Quality Studies of an Invisible Chaos-Based Watermarking Scheme with Message Extraction

Jacques M. Bahi, Jean-François Couchot, Nicolas Friot, and Christophe Guyeux*
FEMTO-ST Institute, UMR 6174 CNRS
Computer Science Laboratory DISC
University of Franche-Comté
Belfort, France
{*jacques.bahi,jean-francois.couchot, nicolas.friot, christophe.guyeux*}*@femto-st.fr*

Kamel Mazouzi*
HPC Mesocenter
University of Franche-Comté
Besançon, France
kamel.mazouzi@univ-fcomte.fr
**Authors are cited in alphabetic order*

***Abstract*—This paper takes place in the field of invisible chaos-based watermarking schemes. It addresses the quality study of an already pyblished algorithm by focusing on three class of properties. Its robustness is experimentally shown against classical attacks on a large set of image instances and image transformations. It correctness and completness are formally proven. Due to this main advantages, this process is fitted for practical use.**

***Keywords*-Invisible Watermarking; Chaos; Robustness; Correct and Complete message extraction.**

Recently, chaos has become an usual technique to define schemes used for encryption or watermarking [2], [5]. In this context, embedded watermarks can be either visible or invisible. In the former case, the mark overlays the image host and is thus visible. In the later case, the mark is embedded in such a way that the differences between the original host and the watermarked one are perceptually unnoticeable.

Our invisible chaos-based watermarking scheme proposed in this research work uses such kind of techniques. It is an extension of a previously released information hiding method [3] initialy used for steganography, and adapted here in a watermarking framework in the direction of quality analysis. Its robustness facing geometrical attacks and signal processing is studied.

The remainder of this document is organized as follows. In Section I, the watermarking scheme is given. Furthermore, the correctness and the completeness of the approach has been proven. This is the first contribution of the paper. Then, in Section II, the robustness of $\mathcal{CI}$ is studied facing geometrical and signal processing attacks through a large number of experiments. This is the second contribution. The documents ends with a conclusion section, where our contribution is summarized and intended future researches are presented.

I. The Information Hiding Scheme $\mathcal{CI}$

This section recalls basics of $\mathcal{CI}$ formerly defined in [3].

The set of all $k-$strategies is furthere denoted as to $\mathbb{S}_\mathsf{k}$. For $\mathsf{k} \in \mathbb{N}^*$, a $k-$*strategy* is a sequence which elements belong into $[\![0, \mathsf{k}-1]\!]$. The term "strategy" will be used instead of $k-$strategy when the context will easily allow to recover k. The following notations are also used in the sequel: $[\![0;N]\!] = \{0, 1, \ldots, N\}$, $\mathbb{B} = \{0, 1\}$, S^n denotes the n^{th} term of a sequence S, and V_i is for the i^{th} component of a vector V.

A. Notations used for $\mathcal{CI}$

- $x^0 \in \mathbb{B}^\mathsf{N}$ is the representation of the chosen Least Significant Coefficients (LSCs) of a given host media where the watermark will be embedded. They are expressed as a vector of N Boolean values. In this work and in experimentations, x^0 is defined as the Least Significat Bits (LSBs) of the host content, since we want the scheme to produce an invisible watermark.
- $m^0 \in \mathbb{B}^\mathsf{P}$ is the watermark message to embed into x^0. This is a vector of P Boolean values.
- $S_p \in \mathbb{S}_\mathsf{N}$ is a strategy called **place strategy**. Intuitively, this sequence defines which element of x is modified at each iteration.
- $S_c \in \mathbb{S}_\mathsf{P}$ is a strategy called **choice strategy**. It defines which element of the watermark is embedded at each iteration.
- Lastly, $S_m \in \mathbb{S}_\mathsf{P}$ is a strategy called **mixing strategy**. This sequence gives which element of the watermark is switched at each iteration.

In what follows, x^0 and m^0 are sometimes replaced by x and m for the sake of brevity, when such abridge does not introduce confusion.

B. The $\mathcal{CI}$ scheme

With this material and for all (n, i, j) in $\mathbb{N}^* \times [\![0;\mathsf{N}-1]\!] \times [\![0;\mathsf{P}-1]\!]$, the $\mathcal{CI}$ process is defined by:

$$\left\{ \begin{array}{l} x_i^n = \left\{ \begin{array}{ll} x_i^{n-1} & \text{if } S_p^n \neq i \\ m_{S_c^n}^{n-1} & \text{if } S_p^n = i. \end{array} \right. \\ \\ m_j^n = \left\{ \begin{array}{ll} m_j^{n-1} & \text{if } S_m^n \neq j \\ & \\ \overline{m_j^{n-1}} & \text{if } S_m^n = j. \end{array} \right. \end{array} \right.$$

where $\overline{m_j^{n-1}}$ is the Boolean negation of m_j^{n-1}. The stego-content is the Boolean vector $y = x^l \in \mathbb{B}^{\mathsf{N}}$ provided the following constraints are applied:

1) The number l of iterations is sufficiently large (see details below).
2) Let $\Im(S_p)$ be the set $\{S_p^1, S_p^2, \ldots, S_p^l\}$ of cardinality k, $k \leq l$ (repetitions are removed in a set). This set contains all the elements of x that have been modified along the iteration process. Let us consider $\Im(S_c)_{|D}$ defined by $\{S_c^{d_1}, S_c^{d_2}, \ldots, S_c^{d_k}\}$ where d_i is the last iteration that has modified the element $i \in \Im(S_p)$. We require that this set is equal to $[\![0; \mathsf{P}-1]\!]$.

Let us discuss the constraints given above. The first one implies that the number of iterations is greater than a given threshold. This requirement has both practical and theoretical reasons. Theoretically speaking, the ability to successfully pass statistical tests is directly linked to this number of iterations. And, in practice, this value is bounded by the size of the host content. The second constrain, for its part, addresses the method's completeness and correctness, as detailed below.

C. Correctness and Completeness Studies

Without attack, the scheme has to ensure that the user can always extract a message and that this latter is the watermark, provided the user has the correct keys. These two demands correspond respectively to the study of completeness and of correctness for the proposed approach. To achieve this study, let us firstly prove the following assessment.

Proposition 1: In section I-B, item 2 is a necessary and sufficient condition to allow message to be extracted from the host.

Proof: For sufficiency, let d_i be the last iteration (date) the element $i \in \Im(S_p)$ of x has been modified:

$$d_i = \max\{j | S_p^j = i\}.$$

Let $D = \{d_i | i \in \Im(S_p)\}$. The set $\Im(S_c)_{|D}$ is thus the restriction of the image of S_c to D.

The host that results from this iteration scheme is thus $(x_0^l, \ldots, x_{\mathsf{N}-1}^l)$ where x_i^l is either $x_i^{d_i}$ if i belongs to $\Im(S_p)$ or x_i^0 otherwise. Moreover, for each $i \in \Im(S_p)$, the element $x_i^{d_i}$ is equal to $m_{S_c^{d_i}}^{d_i-1}$. Thanks to constraint 2, all the indexes $j \in [\![0; \mathsf{P}-1]\!]$ belong to $\Im(S_c)_{|D}$. Let then $j \in [\![0; \mathsf{P}-1]\!]$ s.t. $S_c^{d_i} = j$. Thus we have all the elements m_j of the vector m. Let us focus now on some $m_j^{d_i-1}$. Thus the value of m_j^0 can be immediately deduced by counting in S_c how many times the component j has been switched before $d_i - 1$.

Let us focus now on necessity. If $\Im(S_c)_{|D} \subsetneq [\![0; \mathsf{P}-1]\!]$, there exist a $j \in [\![0; \mathsf{P}-1]\!]$ that does not belong to $\Im(S_c)_{|\Im(S_p)}$. Thus m_j is not present in x^l and the message cannot be extracted. ∎

When the constraint 2 is satisfied, we obtain a scheme that always finds the original message provided the watermarked media has not been modified. In that context, correctness and completeness are established.

Thanks to constraint 2, the cardinality k of $\Im(S_p)$ is larger than P. Otherwise the cardinality of D would be smaller than P and similar to the cardinality of $\Im(S_c)_{|D}$, which is contradictory.

One bit of index j of the original message m^0 is thus embedded at least twice in x^l. By counting the number of times this bit has been switched in S_m, the value of m_j can be deduced in many places. Without attack, all these values are equal and the message is immediately obtained. After an attack, the value of m_j is obtained as mean value of all its occurrences. The scheme is thus complete. Notice that if the cover is not attacked, the returned message is always equal to the original due to the definition of the mean function.

D. Deciding whether a Media is Watermarked

Let us consider a media y that is watermarked with a message m. Let us consider y' that is an altered version of y, *i.e.*, where some bits have been modified. Let m' be the message that is extracted from y'.

Let us now check how far the extracted message m' is from m. To achieve this, let us consider $M = \{i | m_i = 1\}$ of the Boolean vector message m and similarly the set M' for the message m'. Most of similarity measures depend on the functions a, b, c, and d, all from $\mathbb{B}^{\mathsf{P}} \times \mathbb{B}^{\mathsf{P}}$ to $\mathbb{N}$, and respectively equal to $a(m, m') = |M \cap M'|$, $b(m, m') = |M \setminus M'|$, $c(m, m') = |M' \setminus M|$, and $d(m, m') = |\overline{M} \cap \overline{M'}|$ ($|\Gamma|$ and $\overline{\Gamma}$ respectively denote the cardinality and the complementary of any set Γ). In what follows a, b, c, and d respectively stand for $a(m, m')$, $b(m, m')$, $c(m, m')$, and $d(m, m')$.

According to [4] the Fermi-Dirac measure $\mathcal{M}_{FD}$ is the one that has higher discrimination power, *i.e.*, which allows a clear separation between correlated vectors and uncorrelated ones. The measure is recalled hereafter with respect to the previously defined scalars a, b, and c.

$$\begin{array}{rcl} \mathcal{M}_{FD}(\varphi) & = & \dfrac{F_{FD}(\varphi) - F_{FD}(\frac{\pi}{2})}{F_{FD}(0) - F_{FD}(\frac{\pi}{2})}, \\ F_{FD}(\varphi) & = & \dfrac{1}{1 + \exp(\dfrac{\varphi - \varphi_0}{\gamma})}, \end{array}$$

where $\varphi = \arctan(\dfrac{b+c}{a})$, φ_0 is $\pi/4$, and γ is 0.1.

The distance between m and m' is then computed as $1 - \mathcal{M}_{FD}(m, m')$ and is thus a real number in $[0; 1]$. The Table I gives an example of the relation between modified watermarked and the computed distance. If such a distance is behind a threshold, y' will be declared as watermarked and not watermarked otherwise. Next section presents a practical evaluation of this approach.

0.0	0.03	0.24	0.45	0.66	0.77	0.89	0.94	0.99

Table I
RELATION BETWEEN WATERMARKS DEGRADATION AND THE DISTANCE.

II. ROBUSTNESS STUDY OF $\mathcal{CI}$

This section is devoted to the robustness study of our scheme. This one has to ensure that the watermark withstands against different types of active attacks that modify the watermarked image.

For the whole experiment, a set of 100 images is randomly extracted from the database taken from the BOSS contest [1]. In this set, each cover is a 512×512 grayscale digital image. The considered watermark m is given in Table I. Testing the robustness of the approach is achieved by successively applying on watermarked images attacks like cropping, compression, geometric transformations,. . .

Remark 1: On the following figures, the difference percentage corresponds to the distance between the retrieved and the original watermarks.

Robustness of the approach is evaluated by applying different percentages of cropping: from 0.25% to 90%. Results are given in Fig. 1, which presents effects of such an attack. All the percentage differences are so far less than 97% and thus robustness is established.

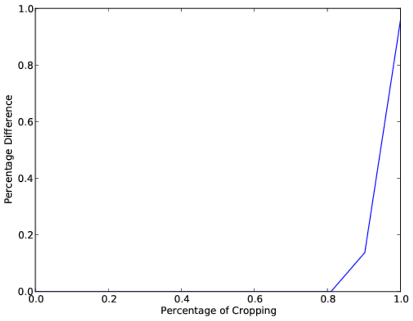
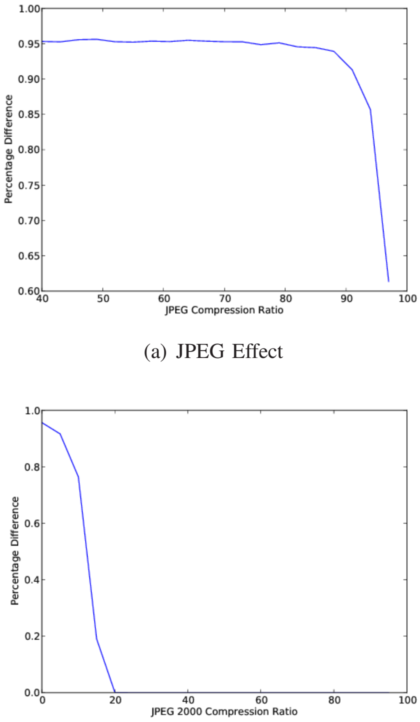


Figure 1. Cropping Results

Robustness against compression is addressed by studying both JPEG and JPEG 2000 image compression. Results are respectively presented in Fig. 2(a) and Fig. 2(b). It is not hard to see that robustness is well established for JPEG2000 compression: for all the ratio larger than 10%, the watermark is retrieved. However, our scheme is not robust against JPEG compression for a ratio inferior to 90%.

A potential solution in order to improve this result should be to insert the watermark in least significant coefficient of the image described in frequency domain as for example with discrete cosine transform or with wavelet transform. This study will be described in future works.



(a) JPEG Effect

(b) JPEG 2000 Effect

Figure 2. Compression Results

Among geometric transformations, we focus on rotations, *i.e.*, when two opposite rotations of angle θ are successively applied around the center of the image. In these geometric transformations, angles range from 2 to 60 degrees. Results are presented in Fig. 3. Thanks to an efficient embedding, our scheme is resistant to all these types of attacks.

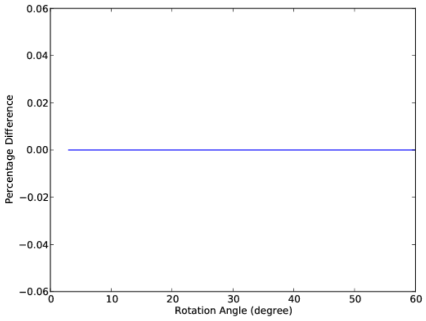
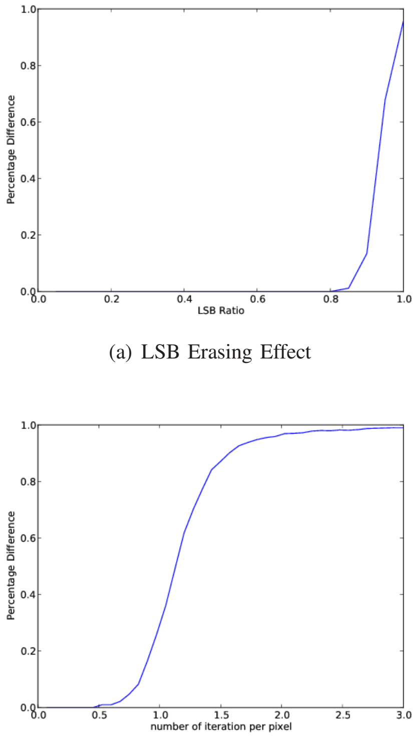


Figure 3. Rotation Attack Results

Let us first recall that this scheme has defined x as the LSBs of the host and is thus based on LSBs modifications. This part focuses on two types of attacks modifying these set of LSBs (see Fig 4). The former consists in setting to zero a subset of this one. Results are expressed in Fig. 4(a) and show that the scheme is robust, unless 95% of the LSBs is erased. In this case the image is really damaged. The latter consists in applying again this scheme on the watermarked image but with another message. Results of Fig. 4(b) show that this scheme is robust against that type of attack, provided the number of iterations is lesser than 1.75 times the number of pixels. With more iterations, the image is dramatically modified: more than 50% of the LSB is switched.



(a) LSB Erasing Effect

(b) Applying Algorithm twice

Figure 4. LSB Modifications

Finally a receiver Operating Characteristic approach has been implemented to find the most adapted threshold w.r.t. the separation between watermarked images and other ones.

The Figure 5 is the Receiver Operating Characteristic (ROC) curve. This curve is close to the ideal one that is without False Positive and False Negative answer. The threshold with best results is a distance equal to 0,97. With such a value, we can give some confidence intervals for most of evaluated attacks. The approach is resistant to all the cropping where percentage is less than 90, to JPEG200 compression where quality ratio is greater than 5%, to all the rotation attacks, to LSB erasing when less than 95% are set to 0, a second application of the scheme with less than 1.75 iterations per pixel.

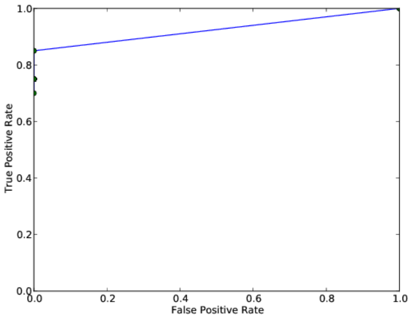


Figure 5. ROC Curves for DWT or DCT Embeddings

III. Conclusion and Future Work

In this research work, a complete quality study of our scheme [3] has been given, namely robustness, completness, and correctness. This scheme is now ready for practical use.

To improve again the robustness, notice that the definition of x can be changed as follows: at worst, any process which always returns the same set of bits for a given image would return an amanable vector. However, the most fine would be the set of bits whose modifications minimize a distortion function. Following such idea we plan to combine this bit selection step with feature extraction function. We are aware that in the field of information hiding it is known that embedding in LSBs is not a good choice. They are used in this work in order to validate the concept which remains valid with an other choice. So in future work these better choice will be explored.

References

[1] Patrick Bas, Tomás Filler, and Tomás Pevný. ”break our steganographic system”: The ins and outs of organizing boss. In Tomás Filler, Tomás Pevný, Scott Craver, and Andrew D. Ker, editors, *Information Hiding*, volume 6958 of *Lecture Notes in Computer Science*, pages 59–70. Springer, 2011.

[2] Zhao Dawei, Chen Guanrong, and Liu Wenbo. A chaos-based robust wavelet-domain watermarking algorithm. *Chaos, Solitons and Fractals*, 22:47–54, 2004.

[3] Nicolas Friot, Christophe Guyeux, and Jacques M. Bahi. Chaotic iterations for steganography - stego-security and chaos-security. In Javier Lopez and Pierangela Samarati, editors, *SECRYPT*, pages 218–227. SciTePress, 2011.

[4] Maria Rifqi, Marcin Detyniecki, and Bernadette Bouchon-Meunier. Discrimination power of measures of resemblance. In *IFSA'03*, 2003.

[5] Zhao Yantao, Ma Yunfei, and Li Zhiquan. A robust chaos-based dct-domain watermarking algorithm. In *Computer Science and Software Engineering, 2008 International Conference on*, volume 3, pages 935 –938, dec. 2008.