Nello Balossino; Davide Cavagnino; Marco Grangetto; Maurizio Lucenteforte; Sergio Rabellino. A high capacity reversible data hiding scheme for radiographic images. COMMUNICATIONS IN APPLIED AND INDUSTRIAL MATHEMATICS. 4 pp: 1-14.
DOI: 10.1685/journal.caim.444

Communications in Applied and Industrial Mathematics, ISSN 2038-0909, vol, issue, (YYYY) 1–14

DOI: 10.1685/YYYYCAIMXXX

A High Capacity Reversible Data Hiding Scheme for Radiographic Images

Nello Balossino, Davide Cavagnino, Marco Grangetto,
Maurizio Lucenteforte, Sergio Rabellino

Università di Torino, Computer Science Department
{nello.balossino, davide.cavagnino, marco.grangetto, maurizio.lucenteforte, sergio.rabellino}@unito.it

Communicated by Associated editor

Abstract

Digital watermarking has been growing in popularity in latest years and plays a key role in the protection of intellectual property rights of multimedia content. In particular, reversible fragile watermarking is a feasible technique when security aspects like authentication, data integrity, and recovery of original data are required. This paper proposes a data hiding scheme to guarantee all the mentioned requests for radiographic images which are transmitted through a generic transmission channel. In particular, it is possible to ensure authenticity and to verify the integrity of the transmitted digital image. Moreover, both random errors and malicious attacks, that are likely to occur in an unprotected communication environment such as e-mail and Web, are jointly identified. The proposed scheme covers the combined use of digital signature techniques and fragile and reversible watermarking in the spatial domain to embed the patient data and the digital signature in the host image, ensuring high-capacity and high visual quality.

Keywords: Fragile image watermarking, data hiding, data authentication.

AMS Subject Classification: 68N04, 68U04, 65Y04

1. Introduction

Digital watermarking and steganography techniques are used in application contexts like digital rights management, authentication and fingerprinting. These methods embed some information in a digital object (called host object) in an imperceptible way by slightly changing the original data in the object itself. It is important to highlight that in such a case the original data has been modified, with potential impact on any subsequent usage of the object Watermarking techniques generally realize a compromise between the capacity of the watermark (number of bits of information that it is possible to insert in the image, also called payload) and its imper-

Received YYYY MM DD, in final form YYYY MM DD
Published YYYY MM DD

ceptibility: increasing the watermark capacity generally amplifies the image noise so that the artefacts produced by the algorithm become more visible. In some fields of application the end user cannot tolerate to deal with imprecise information that can be misleading. This is the case, for example, of a radiologist who is examining a radiographic image to obtain a medical diagnosis: obviously, a judgement based on corrupted data is not acceptable in the scientific and medical fields.

Watermarking algorithms may be subdivided into three main categories: fragile, semi-fragile and robust. In the first case the inserted watermark is designed to be altered when a modification (even minimal) is applied to the watermarked image, thus proving that the image has been compromised. Watermarks in the semi-fragile category can survive to a process of slight changes of the image, like a high quality JPEG compression. Robust watermarks are devised to resist processing operations aimed at their removal, and are typically used in contexts like copyright protection or tracking of origin.

Another important feature of watermarking techniques is the reversibility, meaning that it must be possible to reconstruct the original content from the watermarked one. In such a way any evaluation can be performed on the original data, while simultaneously allowing to carry some amount of information (the watermark). For medical applications reversible watermarking algorithms are clearly essential. Reversible watermarking, also known as invertible or lossless, is designed to be applied mainly in contexts where the authenticity of a digital image has to be provided and the original content is needed at the decoding side.

It is possible to develop techniques that allow the embedding of patient data and/or diagnosis along with the information used for authentication and reversibility. In this way the receiver may retrieve the mentioned data after a successful authentication; moreover he will be able to recover the original host image.

In this paper a reversible watermarking method is presented and discussed. The remainder of this paper is organized as follows: in section 2 an overview of reversible watermarking with an emphasis on medical image applications is presented; section 3 introduces our method, while experimental results are given in section 4. In section 5 our conclusion are drawn.

2. Reversible watermarking for medical image authentication

The majority of the published works in the field of reversible watermarking can be divided into three classes: approaches based on correlation, spatial domain and transform domain.

In [1] a spatial domain method for watermarking medical images is presented; the algorithm is non-reversible because the pixel's Least Significant Bits (LSBs) are substituted with the watermark bits, but are not saved in any way to be restored later. Moreover the method proposes to encrypt the patient data to be inserted into the watermark, but the security of the encryption function is not discussed.

The paper [2] discusses the requirements for the storage and transmission of patient medical data, presenting watermarking as a possible solution to some of these requirements. This work also highlights that in the medical context a watermarking system must take into account some user functionalities like integrity and authentication along with reversibility and imperceptibility.

Among the spatial domain approaches the algorithm proposed by Tian [3,4] is one of the most well-known. This technique is based on the evaluation of differences between neighbouring pixel, that are used to implement the so called difference expansion (DE). DE allows one to achieve jointly high-capacity and high visual quality. Taking hint from this approach, similar methods have been proposed, using DE applied to three or more pixels, embedding two or more bits. Alattar in [5] embeds two bits in each triplet of pixels, while in [6] he proposes an extension that hides triplets of bits in the difference expansion of quads of 2×2 adjacent pixel values. Finally, in [7], an additional generalization of the previous algorithm is proposed; in this case DE works on groups of N pixel values arranged in a vector, achieving a maximum capacity of 1 bit per pixel (bpp).

In [8], Ni et al. propose a reversible data hiding algorithm with high-capacity and high visual quality. The algorithm is based on histogram modification. It first finds a zero point (i.e. no value of that gray level is present in the host image) or a minimum point in case that a zero point does not exist, and then the peak point (most frequent gray level in the host image). The number of bits that can be embedded into the image equals to the frequency value of the peak point. Consider now the marked image and let a be the peak point, and b the zero point (or minimum point), with $a < b$. The first step in the embedding process (after scanning in sequential order) is to increase by 1 the value of pixels between $a + 1$ and $b - 1$. As a result the range of the histogram is shifted to the right-hand side by 1, leaving the value $a + 1$ empty. The image is scanned once again in the same sequential order and when a value of a is encountered, such value is incremented by 1 if the bit value of the data to embed is 1; otherwise, the pixel value is left unaltered. At the decoding side, the algorithm scans in sequential order (the order used in the embedding phase) the marked image. When a pixel having grayscale value $a + 1$ is encountered, a bit '1' is extracted and the pixel value is decremented by 1. If a pixel with value a is encountered, a bit '0' is extracted. The described algorithm is applied in the simple case of one pair of minimum point and maximum point. A possible extension of the proposed method considers the case of multiple pairs of maximum and minimum points. The multiple pair case can be treated as a recursive application of the technique. The authors show that the proposed algorithm guarantees a lower bound of the PSNR as high as 48 dB on monochromatic images with 8 bpp.

In [9] a reversible watermarking algorithm, that can be applied to DICOM images from various sources, e.g. X-Ray, magnetic resonance, computer tomography, is presented. This algorithm implements a dual layer watermarking to increase the storage capability, being able to save into the image both patient and user data and also Cyclic Redundancy Codes (CRC) for the localization of tampered areas. The algorithm works on small data blocks (of size 2×2 pixels) modifying some pixels if they satisfy a condition w.r.t. a secret key selected pixel: the process is stated as reversible, even if in the paper the extraction step is not clearly explained. The only constraint of the algorithm is that pixels cannot span the whole dynamic range, but typically the higher gray levels (9 in the case examined in the paper) cannot be used. The localization capability is obtained using CRC-16 on blocks of 16×16 pixels. In the examples presented, the quality of the watermarked image varies from a PSNR of 34 dB to 35 dB.

In [10] it is described a method for the insertion of hidden data in an image; the algorithm stores data bits into the discrete wavelet transform coefficients of the medium (LH and HL) and high (HH) frequencies, and provides for the reconstruction of the original image (reversibility). The algorithm substitutes some bits (identified by bitplanes) in the coefficients with the payload data to insert. In order to guarantee reversibility, the substituted bits are compressed and inserted as part of the payload. A compromise between achievable compression of the substituted bits (related to the bitplane) and quality of the resulting image has to be made at the embedding stage. Moreover, the algorithm takes into account the possible pixel overflow, as this situation may happen when transform coefficients are modified.

3. Proposed algorithm

Our method gets foundation on a histogram analysis of some radiographic images in DICOM format. We noticed that not all gray levels are used in the representation of the image; indeed, the histograms follow a sort of "comb" representation, in which highly populated gray levels are adjacent to empty levels. As an example, the histogram of a thoracic DICOM image is displayed in figure 1(a), while the comb structure for the same image is emphasized in figure 1(b) where a limited range of the histogram is shown.

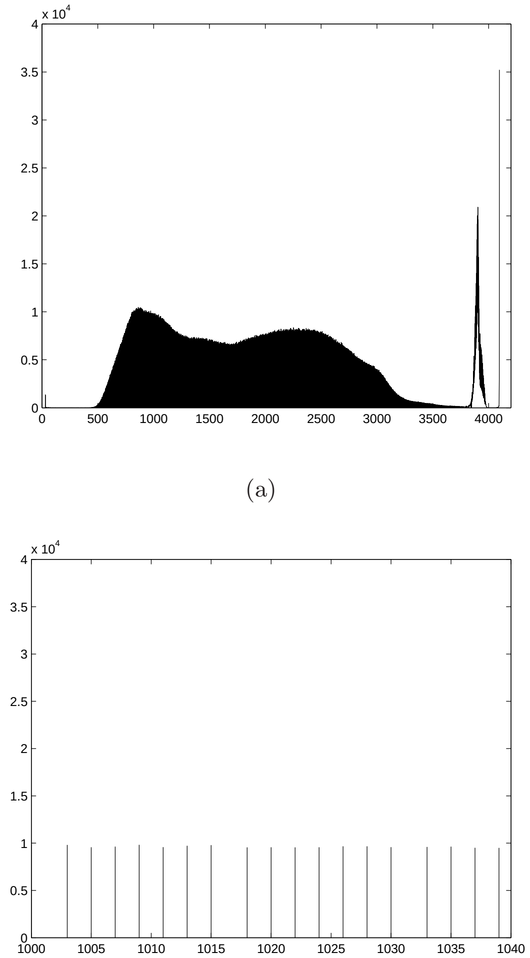
This comb behaviour has been reported in many different kinds of radiographic images, from thoracic to dental, produced by different hardware. We can easily assume this behaviour is due to the post-processing steps responsible of image equalization and contrast enhancement. Such algorithms warp the image histogram, thus potentially leaving some level values unused.

The method proposed in this paper is a generalization of [8], where a modification in the image histogram is required before the insertion process. In [8], if a highly populated level l_i is not adjacent to an empty level, it is necessary to increase by one unit the values of the pixels belonging to the intermediate luminance values, in order to set to zero the frequency of the level $l_i + 1$. On the other hand, we note that the comb shape of the histograms of the DICOM images analysed in this paper, permits to avoid modifying the histogram. Indeed, it is possible to identify directly the levels having a non-zero frequency that are adjacent to an unused level. We will call these levels *changeable*. The advantages are twofold: the algorithm of insertion/extraction is greatly simplified and the introduced distortion is much smaller (fewer pixels in the image are changed, i.e, only those associated to a watermark bit 1 insertion). Our algorithm requires that the list of changeable levels is coded as side information, so as to allow the decoder to retrieve both the watermark and host image.

The list of changeable levels is encoded as a sequence of differences from one level to the next one, starting from lowest changeable level l_0. The level l_0 is stored in the LSBs of the first pixels of the image, and the original LSBs are kept as part of the watermark (for extraction and reconstruction purposes).

The essential information to be encoded in the pixels of gray level l_0 are the LSBs of the first pixels, the number of differences with their coding length, and the sequence of differences to reconstruct the list of changeable levels. From the list it is possible to compute the pixel values in which the remaining part of the watermark is stored. The watermark may contain payload data (for example patient data and diagnosis eventually encrypted), and a message authentication code for integrity and authentication purposes or a digital signature if non-repudiation is required.

Given that the capacity of a gray level may be limited, we consider as changeable levels only those whose frequency exceeds a given minimum threshold f_{min}. This allows for a trade-off between capacity and overhead



(a)

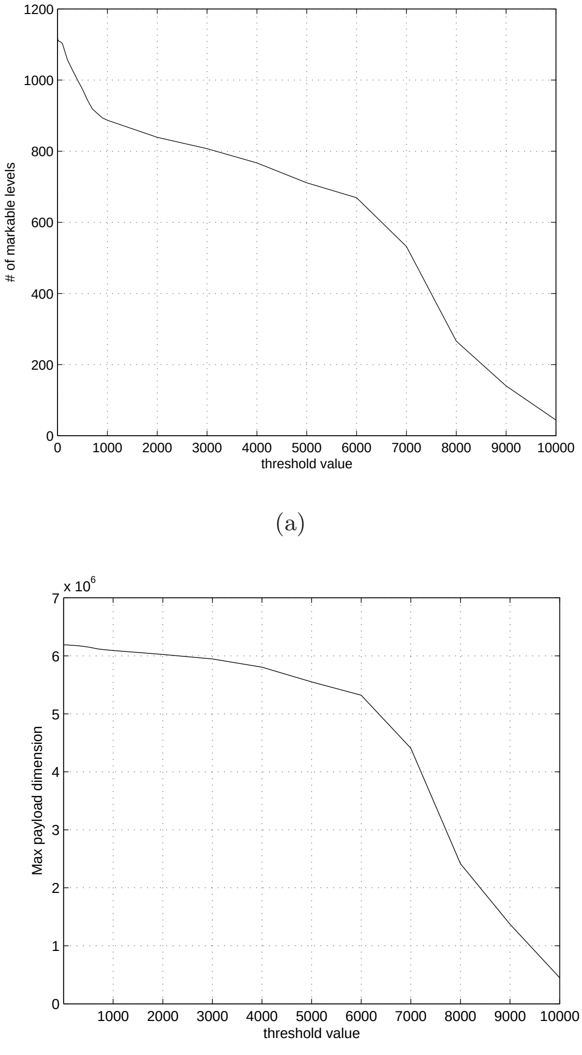
(b)

Figure 1. (a) Histogram of a DICOM image and (b) a magnification of a region of the histogram.

in encoding the differences.

Figure 2 shows (a) the number of changeable levels and (b) the corresponding payload size (i.e. the number of marked pixels) versus the minimum frequency threshold of such levels for a chest X-ray image with 12 bpp (i.e. 4096 grayscale levels), taken using a Kodak DR 5100 system.

Figures 4 and 5 show high level block diagram for the embedding and extraction processes. In the following subsections we figure out the pseudocode of the insertion and recover processes. We assume that the frequency of level l_0 is adequate to store all the data regarding the differences to encode the changeable levels; if not, the level l_1 should be used like the previous one to continue the encoding of the differences, and so on.



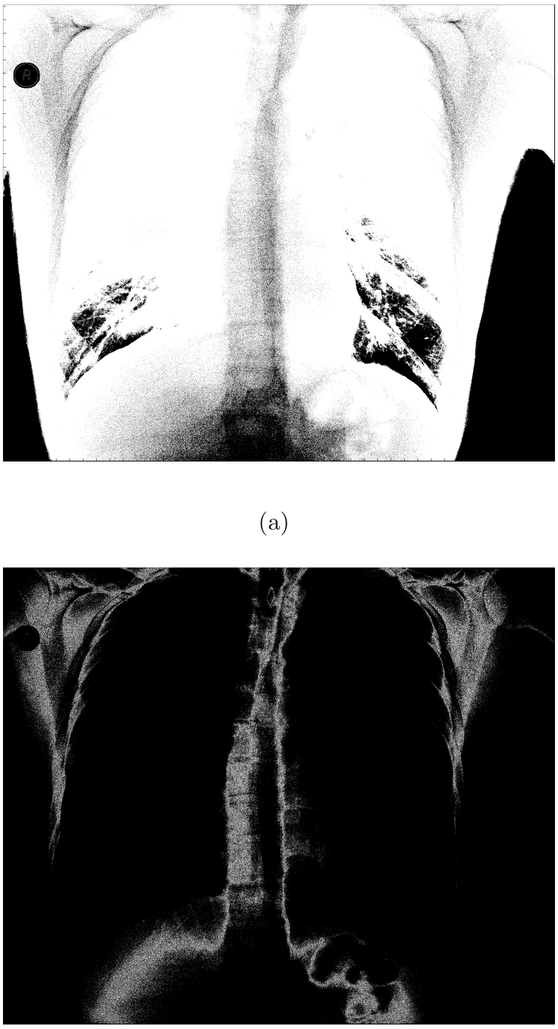
(a)

(b)

Figure 2. (a) Number of watermarkable levels and (b) maximum payload (i.e. number of watermarkable pixels) vs. threshold of minimum frequency f_{min}.

3.1. *Pseudo-code for watermark insertion process:*

```
1. read the host image $I_o$;
2. determine the color depth $k$;
3. extract the first $k$ pixels and store their LSBs in a variable $S$;
4. calculate the histogram of the remaining pixels of $I_o$: the histogram values
   will be $f(0), f(1), \ldots, f(2^k - 1)$ for the $k$ bit-depth image;
```



(a)

(b)

Figure 3. Markable pixels (in white) for (a) $f_{min} = 4000$, (b) $f_{min} = 10000$.

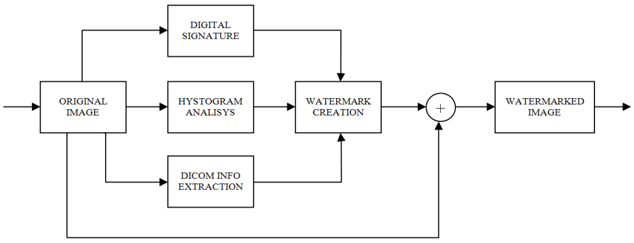


Figure 4. Block diagram of the watermark insertion process.

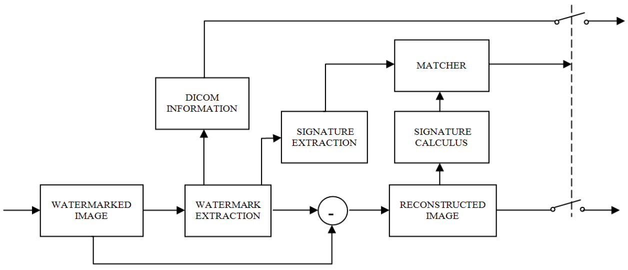


Figure 5. Block diagram of the extraction/recovery process.

5. determine a list of levels $l_0, l_1, l_2, \ldots, l_n$ such that $f(l_i) >= f_{min}$ and $f(l_i + 1) = 0$, where f_{min} is a minimum value that allows to reduce the levels encoding overhead;
6. compute the list of delta values $\Delta_i = l_i - l_{i-1}$ with $i = 1..n$;
7. encode l_0 in the LSBs of the first k pixels;
8. compute the bit string B as the concatenation of $W_L, S, L, n, \Delta_1, \Delta_2, \ldots, \Delta_n, PayloadData, Signature$ where
   (a) W_L is the bit length of the following watermark, represented on 32 bits,
   (b) L are 8 bits expressing the bit length of the encoding of each Δ,
   (c) n is the 8 bit representation of the number of the Δ values,

```
    (d) PayloadData are the data to be stored in the image (eventually encrypted)
    (e) Signature is the digital signature of the host image along with the PayloadData;
 9. insert the bits of $B$ into the pixels (in raster scan order from the $(k+1)$-th) having a gray level $l_0$, and then using the pixels with levels $l_1, l_2, \ldots, l_n$ in raster scan order: the insertion of a bit of value $b$ into a pixel of gray level $l_i$ will assign the gray level $l_i + b$ to the pixel;
10. write the obtained image $I_w$.
```

Note that the maximum bit capacity of the watermark is

$$C = \sum_{i=0}^{n} f(l_i) \tag{1}$$

3.2. *Pseudo-code for extraction and recover process:*

```
1. read the watermarked image $I_w$;
2. determine the color depth $k$;
3. extract the first $k$ pixels;
4. decode $l_0$ from the LSBs of the first $k$ pixels;
5. from the pixels having gray levels $l_0$ and $l_0 + 1$, in raster scan order, decode the bits $b$ according to the formula $b = l - l_0$ where $l$ is the gray level of the pixel; set the value of each of these pixels to $l_0$;
6. the extracted bit sequence allows to obtain:
    (a) $W_L$: these 32 bits allow to determine the length of the whole watermark;
    (b) $S$: use these $k$ bits to restore the LSBs of the first $k$ pixels;
    (c) $L$: these 8 bits define the bit length of the following $\Delta$ values;
    (d) $n$: these 8 bits define number of the following $\Delta$ values;
    (e) $\Delta$ values: from the previous information it is possible to decode $\Delta_1, \Delta_2, \ldots, \Delta_n$ and thus compute $l_1, l_2, \ldots, l_n$;
7. from the remaining pixels with gray level $l_0$ or $l_0 + 1$ and then, in raster scan order, from $l_1, l_1 + 1, l_2, l_2 + 1, \ldots, l_n, l_n + 1$ it is possible to extract the PayloadData and the Signature; each pixel is restored to its original value;
8. the Signature is used to verify the authenticity of the image and of the data in the payload.
```

4. Experimental results

In this section we describe the results of the application of the proposed algorithm to a large set of medical images coming from different sources. Moreover, we compare the obtained results with the performance of the algorithm [10].

Given that our algorithm works in the spatial domain, it is effortless to estimate the watermarked image quality (in terms of PSNR) from the gray level histogram of the original host image. Let's assume to insert a binary watermark obtained from the optimal compression (or, analogously, encryption) of a stream: in this scenario the resulting binary sequence will have an entropy of 1 bpp, corresponding to an average frequency of 0.5 for both symbols 1 and 0. According to the previously presented algorithm, W_L is the watermark length in bit: on average, only one half of such information will force a gray level modification. These considerations allow to estimate very precisely the PSNR of the marked image as a function of W_L. In particular, the *mse* of the watermarked image with resolution $rows \times columns$ can be computed as:

$$mse = \frac{W_L}{2 \cdot rows \cdot columns} \tag{2}$$

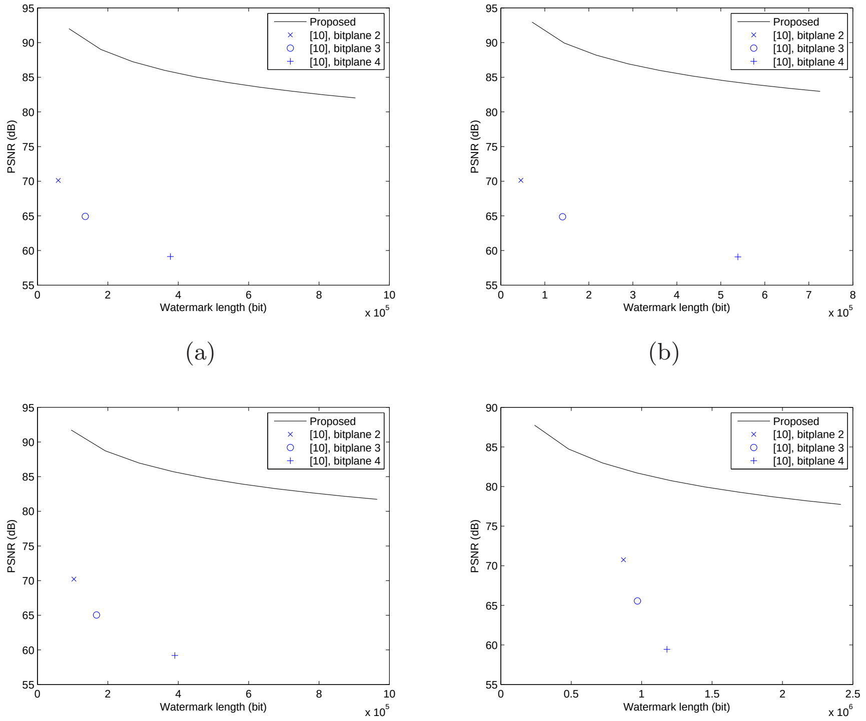
The Peak Signal-to-Noise Ratio (PSNR) of the watermarked image with respect to the host image is calculated as:

$$PSNR = 10 \cdot log_{10} \frac{maxgray^2}{mse} \tag{3}$$

where $maxgray = 2^{bitdepth} - 1$ is the maximum pixel value that can be represented using *bitdepth* bits.

In figure 6 the PSNR obtained with the proposed algorithm on 4 medical images is shown as a function of W_L. For comparison we report the results achievable with [10], as well. The watermark length of [10] is determined by the payload space, remaining after the insertion of the compressed original bitplane, and therefore is affected by the statistical properties of the removed bitplane bits. In figure 6 it can noted that the proposed algorithm outperforms [10] by at least 20 dB in a large range of watermark lengths.

In order to provide a reliable and complete evaluation, the proposed algorithm has been tested on a large database of 127 images provided by different medical units using equipments from different vendors. We inserted a watermark collecting the patient data, other information from the DICOM header and a 1024 bits digital signature. Obviously the watermark also contained the information necessary for the watermark extraction (as



(a)

(b)

(c)

(d)

Figure 6. PSNR (dB) vs inserted watermark size (bit).

described in section 3). The obtained results are reported in Table 1. The column *Watermark size* reports the size of the embedded data, consisting of information present in the DICOM header of the image along with the side information necessary for the extraction/verification process. The column *PSNR* contains the corresponding image quality; on the other hand *Available payload size* indicates the total amount of bits that can be stored in the image. It can be seen from the table that the available amount of data that can be inserted is always sufficient to contain the watermark.

Table 1. Results summary for a set of 127 DICOM images.

	Watermark size	PSNR	Available payload size
Min	36116	97.23	61020
Max	55239	100.51	5552632
Average	46139.94	99.62	1916323.31

5. Conclusions

In this paper it is presented an algorithm that can be used at the same time for the fragile watermarking of a digital image and for the insertion of useful data. The advantage of the method is that, in case no modification to the image occurs, the original host image may be reconstructed, making this algorithm reversible. Moreover, the authentication information (i.e. a signature or a Message Authentication Code) is contained in the image itself, as with fragile watermarked images.

The reversibility of the algorithm is particularly important for medical images: an algorithm that guarantees the originality of an image but performs even small pixel alterations as a result of its operations, is typically useless in the medical field.

The proposed algorithm exploits a characteristic of some radiographic images; it has been noted that the gray level histogram has a comb structure: often an highly populated gray level has neighbouring levels having null frequency. For this type of images it is possible to use this redundancy to insert authentication, user data and the required information to implement a reversible process.

We have compared the proposed algorithm with another reversible watermarking algorithm: the results show a very high quality of the images watermarked with the proposed method. The performance of our algorithm is high for the images having an histogram comb structure, but will suffer for images not having this characteristic: this is the reason of the better performance w.r.t. the algorithm we compared, which will work better with other kinds of images.

A further improvement of the developed algorithm can be obtained taking into account runs of zero frequency levels. For example, if a highly populated gray level l is followed by 7 unused grey levels, then 3 bits can be coded modifying a single pixel of grey level l. In this scenario, the list of delta values must be extended with the associated run length. This improvement increases the payload reducing the watermarked image quality, but due to the reversibility of our algorithm, this is not an issue.

Acknowledgments

The authors would like to thank AOU Molinette, ASL Alba-BRA and Dr. Nolè (dentist) for providing the X-ray images used in the developing phases of the proposed algorithm.

REFERENCES

1. D. Anand and U. C. Niranjan, Watermarking medical images with patient information, in *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, vol. 2, pp. 703–706, IEEE, 1998.
2. G. Coatrieux, H. Maitre, B. Sankur, Y. Rolland, and R. Collorec, Relevance of watermarking in medical imaging, in *Proceedings. 2000 IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, pp. 250–255, IEEE, 2000.
3. J. Tian, Reversible watermarking by difference expansion, in *Proceedings of Multimedia and Security Workshop at ACM*, pp. 19–22, 2002.
4. J. Tian, Reversible data embedding using a difference expansion, *IEEE Trans. Circuits Syst. Video Techn.*, vol. 13, no. 8, pp. 890–896, 2003.
5. A. M. Alattar, Reversible watermark using difference expansion of triplets, in *Proceedings of International Conference on Image Processing (ICIP 03)*, vol. 1, pp. 501–504, 2003.
6. A. M. Alattar, Reversible watermark using difference expansion of quads, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 04)*, pp. 377–380, 2004.
7. A. M. Alattar, Reversible watermark using the difference expansion of a generalized integer transform, *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1147–1156, 2004.
8. Z. Ni, Y. Shi, N. Ansari, and W. Su, Reversible data hiding, *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 3, pp. 354–362, 2006.
9. C. K. Tan, J. C. Ng, X. Xu, C. L. Poh, Y. L. Guan, and K. Sheah, Security protection of DICOM medical images using dual-layer reversible watermarking with tamper detection capability, *Journal of Digital Imaging*, vol. 24, no. 3, pp. 528–540, 2011.
10. G. Xuan, J. Zhu, J. Chen, Y. Shi, Z. Ni, and W. Su, Distortionless data hiding based on integer wavelet transform, *Electronics Letters*, vol. 38, pp. 1646–1648, Dec 2002.