31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Comparison of some reversible watermarking methods in application to medical images

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

Abstract—**Several reversible watermarking schemes have been proposed for images of sensitive content, like medical imaging, for which any modification may affect their interpretation. In this work, we distinguish these methods according to the way watermark insertion is conducted: additive and substitutive. Some of these approaches have been tested on different sets of medical images issued from three distinct modalities: Magnetic Resonance Images, Positron Emission Tomography and Ultrasound Imaging. Comparison analysis has been conducted with respect to several aspects including data hiding capacity and image quality preservation. Experimental results show different limitations which depend on the watermark approach but also on image modality specificities.**

I. Introduction

With the advances of Internet technology, especially in healthcare, images can be cross-exchange in right time allowing new medical practice through for example telediagnosis, teleconsultation services. At the same time, ensuring the security of exchanged medical information, including both protection and reliability guaranties, becomes a challenging concern.

In general speaking, Watermarking allows inserting a message, also called a watermark, in a host document by modifying the host content in an imperceptible way. For image, the message is attached at the signal level by modifying the image gray values. Whence the hosted message and the image are intimately attached and independently of the image file format. By its ability to introduce a protection level the nearest as possible of the data, watermarking has been shown as a complementary mechanism to enhance medical image security [1] in applications devoted to medical images where information is mostly secured by means of cryptographic tools like encryption. Watermarking can rise up medical image reliability by asserting its integrity (I.e. proof that the information has not been modified by non-authorized people) and its authenticity (i.e. an evidence that the information belongs to the correct patient and is issued from the right source). To do so, the embedded message may for instance consist in a digital signature of the image pixels [2]. Nevertheless, besides security aspects, data embedding can be used to make the image more informative through the insertion of meta-data [3].

For medical images, it is widely expected that the watermark shouldn't hinder the qualitative perception of the image. This mandatory constraint implies that the interpretation of the image content by a specialist shall remain unchanged after message insertion. However, the majority of watermarking methods irreversibly alter this content. Such distortions may be low-level when the watermark insertion is weighted by use of a visual perception model [4], but to our knowledge none of these models have been validated in the case of medical images. Consequently, these distortions may introduce a doubt about image validity.

Reversible or lossless watermarking has been proposed to overcome this issue. It allows the user to reconstruct the original image after having extracted the watermark (i.e. by removing image distortion). However, once the watermark has been removed, the image is no more protected, just like for data encryption. So even if removing the watermark is possible, most applications (data storage, transmission and also processing if the watermark doesn't interfere with the result [3]) have a high interest to keep it as long as possible in the image in order to continuously protect the information. Whence, in our view, even for reversible watermarking, the imperceptibility property has to be guaranteed in the medical domain. The reversible property has an interest for watermark content update and also for incoming image processing which can be applied to the original image.

Several reversible watermarking methods have been proposed [5-17]. They introduce more or less visible distortions with varying insertion capacities. Capacity is the amount of information that can be embedded into one image and which is expressed in bit of message per pixel (*bpp*). In this paper, we have tested some of these methods among different sets of medical image issued from different modality: MRI (magnetic resonance imaging), PET (positron emission tomography) and US (ultrasound imaging). Before comparing these methods with respect to the criterions given above in section 3, we provide in section 2 a classification of these methods depending on their additive/substitutive nature. Conclusions are made in section 4.

W. Pan, G. Coatrieux, J. Montagner 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, julien.montagner, christian.roux}@telecom bretagne.eu).

N. Cuppens and F. Cuppens are with the Institut Telecom; Telecom Bretagne; LUSSI Department, 2 rue de la Châtaigneraie, CS 17607, 35576 Cesson Sévigné Cedex France (e-mail: {nora.cuppens, fredric.cuppens}@telecom bretagne.eu).

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

II. REVERSIBLE WATERMARKING TECHNIQUES

Two classes of reversible watermarking methods may be distinguished: additive methods and substitutive methods.

A. *Additive insertion*

In the case of an additive insertion, the message *m* to be embedded is first transformed into a watermark signal *w*, and then added to the host signal *s* leading to the watermarked signal s_w: $s_w = s + w$.

Additive insertion has been primarily applied in the spatial domain in which the image pixel gray level values are limited to a fixed dynamic (2^p possible gray levels for an image of *p* bits depth). Consequently, watermark addition may lead to over/underflows, it means that modified pixel values may fall out of the allowed gray value range $[0\ldots2^p\text{-}1]$. Obviously, such a problem occurs also when embedding is conducted in a transformed domain like in the wavelet or DCT domain.

Different strategies have been proposed to overcome over/underflow problem. One approach introduced in [5] consists in using modulo arithmetic. Insertion equation $I_w = (I + w) \bmod 2^p$ can however lead to a salt and pepper noise due to jumps between congruent values of the dynamic. An improved version of this method has been proposed in [9] where visual distortions are minimized by making use of arithmetic modulo on short cycles, obtained by splitting the signal dynamic in ranges of short size.

An alternative is based on a classification of the signal before embedding. In [6], the proposed reversible watermarking scheme is based on an image signal estimation which is invariant to the insertion process. More clearly the image and its watermarked version will have the same image of reference. In a first time, the estimated image is used to decide whether or not a pixel block can be modified. The image of reference serves a classification procedure for identifying blocks that if modified lead to an over/underflow. Then insertion is conducted on the authorized parts of image by modulating the difference between the original image and its estimated version. As the image of reference is the same for the watermarked image, the decoder can easily retrieve watermarked parts of the image.

A third approach regroups methods that modulate the image histogram in the spatial or a transformed domain. The method suggested by Ni *et al.* in [7] shifts a range of the image histogram. This range is identified by the couple (*zp*,*pp*), where *zp* and *pp* correspond respectively to the gray levels with the smallest ("zero-point") and the highest ("peak-point") number of pixels. This range is shifted by adding or subtracting one gray level from the peak point toward the zero point in order to leave one gray level (a "gap") near the peak point empty. Finally, pixels that belong to the peak point class are moved to the gap or left unchanged for message embedding. Two gray values are used to code the message. Consequently, the alteration is not more important that one gray level for the modified pixels. However, the embedded data cannot be recovered unless the position of initial peak point is known by the decoder. This modulation has been applied in the wavelet domain by Xuan *et al.* [17] where the identification of the couple (*zp*, *pp*) is simplified as integer wavelet coefficients have a "laplacian" distribution centered around '0'.

Leest *et al.* [8] have proposed a similar approach. This latter is based on creating "gaps" at the minimum and maximum luminance values in local histograms of 2×2 pixels blocks. However with this approach, positions of pixels which have the value 0 and 2^p-1 have to be embedded in the image to solve the over/underflow problem. As a consequence, embedding capacity decreases when the numbers of such a pixel increase.

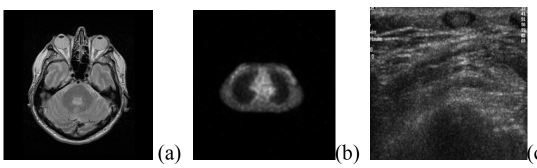
B. *Substitutive insertion*

Substitutive insertion technique differs from the additive in the sense that rather than disrupting the signal by adding a watermark, it comes directly to replace the signal by another one stemmed from a predetermined dictionary signal. For example: the basic LSB scheme removes the pixels' least significant bits by bits of the message to be embedded. To make this scheme reversible, original binary values should be preserved and communicated to the decoder. Unlike this irreversible modulation technique, several solutions have been proposed. We class them into two categories: Lossless Compression Embedding (LCE) techniques and Expansion Embedding (EE) techniques.

Fridrich *et al.* [9] have shown that there exists a bit-plane *B* in the original image *I*, such as *B* can be lossless compressed and disrupted randomly, without visible distortion in *I*. If such a bit-plane exists, it may be replaced by the compressed version of a binary message *m*. The insertion capacity of such a method is $|B|$-$|\text{compress}(B)|$ bits, where $|\cdot|$ denotes the cardinal. With this method the capacity is rather small and since more effective strategies have been proposed.

Xuan *et al.* have proposed an insertion technique on coefficients of the integer wavelet transform [10]. They losslessly compress one or more middle bit-planes of integer wavelet coefficients to save space for data embedding. Celik *et al.* [11] have proposed a generalized LSB substitutive technique, which firstly converts the binary message ($w \in \{0,1\}$) to *M*-ary watermark ($w \in \{0,1,\ldots, M\text{-}1\}$) by arithmetic coding. For example, a watermark *w* can be converted from $(1000101011)_2$ to $(4\ 2\ 1\ 0)_5$, where $M = 5$. Then the lowest *M*-levels of the pixels of the original image are replaced by the *M*-ary watermarks: $p_w = M\lfloor p/M \rfloor + w$, where *p* and p_w represents the host pixel and its watermarked version and, $\lfloor . \rfloor$ the "floor" operator meaning "the greatest integer less than or equal to". The original values are losslessly compressed using the CALIC algorithm [18].

Differently to the above-mentioned LCE techniques, Tian's algorithm [12] may be the first one to use the EE technique for reversible watermarking. EE shifts to the left the binary representation of an integer value *h* to watermark (*h* can be a gray value or a transformed coefficient), thus creating a new virtual LSB that can be used for insertion: $h_w = 2h + b$, where h_w is a watermarked value and b is one bit of the message. To control the insertion distortion, the EE is combined with LSB substitution: $h_w=2\lfloor h/2\rfloor+b$. Tian's scheme [12] applies this solution to the difference x of two adjacent pixels. To minimize the image distortion, they apply EE to the small value of x. As stated previously, for those rests x, the LSB of the original value is directly substituted by one bit of the message. To distinguish at the reader stage which pixels' differences have been expanded, a binary location map L is required. In Tian's scheme L is lossless compressed and added to the embedded message with the original LSBs. Alattar extended this scheme by applying the EE to a generalized integer transform [13]: several bits are embedded into vectors of adjacent pixels.



(a) (b) (c)

Fig. 1. 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.

In the same way, Lee et al. [14] divide a $M \times N$ pixel image into 16x16 pixel blocks, and a watermark is embedded into the high-frequency wavelet coefficients of each block by LSB-substitution or EE technique. Their location map is of small dimension ($(M \times N)/(16 \times 16)$) and doesn't required to be compressed. Always in the same vein, Xuan *et al.* in their scheme [15] introduce a threshold T. If the absolute value of integer wavelet coefficient is lower than T, then EE is applied for data embedding. With this approach, it may be difficult for the reader to distinguish of watermarked and non-watermarked coefficients. To solve this problem, the coefficients which have the absolute values higher or equal to T should be shifted to the left or right according to their signs by T - 1 or T. So all watermarked coefficients that carry the message are in the interval] $-2T + 1$, $2T$ [. With this apporach no more needs of a location map. This is almost the same for the method proposed by Thodi *et al.* [16], which combines Tian's method and this shifting pretreatment for gaining better performance.

	MRI		PET		US	
	C (*bpp*)	PSNR (*dB*)	C (*bpp*)	PSNR (*dB*)	C (*bpp*)	PSNR (*dB*)
[7]	0.26 (0.011)	73.00 (0.46)	0.20 (0.013)	97.98 (0.92)	0.05 (0.053)	52.63 (4.19)
[8]	0.20 (0.007)	75.72 (0.067)	0.22 (0.033)	99.57 (0.29)	0.04 (0.013)	53.19 (0.52)
[6]	0.0031 (0.002)	78.43 (0.84)	0.020 (0.016)	100.79 (1.16)	0.101 (0.032)	48.51 (0.20)

Tab. 1. Capacity, distortion measurements for additive methods: Ni *et al.* [7], Leest *et al.* [8] and Coatrieux *et al.* [6].

All of these methods are known to be fragile, *i.e.* the watermarks will not survive any image alteration. This is why these methods are at first proposed for data integrity control. For this study, we have implemented some of the most recent or original methods, and indicated by their authors as efficient on usual test images such as "Lena", "Baboon" ... Three of these schemes are additives: Ni *et al.* [7], Leest *et al.* [8], Coatrieux *et al.* [6] and two substitutives: Xuan *et al.* [15], Thodi *et al.* [16].

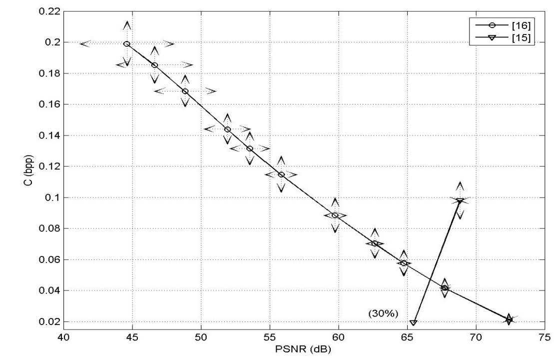


Fig. 2. PSNR/Capacity compromise for substitutive methods [15] and [16] in the case of MRI images.

III. EXPERIMENTS

A. Image database and measures of performance

Experiments were conducted on three modalities: three 12 bits encoded MRI volumes of 79, 80 and 99 axial slices of 256x256 pixels respectively, three 16 bits encoded PET volumes of 234, 213 and 212 axial slices of 144x144 pixels respectively, and, three sequences of 8 bits encoded US images (14 of 480x592 pixels, 9 and 30 of 480x472 pixels respectively). Figure 1 gives some examples of our data set. In the following, the binary message to be embedded is randomly generated based on a uniform distribution.

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

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

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

B. Experimental results

Results are given in figures 2 to 4 and in table 1. They provide the compromise between capacity and distortion depending on the image modality.

If we consider additive schemes in Table1, [7] and [8] allow a watermark capacity close to 0.2 bpp with PSNR about 73-75 dB for MRI, 97-99 dB for PET, but their performances decrease with US images. On the contrary, [6] provides a higher capacity for US images. Such a difference can be explained by the use of the black background by [7] and [8], background which occupies an important place in MRI and PET images. [6] only watermarks non null signal, this is why it has a low capacity for these images.

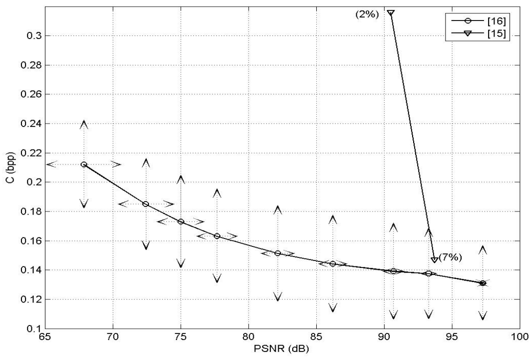


Fig. 3. Averaged PSNR/Capacity compromise for substitutive methods [15] and [16] in the case of PET images.

For MRI and PET images, the results of substitutive methods [15] [16] are less effective than additive methods [7] [8]. For US images, these methods are more efficient, [15] and [16] propose a compromise of 0.14 bpp/48.77dB and 0.22 bpp/48.44 dB respectively. One explanation lies on the nature of the host signal and on the watermark strategy exploited. Some methods keep limited as they require embedding along with the message a lot of information for reconstructing the original image. For some images, [15] was not able to insert a message as the amount of information for reconstruction was more important than the offered capacity. For PET images, only 7% of 659 images can be watermarked with a compromise of C/PSNR = 0.15bpp/93.73dB.

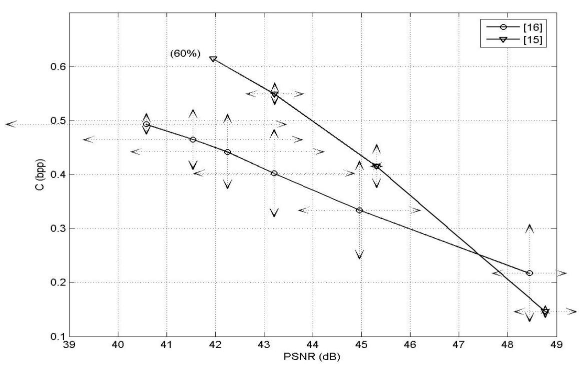


Fig. 4. PSNR/Capacity compromise for substitutive methods [15] and [16] in the case of US images.

From a more applicative point of view, if the constraint is to preserve the image quality at best, [6] is more appropriate for MRI and PET and [8] is better for US images. If the objective is the insertion of a big amount of information: [7] offers a compromise of 0.26 bpp/73 dB for MRI, [8] proposes 0.22 bpp/99.57 dB for PET and at least, for US images [15] proposes a compromise of approximately 0.55 bpp/43.3 dB. Regardless the imaging modality, [16] proposes a satisfactory compromise of 0.02 bpp/72.41 dB, 0.13 bpp/97.27 dB and 0.22 bpp/48.44 dB for MRI, PET and US images respectively.

IV. CONCLUSION

Reversible watermarking is of main concern imagery of sentive content. However, in order to beneficiate of the watermarking's interests which mainly is to provide a continuous protection, it is mandatory to propose reversible methods which minimize distortion and maximize capacity.

In this article, five reversible watermarking methods have been implemented and compared. Each of these methods provides varying results in terms of capacity and invisibility. Different limits have been identified due the specific nature of medical images. As a consequence, the next generation of techniques may consider the signal specificities. Most of methods are fragile and the question of robustness is largely open. Up to now, a few methods have been proposed. This is also one of the upcoming challenges.

REFERENCES

[1] G. Coatrieux, H. Maître, B. Sankur, Y. Rolland, R. Collorec, "Relevance of Watermarking in Medical Imaging," in Proc. of IEEE EMBS Int. Conf ITAB, Arlington, USA, 2000, pp.250-255.

[2] G. Coatrieux, J. Puentes, L. Lecornu, C. Cheze Le Rest, C. roux, "Compliant secured specialized electronic patient record platform," in Proc. of IEEE EMBS Int Conf. D2H2, USA, 2006, pp. 156–159.

[3] 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.

[4] A. Piva, M. Barni, F. Bartolini, V. Capellini, "Exploiting the cross-correlation of RGB-channels for robust watermarking of color images," in *Proc. of IEEE Int. Conf. on ICIP*, vol. I, 1999, pp. 306-310.

[5] 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.

[6] 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.

[7] 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.

[8] A. Leest, M. Veen, and F. Bruekers, "Reversible watermarking for images,"in Proc. of Int Conf. SPIE, Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA, Jan. 2004.

[9] 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.

[10] 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.

[11] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Reversible data hiding," in *Proc. IEEE ICIP*, 2002, vol. 2, pp.157–160.

[12] 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.

[13] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Trans. on Image Processing*, vol. 13, no. 8, pp. 1147–1156, Aug. 2004.

[14] 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.

[15] G.R. Xuan, Y.Q. Shi, C.Y. Yang, Y.Z. Zheng, D.K. Zou, P.Q. Chai, "Lossless Data Hiding Using Integer Wavelet Transform and Threshold Embedding Technique," in proc. of Int. Conf**.** *Multimedia and Expo,* 2005, pp. 1520 – 1523.

[16] 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.

[17] G.R. Xuan, Q.M. Yao, C. Yang, J. Gao, "Lossless Data Hiding Using Histogram Shifting Method Based on Integer Wavelets." IWDW 2006, LNCS-4283 (2006) 323-332.

[18] X. Wu, "Lossless compression of continuous-tone images via context selection, quantization, and modelling," *IEEE Trans.on Image Proc.*, vol. 6, no. 5, pp. 656–664, May 1997.