = \mathcal{G}(\mathcal{H}_I, N, d)$ is the complete set of sequences that could be generated by $\mathcal{G}$. For a fixed I, the only parameters that can be changed are N - the length of the sequence, and d - the probability distribution. Probably d could take two options - Gaussian and Uniform. Another useful option for d might be to generate a list of integers from $1 \cdots N$ in a random order. This may be used for reordering the image coefficients if the algorithm calls for it. No restriction is placed on the length N.
6. Any decomposition of the original image can be used. If decompositions are generated from random sequences only one from the set of possible sequences $\mathbf{s}_N^d$ can be used. If the watermarking algorithm calls for a random sequence (say for re-ordering of coefficients), at any stage of the watermark embedding / extraction process, only random sequences $\mathbf{s}_N^d$ are permitted.
7. Signature to be extracted from the image without subtracting the original image.
8. High detection statistic of the signature with the image in which the existence of the signature is checked, *and* low detection statistic between the signature and the original image. Equivalently, the detection statistic may be considered as the difference between the detection statistics obtained from the image in question and the original.

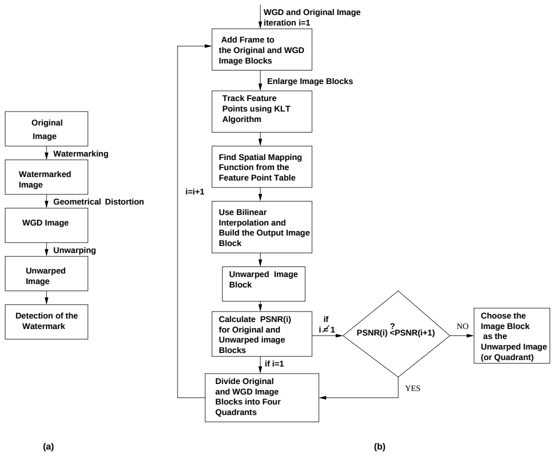

Figure 2: (a) Overall Algorithm (b) Unwarping

3. UNWARPING ALGORITHM

A simplified algorithm of the watermarking process is shown in Figure 2 (a). The watermarked image is attacked by introducing some form of geometrical distortion. The watermarked and geometrically distorted (WGD) image is recovered using the proposed unwarping algorithm which is illustrated in Figure 2 (b). The watermark is detected after undoing the effects of the geometric distortions. In the recovery process, we use polynomial transformation techniques, which determine global alignment of the images [4]. Global methods use sufficient number of matched points for deriving the parameters of any transformation either through approximation or interpolation.

The input to the unwarping process is the WGD image (Figure 2 (b)). WGD and original images are enlarged by taking the mirror image at the boundaries. Since the deformation between the original and WGD images is small, detection of feature points by estimating the affine deformation parameters is not reliable. Because of its higher reliability and accuracy, a pure translation model is preferred for tracking the feature points between the original and WGD images. Hence, the recovery process consists of two main steps: first, the corresponding feature points between the original and WGD images are detected [5]. Then, a spatial mapping function is found using these feature points to unwarp the WGD image [6].

3.1. Detection of Feature Points

A point (x, y) in the first image I moves to point (W_x, W_y) in the second image J, where:

$$J(W_x(x,y), W_y(x,y)) = I(x,y) \tag{3}$$

$$W_x(x,y) = \sum_p^n \sum_q^n a_{pq} x^p y^q \tag{4}$$

$$W_y(x,y) = \sum_p^n \sum_q^n b_{pq} x^p y^q \tag{5}$$

Given the original image I and the WDG image J, the problem is determination of the parameters in the deformation matrix W and d, where $\mathbf{d} = [a_{00} \quad b_{00}]^T$. Since the change between the original and WGD images is small, it is safer to set the higher order parameter matrix to the zero matrix during feature point detection process. The problem becomes the determination of the parameters that minimize the dissimilarity ϵ.

$$\epsilon = \int \int_W [J(W_x(x,y), W_y(x,y)) - I(x,y)]^2 \mathrm{dxdy} \qquad (6)$$

where W is the given feature window. After expanding to Taylor series, d is determined by solving the equation $Z\mathbf{d} = \mathbf{e}$ where:

$$Z = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix} \qquad (7)$$

The eigenvalues of Z determine the selection of feature points, where d gives information about the displacement of the feature points in the second image. The feature points with large eigenvalues correspond to high texture areas that can be matched reliably. These feature points are then used for determination of the spatial mapping function given in the next section.

3.2. Spatial Transformation

The spatial transformation is a mapping of a point (x, y) in image (J) to its warped position (i, j) in the image (I):

$$(i, j) = (W_x(x,y), W_y(x,y)) \qquad (8)$$

Given any point (x, y) in the output image, the coordinates of the corresponding point in the input image can be generated using the warping functions W_x and W_y.

The problem becomes the determination of the coefficient values a and b where n is taken 2 in our approach. In this system, the number of equations is greater than the number of unknown values and there are going to be errors for some points. Consider the matrix equations:

$$\mathbf{i} = \mathbf{Xa} + \mathbf{e} \quad \mathbf{j} = \mathbf{Xb} + \mathbf{e} \qquad (9)$$

where

$$\mathbf{i} = [i_1 i_2 ... i_m]^T \qquad \mathbf{e} = [e_1 e_2 ... e_m]^T$$
$$\mathbf{a} = [a_{00} a_{10} ... a_{22}]^T \quad \mathbf{b} = [b_{00} b_{10} ... b_{22}]^T$$

m is the number of feature points. a and b are the warping coefficients and e is the error vector. To solve these matrix equations, we use the least square error solution by computing $e^T e$:

$$\mathbf{e^T e} = (\mathbf{i} - \mathbf{Xa})^\mathbf{T}(\mathbf{i} - \mathbf{Xa}) \quad \mathbf{e^T e} = (\mathbf{j} - \mathbf{Xb})^\mathbf{T}(\mathbf{j} - \mathbf{Xb}) \; (10)$$

Differentiating $e^T e$ with respect to a and setting it equal to zero, and in a like manner repeating the process for b will result:

$$\mathbf{a} = (\mathbf{X^T})^{-1}\mathbf{X^T i} \quad \mathbf{b} = (\mathbf{X^T})^{-1}\mathbf{X^T j} \qquad (11)$$

	PSNR_1	PSNR_2	s_{d_1}	s_{d_2}
Girl	23.55	30.78	2.56	15.01
Baboon	18.5737	24.42	3.84	45.87
Pepper	20.2871	29.1292	1.66	24.46
Barbara	19.3797	25.6801	1.28	23.33
Lena	19.2658	29.3814	3.07	18.07

Table 1: PSNR and detection statistics for StirMarked images. PSNR_1 and s_{d_1} are respectively the PSNR and watermark detection statistics (expressed as the number of standard deviations) of the StirMarked images. PSNR_2 and s_{d_2} are the corresponding quantities for the unwarped images.

Since the coordinates computed from the unwarping function will not in general be integer values, the grey level must be interpolated from the grey levels of the surrounding pixels. In this work, we use bi-linear interpolation where the estimate is made by considering four neighboring input pixels. The recovery process is implemented adaptively for different blocks of the image to overcome different distortions such as global and local attacks and to increase the accuracy. For this purpose a quadtree representation is used. At each step the image blocks are further divided to quadrants if the resulting PSNR at the new step is less than that of previous step. The experiments and results are given in the next section.

4. RESULTS AND CONCLUSIONS

In our experiments, we geometrically distort watermarked images by using the StirMark algorithm. The difference between the original and WGD images as well as the difference between the original and unwarped images are shown in Figure 3. In Table 1 PSNR and detection statistic values for the distorted and recovered images are given. For each image block the best 100 feature points are tracked and used for the spatial transformation process. Dividing the whole image into quadrants improves the performance of the unwarping process (the improvement being measured in terms of PSNR). The highest PSNR values are obtained when the image is divided into four quadrants. The watermarking method used [1] is briefly outlined in the Appendix.

Application of the proposed unwarping method, in conjunction with the watermarking method outlined in the appendix, is shown to improve the performance of the watermarking scheme significantly, as seen in Table 1. The improvement in detection statistic for the Barbara image for example (from 1.28 to 23.33) implies a reduction in false alarm probability from $Q(2.56) = 0.0052$ to $Q(15.01) \approx 3 \times 10^{-51}$. Note that even lower false alarm probabilities are obtained for other images.

APPENDIX

The cover image I was decomposed into 4 subbands using the 10-tap Daubechies filter. Only the lowest frequency subband was used for embedding the watermark. The selected coefficients further undergo a key based transform to obtain the coefficients $\mathbf{c}$. The binary $(\pm\frac{\Delta}{4})$ signature sequence $\mathbf{s}$

Figure 3: First column : original images. Second Column : Difference between original and StirMarked images. Third Column : Difference between original and unwarped images

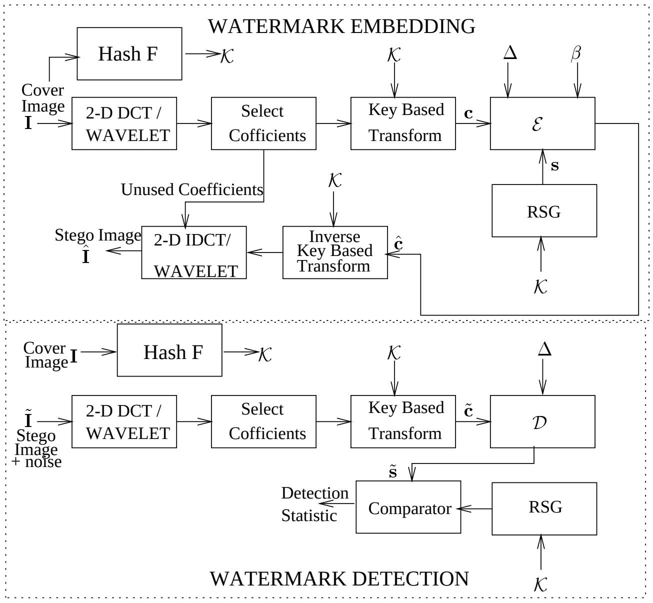

Figure 4: Block Diagram of the Watermark Embedding and Detection

is generated from the key $\mathcal{K}$. The algorithm for the watermark embedder $\mathcal{E}$ (characterized by two parameter Δ and β) which embeds $\mathbf{s}$ in $\mathbf{c}$ to obtain the watermarked coefficients $\hat{\mathbf{c}}$ is

$$
\begin{aligned}
\mathbf{p} &= \mathcal{D}(\mathbf{c}) && (12)\\
e(k) &= s(k) - p(k)\\
e(k) &= (\text{rem}\left(\frac{c(k)}{\Delta}\right) > \frac{\Delta}{2}) \quad ? \quad -e(k) \quad : \quad e(k)\\
\hat{c}(k) &= (c(k) \geq 0) \quad ? \quad c(k) + e(k) \quad : \quad c(k) - e(k)
\end{aligned}
$$

The algorithm for the detector $\mathcal{D}$ is given by

$$
\begin{aligned}
q(k) &= \text{rem}(\frac{|\tilde{c}(k)|}{\Delta}), \quad k = 1 \cdots D\\
\tilde{s}(k) &= (q(k) \geq \frac{\Delta}{2}) \quad ? \quad (\frac{3\Delta}{4} - q(k)) \, : \, (q(k) - \frac{\Delta}{4})
\end{aligned}
$$

In the above equations, $x = (\text{Condition}) \quad ? \quad x_1 \quad : \quad x_2$, stands for "If Condition is true $x = x_1$, else, $x = x_2$".

The choice of the parameters Δ and β depends on the expected distortion of the coefficients $\hat{\mathbf{c}}$, and the *permitted* distortion of the coefficients $\mathbf{c}$. Optimal choice of the parameters is discussed in Ref. [7] In our simulations, Δ was chosen as 100, and β was chosen between 7 (for smooth images like Girl) to as high as 20 for highly textured images like Baboon. The modified coefficients $\hat{bmc}$ (after the inverse key based transform) together with the unmodified coefficients of the other 3 subbands are used to synthesize the watermarked image $\hat{I}$. Detection of the watermark in an image $\tilde{I}$ is performed by obtaining its corresponding coefficients $\tilde{\mathbf{c}}$ and then obtaining $\tilde{\mathbf{s}}$ as the output of the detector. The original signature $\mathbf{s}$ and the recovered signature $\tilde{\mathbf{s}}$ are correlated to obtain the detection statistic.

5. REFERENCES

[1] M.Ramkumar, A.N. Akansu, "A Robust Protocol for Proving Ownership of Still Images", submitted to the *IEEE Trans. on Multimedia*, November 1999.

[2] M. Kutter and F. A. P. Petitcolas. "A Fair Benchmark for Image Watermarking Systems", *Electronic Imaging: Security and Watermarking of Multimedia Contents*, vol. **3657**, pp. 226-239, San Jose, CA, USA, January 1999.

[3] P.M.J.Rongen,M.J.J.J.B. Maes, C.W.A.M. van Overveld, "Digital Image Watermarking by Salient Point Modification: Practical Results," Proceedings of SPIE, Security and Watermarking of Multimedia Contents, San Jose, CA, vol **3657**, pp 273-282, January 1999.

[4] L.G. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys*, Vol. 24, No. 4, pp. 325-376, December 1992.

[5] J. Shi and C. Tomasi, "Good Features to Track," *CVPR*, 1994.

[6] David Vernon, "Machine Vision," Prentice Hall, 1991.

[7] M.Ramkumar, A.N. Akansu, "Self-Noise Suppression Schemes for Blind Image Steganography", SPIE's International Symposium on Voice, Video and Data Communications, Multimedia Systems and Applications (Image Security), Vol **3845**, Boston, MA, Sep. 1999.