An Additive and Lossless Watermarking Method Based on Invariant Image Approximation and Haar Wavelet Transform

W. Pan, Ph.D. Student, G. Coatrieux, N. Cuppens, F. Cuppens, *IEEE Members*, and Ch. Roux, *IEEE Fellow*

***Abstract*—In this article, we propose a new additive lossless watermarking scheme which identifies parts of the image that can be reversibly watermarked and conducts message embedding in the conventional Haar wavelet transform coefficients. Our approach makes use of an approximation of the image signal that is invariant to the watermark addition for classifying the image in order to avoid over/underflows. The method has been tested on different sets of medical images and some usual natural test images as Lena. Experimental result analysis conducted with respect to several aspects including data hiding capacity and image quality preservation, shows that our method is one of the most competitive existing lossless watermarking schemes in terms of high capacity and low distortion.**

I. Introduction

WATERMARKING sensitive documents like medical images induces concomitant risks of interferences with image interpretation as message embedding consists in modifying the image pixel gray level values. To avoid the loss of details that are important for the analysis of the image, reversible or lossless watermarking offers the possibility to reconstruct the original image from its watermarked version by reversing the watermarking distortion. However, it must be noted that if the reversibility property is very helpful as the watermarking invisibility constraint is relaxed, it may also introduce discontinuity in data protection. In fact, the image is no longer protected once the watermark removed similarly to data encryption. So even if watermark removal is possible, imperceptibility has to be guaranteed as most applications have a high interest to keep the watermark in the image as long as possible, continuously protecting the information [1].

Since 1999, several reversible watermarking methods have been proposed [2-11]. According to the message insertion approach, we distinguish two main classes: additive methods and substitutive methods. With additive insertion, the message m to be embedded is first transformed into a watermark signal w next added to the host image I, leading thus to the watermarked image I_w. Without paying careful attention, the resulting watermarked image may include pixels with gray values outside the allowable image dynamic range (e.g. $[0\ldots2^p - 1]$ for a p bit depth image with positive integer values), introducing underflows (negative values) or overflows (values greater than $2^p - 1$).

To overcome these issues of over/undeflow different additive approaches have been proposed. The method suggested by Honsinger *et al.* [2] consists in using arithmetic modulo: $I_w = (I + w) \bmod V_{max}$, where V_{max} corresponds to the maximum value of the signal dynamic range. Thus, the original image can be retrieved simply by subtracting the watermark. The main drawback of this solution is that it may introduce a salt and pepper noise due to jumps between congruent values of the image histogram. To reduce the salt-and-pepper visual artifacts, De Vleeschouwer *et al.* [3] proposed a histogram cyclic modification technique that handles short ranges of the signal dynamic. The method suggested by Ni *et al.* in [5] refers to histogram shifting modulation. It shifts a range of the image histogram to create a 'gap' near the histogram maxima. Pixels with gray values associated to the class of the histogram maxima are then shifted to the gap or kept unchanged to encode '0' or '1'. Obviously, in order to restore exactly the original image, the watermark reader needs to be informed of the pixels' position with gray values that have been shifted out of dynamic range. This requires the embedding of an overhead which reduces the watermark capacity. This modulation has been applied in the wavelet domain by Xuan *et al.* [6]. To reduce the size of the overhead, the method in [4] conducts a classification to identify parts of the image that can be watermarked in additive way. This process makes use of an approximated image derived from the image itself and which has the property to be identical for its watermarked version. The approach we propose in this paper follows the same strategy but conducts embedding in the wavelet domain.

The second class of methods refers to substitutive insertion techniques which come directly replace the signal by another one stemmed from a predetermined dictionary. We suggest classifying these methods into two subcategories: Lossless Compression Embedding techniques (LCE) and Expansion Embedding techniques (EE). The method proposed by Fridrich *et al.* [7] is one of the first LCE schemes. It substitutes one or more bit planes, by its compressed version concatenated with the message to be embedded. Xuan *et al.* extend this scheme [8] by applying compression on one or more middle bit-planes of integer wavelet coefficients to save space for data insertion. Differently to the above-mentioned LCE techniques, EE technique expand the dynamic of the

W. Pan, G. Coatrieux and Ch. Roux are with the Institut Telecom; Telecom Bretagne; Unite INSERM 650 Latim, Technopole Brest-Iroise, CS 83818, 29238 Brest Cedex 3 France (e-mail: {wei.pan, gouenou.coatrieux }@telecom-bretagne.eu).

N. Cuppens and F. Cuppens are with the Institut Telecom; Telecom Bretagne; UMR CNRS 3192 Labsticc, 2 rue de la Châtaigneraie, CS 17607, 35576 Cesson Sévigné Cedex France.

signal by shifting to the left the binary representation of signal samples, thus creating a new virtual least significant bit (LSB) which can be used for data insertion. LSB Substitution is applied on sample that cannot be expanded because of the limited signal dynamic or in order to preserve the quality of the watermarked image. Again some overhead needs to be embedded to store original LSBs. Tian *et al.* [9] apply EE to sub-band coefficients of the 1-D integer Haar wavelet transform of the image. In the same view, Lee *et al.* [10] divide the image into pixel blocks, and a watermark is embedded into the high-frequency wavelet coefficients of each block by LSB-substitution or EE. In order to enhance performances, Thodi *et al.* [11] apply histogram shifting to coefficients unexploited by EE for data embedding.

In this paper, we improved the method proposed in [4] by combining the invariant image classification with embedding in the wavelet transform of the image. Instead of modulating the difference between the image and its approximated version, as originally proposed, we directly embed de message into the conventional Haar wavelet coefficients of parts of the image identified as watermarkable. The proposed method has been tested on different sets of medical images from different medical image modalities but also some usual natural images in comparison with other methods.

The rest of the paper is organized as follows. The proposed scheme and its main parameters are presented in section II. Section III sums up performances analysis of our method and its comparison with two other recent approaches proposed in [10] and [11] in terms of imperceptibility and capacity of the watermark. Conclusions are given in section IV.

II. PROPOSED LOSSLESS WATERMARKING SCHEME

The method we propose is additive and leans on two main steps. The first one corresponds to a signal classification for the purpose of identifying parts of the image that can be watermarked in an additive way without introducing over/underflows. The classification we use is invariant to the second step of our approach which corresponds to the watermark embedding process. In that way, as the reader can retrieve parts of the image that have been additively watermarked, it just has to subtract the watermark to recover the original image. In this section, we present the classification process and then introduce the procedure for message embedding. If classification is conducted in the spatial domain, message embedding is performed in the wavelet domain.

A. Prevention of Over/underflows: classification

Our classification process works in the spatial domain and considers a specific watermark pattern W. It exploits an approximated image $\hat{I}$ derived from the image I to be watermarked. The originality of our approach resides in the fact that $\hat{I}$ remains unchanged after the watermark has been added to the image. Whence, we use the gray values of $\hat{I}$ to identify parts of I that can be or not watermarked additively.

To illustrate this classification process, let us consider a 2x2 pixels block $B = [x_0, x_1; x_2, x_3]$. B will be watermarked by adding or subtracting a watermark pattern $W = [w_0, w_1; w_2, w_3]$, only if the watermarked version $B_w = B + W$ does not lead to the introduction of over/underflows.

There exist many ways for approximating pixels of a block B. In our concern, B can be approximated through a linear function g of its pixels gray level values. The invariance constraint imposes to g the following property:

$$\hat{B} = g(B) = g(B \pm W) = g(B_w) \quad (1)$$

Where, $\hat{B} = [\hat{x}_0, \hat{x}_1; \hat{x}_2, \hat{x}_3]$ is the approximation of B. Once this requirement satisfied, the watermark reader will retrieve exactly $\hat{B}$.

If the watermark pattern is defined as $W = [+\Delta\ -\Delta;\ -\Delta\ +\Delta]$, where Δ is a positive integer value, then each pixel x_i of B can be approximated though a linear combination of its pixels:

$$\hat{x}_0 = (2x_0 + x_1 + x_2)/4\ ;\ \hat{x}_1 = (2x_1 + x_0 + x_3)/4$$
$$\hat{x}_2 = (2x_2 + x_0 + x_3)/4\ ;\ \hat{x}_3 = (2x_3 + x_1 + x_2)/4 \quad (2)$$

As stated earlier, the objective of the classification process is to identify watermarkable blocks. For that purpose, we characterize each block B^k by two values determined from its approximated version $\hat{B}^k$: $\hat{B}^k_{\min} = \min_{i=0..3}(\hat{x}_i,\ \hat{x}_i \in \hat{B}^k)$ and $\hat{B}^k_{\max} = \max_{i=0..3}(\hat{x}_i,\ \hat{x}_i \in \hat{B}^k)$. From this block characterization and considering the N_o and N_u blocks that, if watermarked by the addition or subtraction of W, lead to an overflow or underflow respectively, we can identify two thresholds T_{min} and T_{max}, which makes possible to distinguished these blocks from the others. T_{min} and T_{max} are defined as:

$$T_{min} = max_{\ n=0...Nu}(\hat{B}^n_{\min});\ T_{max} = min_{\ m=0...No}(\hat{B}^m_{\max}) \quad (3)$$

Therefore, a block B^k is considered watermarkable if it satisfies the following constraints:

$$\hat{B}^k_{\min} > T_{min} \text{ and } \hat{B}^k_{\max} < T_{max} \quad (4)$$

In some rare cases, the values of T_{min} and T_{max} may vary due to the embedding process. If this occurs the reader needs to be informed of the values of T_{min} and T_{max} the embedder identified. Thus, for perfect recovery of the original image, T_{min} and T_{max} needs to embedded along with the message [12].

Based on this first step, watermarkable blocks are identified. Message embedding is possible but at the same time it is constrained by the structure of W. In the next section, we illustrate how W can be used for embedding data in the wavelet representation of watermarkable blocks.

B. Embedding in the wavelet domain

The scheme we propose embeds the message into the wavelet coefficients of the image associated to watermarkable blocks while preserving the invariance of the above classification. Because of this latter constraint, the watermark pattern W added to pixel block in the spatial domain is predefined. Whence, the distortion that can be applied on one wavelet coefficient is also fixed but is easy to

measure because of the conventional wavelet transform linearity [13].

More clearly, if we consider the Haar wavelet transform then $Haar(B_w)=Haar(B)+Haar(W)=Haar(B)+W_{wave}$, where B, B_w, W and W_{wave} are the original pixel block, its watermarked version, the watermark pattern and its wavelet representation respectively.

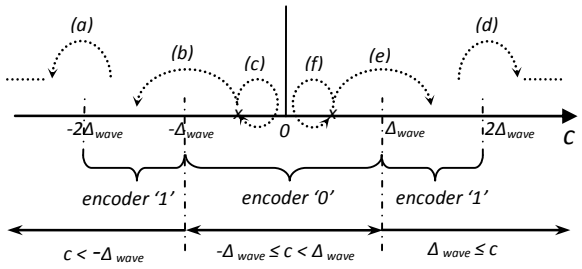


Fig.1. Reversible modulation of the wavelet coefficients for data embedding: if $c < 0$: (a) $c_{w0} = c - \Delta_{wave}$, (b) $c_{w1} = c - \Delta_{wave}$, (c) $c_{w0} = c$; if $c \geq 0$: (d) $c_{w0} = c + \Delta_{wave}$, (e) $c_{w1} = c + \Delta_{wave}$, (f) $c_{w0} = c$.

Specifically, the Haar wavelet transform of $W = [+\Delta\ -\Delta;\ -\Delta\ +\Delta]$ leads to $W_{\text{wave}} = [0\ 0;\ 0\ \Delta_{wave}]$. As a consequence, adding or subtracting W to one block in the spatial domain allows modulating the related *high-high* sub-band coefficient in the wavelet domain. Thus, invariance of the classification is preserved and it is possible to modulate wavelet coefficients in a reversible way. The reversibility property of a modulation ensures that from the watermarked coefficient it is possible both to extract the embedded watermark and to retrieve the original coefficient value.

The reversible modulation we propose to use is similar to histogram shifting [5], but is herein adapted to wavelet coefficients. .

Let us consider one Haar coefficient c issued from the first level *high-high* sub-band (*HH1*) of a watermarkable block. Depending on its value, c will be modulated in c_w, as illustrated in Figure 1, for the embedding of one bit b of the message:

1) If $-\Delta_{wave} \leq c < \Delta_{wave}$, c is considered as a carrier which can encode b:
   a) If $b = 0$, then $c_w = c$ (cases (c) and (f) in figure 1);
   b) If $b = 1$, and if $c \geq 0$, then c is right shifted to $c_w = c + \Delta_{wave}$ - case (e); otherwise, c is left shifted to $c_w = c - \Delta_{wave}$ - case (b);
2) If $c \geq \Delta_{wave}$ or $c < -\Delta_{wave}$, c is considered as a non carrier. To allow the reversibility of the modulation, it should be shifted to the left/right by subtracting/adding Δ_{wave} - cases (a) and (d). The reason is to avoid the overlap between non carrier and carrier coefficients.

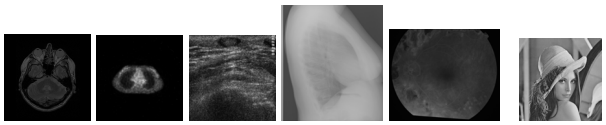
At the decoding stage, once the classifier rebuilt and watermarkable blocks retrieved, carrier coefficients can be re-identified. In fact, if watermarked coefficients belong to the range $[-\Delta_{wave}, \Delta_{wave}[$, they encode '0' and, for those within the ranges $[\Delta_{wave}, 2\Delta_{wave}[$ or $[-2\Delta_{wave}, -\Delta_{wave}[$, they encode '1'. Coefficients outside these ranges are non carrier blocks. Once the message extracted and interpreted, the watermark reader just has to restore the image by adding or subtracting W to the watermarked block following the same rules.

In this example, only the *HH1* wavelet sub-band has been considered. The same procedure, i.e. classification followed by message embedding, can be applied to the *high-low* (*HL1*) and *low-high* (*LH1*) sub-bands but with different watermark patterns: W $[+\Delta\ +\Delta;\ -\Delta\ -\Delta]$ and $[+\Delta\ -\Delta;\ +\Delta\ -\Delta]$ respectively. By watermarking successively each sub-band the capacity can be increased.

III. Experiments

A. Image database and measures of performance

The proposed watermarking method has been tested over several series of images (Figure 2). For medical images, five image modalities have been considered: three 12 bits encoded MRI (magnetic resonance imaging) volumes of 79, 80 and 99 axial slices of 256x256 pixels respectively, three 16 bits encoded PET (positron emission tomography) volumes of 234, 213 and 212 axial slices of 144x144 pixels respectively, three sequences of 8 bits encoded US (ultrasound image) images (14 of 480x592 pixels, 9 and 30 of 480x472 pixels respectively), forty two 12 bits encoded X-ray images of 2446x2010 pixels and thirty 8 bits encoded Retina images of 1008x1280 pixels. We have also considered some multimedia image of reference. In this paper, we report the results achieved with Lena.



(a) (b) (c) (d) (e) (f)

Fig. 2. Image samples from our test set (a) MRI of the head-axial slice of 256x256 pixels, 12 bits encoded. (b) PET –image of 144x144 pixels, 16 bits encoded, (c) US –image of 480x592 pixels encoded on 8 bits, (d) X-ray image of 2446x2010 pixels, 12 bits encoded, (e) Retina image of 1008x1280 pixels, 8 bits encoded, (f) Lena grayscale image of 512x512 pixels, 8 bits encoded.

To objectively quantify algorithms' performances, different indicators have been considered: the capacity rate C expressed in *bpp* (bit of message per pixel of image) and, in order to quantify the distortion between an image I and its watermarked version I_w, the peak signal to noise ratio (*PSNR*):

$$PSNR = 10\log_{10}\left(\frac{NM\left(2^p-1\right)^2}{\sum_{i,j=1,1}^{N,M}\left(I(i,j)-I_w(i,j)\right)^2}\right) \quad (5)$$

where p corresponds to the image depth, N and M correspond to the image dimensions.

In the following experiments, the embedded message is binary and randomly generated according to an uniform distribution.

B. Experimental results

Results are given in Table 1 and Figure 3 in terms of

capacity and distortion.

Performances indicated in Table 1 depend on the parameter Δ and the sub-bands that have been used for embedding. Given results in terms of capacity and distortion correspond to the mean and standard deviations per image. As it can be seen, our scheme allows a watermark capacity close to 0.2 *bpp* with *PSNR* about 51 *dB* for US, 51 *dB* for Retina images and 63 *dB* for X-ray images. However, capacity performances are quite small for MRI and PET images. Such a difference can be explained by the black image background which occupies an important place in MRI and PET images (i.e. 25% pixels of the MRI sample images are equal to '0'). In fact, our method embeds data within blocks that contains some information. Blocks with null gray value are not watermark. Nevertheless, even if the constraint is to preserve the image quality at best, our scheme is always applicable for the insertion of a short message such as a digital signature of 256 bits long to assert image integrity.

	Δ = 1 (only in *HH1*)		Δ = 1 (in *3 sub-bands*)		Δ = 2 (in *3 sub-bands*)	
	C	*PSNR*	*C*	*PSNR*	*C*	*PSNR*
MRI	0.006 (0.004)	78.62 (0.82)	0.011 (0.008)	74.07 (0.9)	0.022 (0.015)	68.25 (0.98)
PET	0.029 (0.02)	101.31 (1.06)	0.057 (0.05)	97 (1.15)	0.051 (0.05)	92.47 (1.24)
US	0.2 (0.02)	51.1 (0.34)	0.3 (0.05)	47.5 (1.2)	0.37 (0.05)	42.18 (1.13)
Retina	0.22 (0.012)	51.19 (0.16)	0.59 (0.046)	46.6 (0.33)	0.55 (0.04)	41.6 (0.34)
X-ray	0.04 (0.02)	73.59 (1.94)	0.09 (0.05)	69.56 (2.9)	0.16 (0.08)	63.87 (2.89)
Lena	0.13	50.12	0.33	45.04	0.51	39.86

Tab. 1. Capacity and distortion measurements for our approach in application to MRI, PET, US, Retina, X-ray and Lena images.

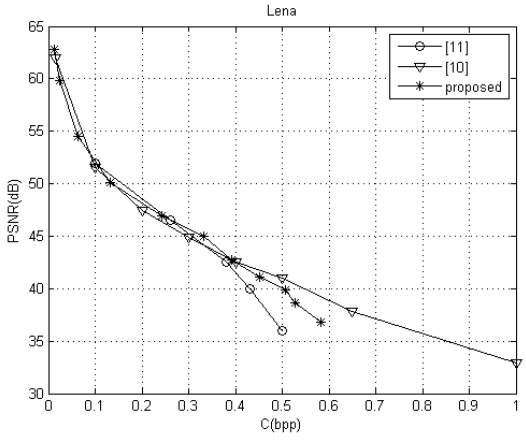


Fig. 3. Comparison of embedding capacity (C) versus distortion (PSNR) with existing reversible schemes [10] and [11] for the test image Lena.

In Figure 3, we compare our technique with two of the best methods proposed in the literature [10] and [11] in application to the grayscale Lena image. As it can be seen, our method gives a compromise from 0.012 *bpp* / 62.82 *dB* to 0.58 *bpp* / 36.76 *dB* simply by modifying the parameter Δ. In general, performances are quite similar to those of methods [10] and [11]. However, it allows slightly better capacities for highest PSNR values. For a *PSNR* of 45 *dB*, it can also achieve 0.33 *bpp*. But when the capacity becomes greater than 0.4 *bpp*, the embedding strength Δ is greater obviously magnifying rapidly the distortion when more than one sub-band is watermarked.

IV. Conclusion

In this article, we have proposed a new lossless watermarking scheme which makes use of an image approximation that is invariant to the insertion process and which is exploited to decide whether or not a pixel block can be modified. Then, message embedding is conducted in the Haar wavelet coefficients. According to the linearity of the wavelet transform, the invariance of the classification is preserved and the message is reversibly embedded.

As illustrated by the experimental results, this embedding scheme offers a high capacity and low distortion for both medical images and natural images. Future work will focus on extending the proposed algorithm to generalized wavelet transform in order to achieve better a capacity/distortion compromise.

References

[1] G. Coatrieux, L. Lecornu, B. Sankur, and Ch. Roux, "A Review of Image Watermarking Applications in Healthcare," in *Proc. of the IEEE EMBC Conf.*, New York, USA, 2006, pp. 4691–4694.

[2] C. W. Honsinger, P. Jones, M. Rabbani, and J. C. Stoffel, "Lossless recovery of an original image containing embedded data," US Patent application, Docket No.:77102/E-D, 1999.

[3] C. De Vleeschouwer, J. F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 97–105, Mar. 2003.

[4] G. Coatrieux, M. Lamard, W. Daccache, J. Puentes, and C. Roux, "A low distortion and reversible watermark application to angiographic images of the retina," in *Proc. of the IEEE EMBC Conf.*, Shanghai, China, 2005, pp. 2224–2227.

[5] Z. Ni, Y. Shi, N. Ansari, and S.Wei, "Reversible data hiding," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2003, vol. 2, pp. 912–915.

[6] Xuan, G. Yao, Q., Yang, C., Gao, J.: Lossless Data Hiding Using Histogram Shifting Method Based on Integer Wavelets. IWDW 2006, LNCS-4283 (2006) 323-33

[7] J. Fridrich, J. Goljan, and R. Du, "Invertible authentication," in Proc. of Int. Conf. SPIE, Security and Watermarking of Multimedia Content, San Jose, CA, Jan. 2001, pp. 197-208.

[8] G. Xuan, J. Chen, J. Zhu, Y. Q. Shi, Z. Ni, and W. Su, "Lossless data hiding based on integer wavelet transform," in *Proc. MMSP*, St. Thomas, Virgin Islands,2002, pp. 312–315.

[9] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.

[10] S. Lee, C. D. Yoo, T. Kalker, "Reversible Image Watermarking Based on Integer-to-Integer Wavelet Transform Information Forensics and Security," *IEEE Trans. Info. Forensics and security*, vol. 2, no. 3, pp. 321 – 330, Sept. 2007.

[11] D. M. Thodi and J. J. Rodriquez, "Expansion Embedding Techniques for Reversible Watermarking,," in *IEEE Trans. Image Processing*, vol.16, no.3, pp. 721-730, March 2007.

[12] G. Coatrieux, C. Le Guillou, J. M. Cauvin, C. Roux, "Reversible watermarking for knowledge digest embedding and reliability control in medical images," *IEEE Trans. on Information Technology in Biomedicine*, vol.13, pp. 158-165, March 2009.

[13] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Trans. Patt. Anal. Mach. Intell.,* 11 (1989), pp. 674–693.