MULTIBIT DIGITAL WATERMARKING ROBUST AGAINST LOCAL NONLINEAR GEOMETRICAL DISTORTIONS

Sviatoslav Voloshynovskiy , Frédéric Deguillaume, and Thierry Pun

CUI - University of Geneva,
24 rue du Général Dufour, 1211 Geneva 4, Switzerland

ABSTRACT

This paper presents an efficient method for the estimation and recovering from nonlinear or local geometrical distortions, such as the random bending attack and restricted projective transforms. The distortions are modeled as a set of local affine transforms, the watermark being repeatedly allocated into small blocks in order to ensure its locality. The estimation of the affine transform parameters is formulated as a robust penalized Maximum Likelihood (ML) problem, which is suitable for the local level as well as for global distortions. Results with the Stirmark benchmark confirm the high robustness of the proposed method and show its state-of-the-art performance.

1. INTRODUCTION

One problem with almost all current watermarking technologies is that they fail to recover a watermark from random bending geometrical distortions, known as the random bending attack (RBA). The RBA was first introduced by F. Petitcolas in the benchmarking tool Stirmark to model printing/scanning artifacts [1]. If today watermarking technologies resist in practice against printing/scanning, unfortunately however the RBA attack still remains an essential problem for almost all existing watermarking methods. The practical danger of this attack consists in the fact that the attacker can apply it against some watermarking technology using the Stirmark benchmarking tool, while preserving visual image quality. Having removed the watermark the attacker can commercially exploit the attacked image, violating copyright laws. The main difficulty to deal with the RBA comes from the basic assumption that all geometrical alterations introduced by the attacker are modeled as a global affine transform. This does not hold for the RBA where the introduced distortions cannot be described using the parameters of a global affine transform only. Moreover, the situation is complicated by the fact that many technologies [2-4] are using a global template in the magnitude spectrum of the image, which does not allow to differentiate local alterations introduced in the case of RBA.

A group of methods are using the assumption about the local character of the RBA [5,6]. However, an exhaustive search is used to recover from this attack. Moreover, no dedicated synchronization structure for the estimation of local distortions is proposed in the above methods, except exhaustive search solutions. This restricts the usage of such methods in commercial and on-line applications due to the high computational complexity of the exhaustive approach.

One way to overcome this problem is to divide the image into segments or cells, and to embed the watermark into each segment. This has been done by Rhoads [7], by Lin et al [8] as well as by Voloshynovskiy et al [9]. A particular case of this approach to watermark generation is the periodical tiling of the same watermark. In fact the idea of repeating the same watermark has several advantages. First, it allows to resist against cropping. Secondly, exploiting the periodical structure of the watermark one can use either the autocorrelation function (ACF) [10] or the magnitude spectrum of the Fourier transform [9] to estimate and recover from global transformations. Unfortunately, all these schemes have the important defect that local random bending alterations and the general class of projective transformations were not integrated in the watermark detector.

The goal of this paper is to show a possible solution of the RBA problem based on the periodical watermarking. The watermarking method we describe here has three novel features: first, an estimation method which is able to estimate and recover from local or non-linear geometrical alterations in images or videos; secondly, a reference watermark which can be used to recover from local geometrical transforms, and which is also encoded and can be used for the verification of the reliability of the local geometrical transform recovering, or for fast detection of the watermark from the given data for a particular key; thirdly, a locally flipped informative watermark that can be used for the estimation and the recovering from RBA, local nonlinear and projective transforms based on local ACF or magnitude spectrum of a given small local region. The locally flipped informative watermark can be additionally used for the estimation and compensation of translation and cropping based on zero-phase condition.

We will use the term of *informative watermark* to refer to the watermark which carries the message, and of *reference watermark* to refer to the watermark which carries information about synchronization, reliability and detection (i.e. watermark presence/absence).

The main idea for these features is to consider the geometrical transforms at a local hierarchical level instead of modeling them as global affine transforms. That allows to approximate a global projective transform as a juxtaposition of local affine transforms; this observation is true for RBA too. In the case of global affine transforms, the parameters of local affine transforms will be the same as the global one, and this allows to utilize the same unified approach for modeling all the above kinds of attacks.

Section 2 introduces the problem formulation. Section 3 describes the embedding of a watermark which is resistant to the above attacks, based on this formulation. In section 4 the method of determination of the global affine transform as well as the local affine transforms approximation are detailed. Results are then given in section 5, showing the high robustness of our approach.

2. PROBLEM FORMULATION

The algorithm we propose consists in the novel use of the two components mentioned above: a specially designed method for the recovering from local geometrical transforms, and a reference watermark. The problem of watermark decoding can be solved by the approach of channel state estimation in order to recover from attacks, as presented in Voloshynovskiy et al [11] for the cases of fading and stationary Generalized Gaussian (sGG) noise; the main emphasis of this paper is to extend this concept to geometrical attacks, and in particular RBA.

First, the method for recovering from local geometrical transforms is a tool which is based on the assumption that global affine or projective transforms as well as RBA can be considered as a set of local affine transforms. This approximation is possible due to the restricted amount of invisible distortions that can be introduced by the random bending to keep the quality of commercial image within acceptable ranges. Keeping in mind the high level of local correlation required in images, the amount of distortions cannot be voluntarily high. This assumption allows to design a special type of watermark and a dedicated procedure for the estimation and compensation of these geometrical distortions. In particular, we propose here two methods of constructing a local watermark based on the flipped watermark itself, or on the repetition of the watermark.

Secondly, the reference watermark is a key-based sequence that is encoded using some error correction codes (ECC) and inserted into the image closely to the positions of the informative watermark. This additional reference watermark helps us:

- for the determination of the watermark presence or absence in the given image for the given key;
- as a pilot for the estimation of a channel state for the optimal design of a matched filter in the decoder for the informative watermark [11];
- for the evaluation of the reliability of the local and global geometrical transforms recovering;
- for the estimation of the reliability of the decoding of the informative watermark.

The informative watermark itself, organized in a special spatial structure, can be used for the last purpose as well. Note that the watermark is not restricted to be of square shape, but can also be of any regular or irregular shape that is then replicated in a special manner (not necessary strictly periodical) over the image.

3. WATERMARKING ALGORITHM

Here we choose the method proposed by Voloshynovskiy et al. and detailed in [9] and illustrated by Figure 1, since it offers a good compromise between the block size needed for the watermark embedding and the assumption about the locality of the affine approximation of the RBA. Obviously, any block-based technology can be adopted for this purpose. We review the steps involved. We first encode the input message using any ECC which has performances similar as those described in the above paper. The resulting codeword is then mapped from {0,1} to {-1,1} using binary phase shift keying (BPSK) and encrypted based on a key-dependent sequence, followed by a spreading over a square block or segment of any shape with some density D based on the same secret key. The reference watermark, also key-dependent, is first encoded using the same ECC and is then added to the above block in the remaining (i.e. orthogonal) spatial locations. The reference watermark also consists of a binary sequence {-1,1} and its length is determined by the embedding density (1-D) as described above. The resulting block is upsampled by a factor 2 to receive a low-pass watermark and then flipped and copied once in each direction, producing a symmetric macroblock. The flipping is performed first to visually decorrelate the structure of the repeated watermark, and secondly to reduce the number of ambiguities during the estimation of the undergone geometrical attacks. Obviously, non-regular upsampling can be used to produce some groups of 2, 3, 4, 5 or more pixels with other grouping configurations to resists against printing/scanning attack. Any "center of symmetry" for flipping, in the general case key-dependent, can be chosen to create the flipped macroblock. Finally, the resulting macroblock is replicated over the whole image size, resulting in a symmetrical and periodical watermark.

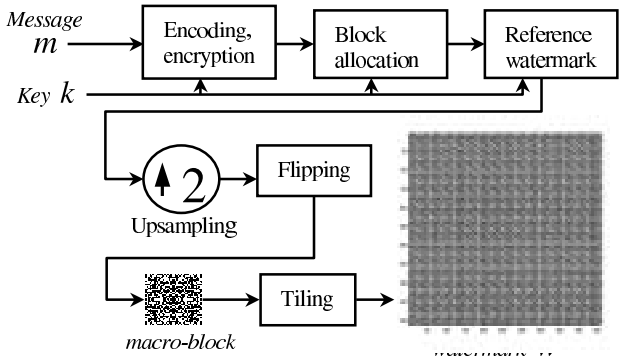

Figure 1. Watermark embedding process: the message m is encoded using some ECC, encrypted, mixed with the reference watermark, and allocated into a block, depending on the secret key k. This block is then upsampled, flipped, and the resulting macro-block is tiled up to the complete image size.

In the case of a square shaped block, the watermark can be expressed as:

$$w_p(x,y) = \sum_{m=0}^{K_x-1}\sum_{n=0}^{K_y-1} w(x-mT, y-nT) \qquad (1)$$

where $K_x = \lceil M/T \rceil$, $K_y = \lceil N/T \rceil$, M,N is the image size, and T is the period of replication along each axis.

The resulting watermark can be slightly pre-distorted in such a way that in every period, the watermark can have some small affine distortion to resist against spatial averaging and removal attack, which is an attack based on the subtraction of the estimated sign of the watermark in the macroblock. This pre-

distortion will not significantly affect the ACF or magnitude spectrum of the watermark used for the recovering of the global affine transform, and will not interfere with the recovering of the local transforms.

Both the cover image and the watermark are first decomposed into a multi-resolution sub-band pyramid using a Wavelet transform and are then added together using a perceptual masking applied to the watermark. Different maskings are applied to the flat regions and to the textured areas of the image during the watermark embedding. The resulting stego image is then back-transformed to the spatial domain, or is directly stored in the compressed domain.

4. RECOVERING FROM GEOMETRICAL ATTACKS

In order to extract the information, we first use either a Maximum Likelihood (ML) estimator, or a penalized ML or a minimum mean square error (MMSE) estimator. In the case when no geometrical transform was applied the message is decoded from the extracted watermark directly. If some geometrical transform was applied, the extracted watermark is processed in order to invert it.

4.1. Recovering from global affine transforms

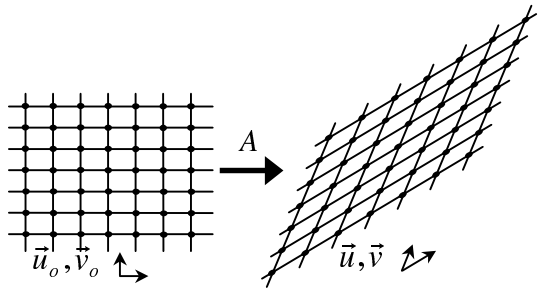

Figure 2. ACF or magnitude spectrum underlying grid, and vectors $\vec{u}_o, \vec{v}_o$ representing it, corresponding to the embedded information, used as a reference (left). Distorted grid and associated vectors $\vec{u}, \vec{v}$ after an affine transform A (right).

An important problem constraining the practical exploitation of watermarking technology consists in the low robustness of the existing watermarking algorithms against general geometrical attacks such as rotation, scaling, cropping, translation, change of aspect ratio and shearing. All these attacks can be uniquely described by a global affine transform. An affine transform can be represented by the 4 coefficients a,b,c,d which form the linear component matrix A, plus a translation component $\vec{v}$:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \vec{v} \qquad (2)$$

The translation component ($\vec{v}$) can be separately recovered, for example based on an intercorrelation between the extracted watermark and the reference watermark mentioned above, and can be ignored in the following developments. Therefore, an affine transform maps each point of cartesian coordinates (x, y) to (x', y'), according to the formula:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = A \cdot \begin{pmatrix} x \\ y \end{pmatrix} \qquad (3)$$

where '.' is the matrix product. Successive combination of n affine transforms A_i, $i = 1...n$ yield another affine transform which can be expressed as $A = A_n \cdot A_{n-1} \cdot ... \cdot A_1$. With respect to the originally embedded watermark w, the resulting watermark w'_p after a global affine transform can be written as:

$$w'_p(x,y) = \sum_{m=0}^{K_x-1} \sum_{n=0}^{K_y-1} w\left(A^{-1}(x,y)^T - (mT, nT)^T\right) \qquad (4)$$

where A is applied to all image blocks. Due to the periodicity of the used watermark, the ACF of the watermark, as well as its magnitude spectrum, result in a structure showing local maxima, or peaks, which is periodical too [9]. The general setup of this work is illustrated in Figure 2 where the underlying grid structure of input points for the proposed algorithm represents the peaks in the ACF or in the magnitude spectrum of a periodical watermark, before and after the application of an affine transform. Peaks are placed at the intersection between lines.

We propose to use an approach based on penalized ML estimation, as a robust approach for the estimation of the applied affine transform:

$$\hat{A} = \arg \min_{A \in \Phi} \left\{ \rho\left(\begin{pmatrix} x' \\ y' \end{pmatrix} - A \begin{pmatrix} x \\ y \end{pmatrix} \right) + \mu \Omega(A) \right\} \qquad (5)$$

where $\hat{A}$ is an estimate of the affine transform within the set of possible solutions $A \in \Phi$, μ is a regularization parameter that controls the trade-off between the feasibility of the estimate with respect to the observed data (first term), and the prior $\Omega(A)$. $\rho(.)$ denotes the cost function, which is a quadratic norm in the case of Gaussian misalignments, or a ℓ_1 norm for Laplacian misalignments; more general cases can be expressed for a generalized Gaussian distribution of misalignments, of which the two previous cases are particular cases. The prior $\Omega(A)$ represents the possible variations of the four parameters a,b,c,d corresponding to the linear transform matrix A. The idea of restricting possible combinations of shearing, flipping and radii between template points in the magnitude spectrum proposed by S. Pereira [2] can be considered as a particular case of this prior.

4.2. Recovering from local non-linear transforms

A different situation is observed for local non-linear transforms. As said before, the RBA and the projective transforms are approximated as local affine transforms, giving the following correspondence between the embedded watermark w and the distorted watermark w'_p:

$$w'_p(x,y) \approx \sum_{m=0}^{K_x-1} \sum_{n=0}^{K_y-1} w\left(A_{mn}^{-1}(x,y)^T - (mT, nT)^T\right) \qquad (6)$$

where A_{mn} is an approximation of the local linear transform applied to the m,n^{th} block.

In order to determine local affine transforms, one can either use local ACF or magnitude spectrums, or exploit the reference watermark information at the block level as shown in Figure 3. We propose to use the same penalized ML estimation of equation 5 as for the global level, assuming that a local affine approximation introduces random misalignments with a Laplacian distribution.

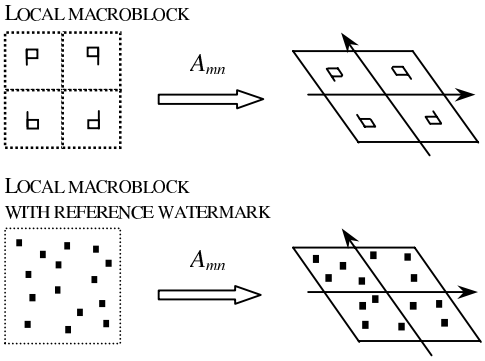

Figure 3. Distortion locally applied on a flipped macro-block, approximated as an affine transform A_{mn} (up). The same, considering the reference watermark (the black squares), which can be used to locally resynchronize the watermark information (down).

5. RESULTS

We tested our approach based on the Stirmark 3.1 benchmark of F. Petitcolas [1] using 6 standard images. The results are shown in Table 1, by marks between 0 (no watermark decoded) to 1 (watermark resisted all attacks). The watermark was decoded from all randomly distorted images, significantly increasing the total score up to 0.996 over 1. Today, no known algorithm presents such a high score.

Applied attack	Stirmark score
Signal enhancement	1,00
Compression (JPEG/GIF)	0,99
Scaling	1,00
Cropping	0,99
Shearing	1,00
Rotation (auto-crop, auto-scale)	0,99
Column and line removal	1,00
Flip	1,00
Random geometrical distortions	1,00
Total average score	**0,996**

Table 1. Stirmark 3.1 benchmarking of our approach.

6. CONCLUSION

We described a new method of digital watermarking which can resist against random bending attack as well as projective transforms, and does not require the recovering of global affine transform or even the repetition of the same watermark pattern. This invention is not limited to resistance against random bending attack, in fact it will become apparent that the watermark can be designed so as to resist against common global affine transformations such as rotation, scaling, and changes of aspect ratio, cropping as well as other types of operations such as filtering, lossy compression, printing/scanning, or detection of watermark in front of a video, web or photo camera, or any other imaging device.

7. REFERENCES

[1] M. Kutter and F.A.P. Petitcolas, "A fair benchmark for image watermarking systems", Proceedings of SPIE: *Security and Watermarking of Multimedia Content*, vol. 3657, pp. 219-239, San Jose, CA, USA, January 1999.

[2] S. Pereira and T. Pun, "Fast Robust Template Matching for Affine Resistant Watermarks", Lecture Notes in Computer Science*: Third International Workshop on Information Hiding*, Springer, vol. 1768, pp. 199-210, 1999.

[3] M. Barni, F. Bartolini, V. Cappellini and A. Piva, "Metodo e sistema di marchiatura o cosiddetto watermarking di immagini digitali" ("A method and a system for digital image watermarking"), Italian Patent FI99A000090, filed April 1999.

[4] M. Barni, F. Bartolini, V. Cappellini, A. De Rosa and A. Piva, "Metodo di rivelazione di un marchio in immagini digitali" ("A method for detecting watermarks in digital images"), Italian Patent FI99A000091, filed April 1999.

[5] P. Bas, J.M. Chassery and B. Maco, "Robust watermarking based on the warping of predefined regular triangular patterns", Proceedings of SPIE: *Security and Watermarking of Multimedia Content II*, San Jose, CA, USA, January 2000.

[6] J.-L. Dugelay and F.A.P. Petitcolas, "Image watermarking: possible counterattacks against random geometric distortions", Proceedings of SPIE: *Security and Watermarking of Multimedia Content II*, vol. 3971, pp. 24-26, San Jose, CA, USA, January 2000.

[7] G.B. Rhoads, "Steganography systems", International Patent WO 96/36163 PCT/US96/06618, November 1996.

[8] C. Lin, M. Wu, J. A. Bloom, I.J. Cox, M.L. Miller, Y.M. Lui, "Rotation, Scale, and Translation Resilient Public Watermarking for Images", Proceedings of SPIE: *Security and Watermarking of Multimedia Contents II*, vol. 3971, pp. 90-98, San Jose, CA, USA, January 2000.

[9] S.Voloshynovskiy, F. Deguillaume and T. Pun, "Content adaptive watermarking based on a stochastic multiresolution image modeling", EUSIPCO2000, X European Signal Processing Conference, Tampere, Finland, September 2000.

[10] M. Kutter, "Watermarking resistant to translation, rotation and scaling", SPIE International Symposium on Voice, Video, and Data Communication, November 1998.

[11] S. Voloshynovskiy, F. Deguillaume, S. Pereira and T. Pun, "Optimal adaptive diversity watermarking with state channel estimation", Proceedings of SPIE*: Security and Watermarking of Multimedia Content III*, vol. 4314, San Jose, CA, USA, 22-25 January 2001.