Signal Processing: Image Communication 47 (2016) 160–169

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

A crypto-watermarking system for ensuring reliability control and traceability of medical images

Dalel Bouslimi [a,b,*], Gouenou Coatrieux [a,b]

[a] *Telecom Bretagne, Département Image et Traitement de l'Information, Brest 29238, France*
[b] *INSERM U1101, Laboratoire de Traitement de l'Information Médicale, Brest 29238, France*

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form
28 March 2016
Accepted 31 May 2016
Available online 16 June 2016

Keywords:
Decryption
Medical image
Reliability
Security
Traceability
Watermarking

ABSTRACT

In this paper, we propose a novel crypto-watermarking system for the purpose of verifying the reliability of medical images and tracing them, i.e. identifying the person at the origin of an illegal disclosure. This system couples a common watermarking method, based on Quantization Index Modulation (QIM), and a joint watermarking-decryption (JWD) approach. At the emitter side, it allows the insertion of a watermark as a proof of reliability of the image before sending it encrypted; at the reception, another watermark, a traceability proof, is embedded during the decryption process. The scheme we propose makes interoperate these two watermarking approaches taking into account risks of interferences between embedded watermarks, allowing the access to both reliability and traceability proofs. Experimental results confirm the efficiency of our system, and demonstrate it can be used to identify the origin of a disclosure even if the image has been altered.

1. Introduction

The rapid evolution of information and communication technologies has boosted the sharing, process and remote access to medical images [1]. If these technologies advances offer significant benefits for medical providers and patients, it is obvious that they also increase security issues in terms of [2]:

- Confidentiality – the property which guarantees that, in normal conditions, only authorized users have access to the information.
- Availability of personal medical information – the ability of an information system to be used under the normal conditions of access and exercise.
- Reliability – it corresponds to the confidence one can have into the information. It has two aspects: i) Integrity – a proof that a piece of information has not been modified in a non- authorized way; ii) Authentication – a proof of the information origins and of its attachment to one patient. Reliable pieces of information can be used confidently by the physician.

Another security concern is data traceability which aims at identifying the persons at the origin of an information disclosure.

Nowadays and in practice, protection of medical images is achieved by cryptographic means, especially encryption and digital signatures. The kind of protection that encryption mechanisms offer is however mostly "*a priori*" in the sense that an image is protected until its decryption or, equivalently, before the access to its content is granted. Several questions appear then about the security of the image once decrypted. Here comes the interest for an "*a posteriori*" protection that allows the user accessing the image content while maintaining it protected. A kind of protection the watermarking technology is appropriated for [3]. Basically, when it is applied to images, watermarking modifies or modulates the image pixels' gray level values in an imperceptible way in order to encode or insert a message (i.e. the watermark). The message can be a set of security attributes [3] that will protect the image while it is accessed. There exist three types of watermarking methods proposed for medical imagery [4]. The first class differentiates regions of interest (ROI) within images and then sets out to watermark within region of non-interest (RONI) [5]. Since only the non-relevant parts of the image are manipulated, the diagnostic capability is not compromised. The second class of approaches corresponds to lossless or reversible watermarking [6]; schemes in which the watermark can be removed with the exact recovery of the original image. It is however important to point out that once the watermark is removed, the image is not protected. The third class of approaches consists to use common lossy-watermarking methods while minimizing the distortion. In this case, the watermark replaces or alters some image details like the least significant bits (LSB) [7] or the Discrete cosine transform

* Corresponding author at: Telecom Bretagne, Département Image et Traitement de l'Information, Brest 29238, France.
E-mail address: Dalel.Bouslimi@telecom-bretagne.eu (D. Bouslimi).

http://dx.doi.org/10.1016/j.image.2016.05.021
0923-5965/

(DCT) coefficients [8].

Different approaches have been proposed in order to take advantage of the complementarity of these two mechanisms in terms of *a priori*/*a posteriori* protection. They mostly focus on copyright protection, where the objective is to identify the person (issuer or recipient) at the origin of an illegal distribution by mean of a watermark (a fingerprint).

Technically, depending on the way that watermarking and encryption are merged, we suggest to discriminate four main categories of methods:

- "*Watermarking Followed by Encryption*" (WFE) [9–11] where a watermark is embedded within a copy before encrypting it. Inserted data are only available in the spatial domain after decryption.
- "*Encryption Followed by Watermarking*" (EFW) [12–19] – Some methods proposed in [12,13] exploit asymmetric homomorphic encryption. Based on the homomorphic property [20], an operation in the encrypted domain on the cipher text (e.g. '$\times$', '+' …) is translated into another operation on the plain text (e.g. '+', 'x' …). It is then possible to modify an encrypted image and consequently insert a watermark into the spatial domain. With this kind of approach, the embedded watermark is only available in the spatial domain. Another strategy described in [14] consists in inserting a message into the encrypted image by substituting one bit of a pixel block with one of the message to be embedded. At the recipient side, each block of the image is decrypted twice testing the two possible values of the modified bit (i.e. 0 or 1). The decrypted block with the minimum standard deviation is the original block. Herein, the message is extracted before the image decryption by reading the modified or watermarked bits and is removed from the image during the decryption process. In [15], Zhang applies a lossless compression onto the LSBs of an image encrypted with a stream cipher algorithm. The space gain is then used for information insertion. The method proposed in [16] relies on the insertion into the image of a predefined watermark (a pre-watermark) before the encryption process. Message insertion and extraction are then conducted into the encrypted image. It is the impact of this data hiding process onto the "pre-watermark" that gives access to the message after image decryption. A novel concept, called Vacating Room Before Encryption (VRBE), has been proposed [17–19]. Its principle is to reversibly watermark an image before encrypting it so as to leave some free space into the encrypted domain for message embedding.
- "Joint *Watermarking/Decryption*" (JWD) [21–24] – which aims at reducing time computing and complexity on the server side by conducting embedding of the fingerprint (the recipient or client identifier) during the decryption process. To do so, the same encrypted content is sent to all clients, but each of them receives a different decryption key. Processes of watermarking and decryption are jointly conducted. As the decryption keys are different, the decrypted contents are also different. The differences between the original content and its decrypted version correspond to the watermark/fingerprint that identifies the client.
- "*Joint Watermarking/Encryption*" (JWE) [3,25] – these methods conduct data encryption and data watermarking in a single operation. It allows the insertion of two messages, which can be read/ extracted in the encrypted and spatial domains.

Approaches, which belong to *JWE* and *WFE* categories allow essentially ensuring data reliability, but not data traceability when the recipient is not known *a priori*, i.e. before the encryption process. In such a case, the recipient identifier, that allows identifying him if he or she illegally redistributes the image, cannot be embedded into the transmitted image since the watermarking process is conducted before or during the encryption process. Unlike these approaches, those of *EFW* and *JWD* categories have been essentially proposed to ensure data traceability in the context of copyright protection. The majority of *EFW* approaches is based on asymmetric homomorphic encryption; kind of algorithms not suited for medical images due to the fact they represent a large amount of information.

In this work, we focus on providing the reliability and the traceability of medical images while ensuring their confidentiality. As stated above, reliability and traceability are key issues for medical data sharing, compared to the multimedia domain where copyright issues are of major concern. To do so, we propose to merge a common lossy watermarking method, which is QIM (Quantization Index Modulation) [26], for the embedding before the encryption process of a watermark containing an image reliability proof, with a JWD approach in order to insert a second watermark containing a traceability proof. Among JWD algorithms, we decided to work with the joint watermarking-decryption scheme: ST-DM Secure Embedding proposed by Piva et al. [22], due to the fact that it relies on QIM. Due to the fact the second watermark insertion can interfere with the first one, we study the interferences in between watermarks. In particular, we establish the constraints to satisfy so as to avoid such watermark interferences and to be able to extract both watermarks and the security attributes they carry on.

The rest of this paper is organized as follows. In Section 2, we independently present the watermarking and the JWD approaches we used before introducing their combination in Section 3. In Section 4, we determine the probability the insertion of the second message during the decryption process interferes or modifies the message embedded before the encryption process. Section 5 presents some experimental results considering ultrasound and retina images. Before concluding, we discuss the overall performance of the proposed system in Section 6.

2. Cryptographic and watermarking primitives

2.1. Watermarking primitive: binary QIM modulation

QIM is today one of the most popular substitutive watermarking technique due to its robustness and its simplicity [26]. It relies on quantifying an image component (e.g. a group of pixels or of transformed coefficients) according to a set of quantizers based on codebooks in order to insert a binary message. Basically, to each message m_i, $m_i \in \{0,1\}$, QIM associates the elements of a codebook C_{m_i} such as:

$$C_{m_i} \cap C_{m_j} = \varnothing, i \neq j \tag{1}$$

Substituting one image component X by its nearest element in the codebook C_{m_i} allows the insertion of m_i.

An implementation of QIM firstly generates a randomly unit normal vector $U \in \mathbb{R}^N, U = [u_1, \ldots, u_i, \ldots, u_N]$, based on the watermarking key K_w. As depicted in Fig. 1, U is then divided into non overlapping intervals of equal size $\Delta/2$, where Δ represents the quantization step, an user defined QIM parameter. To satisfy (1), each interval is associated to a codebook C_{m_i}. The bit m_i is then embedded by moving the vector of pixels $X \in \mathbb{N}^N, X = [P_1, \ldots, P_i, \ldots, P_N]$, in such a way that its projection on U corresponds to the center of the nearest interval that encodes the desired bit (see Fig. 1). In such a case, the codebook C_{m_i} can be defined as follows:

$$C_{m_i} = \{c_{m_i,k}\} = \{(2k + m_i)\Delta, k \in \mathbb{Z}\} \tag{2}$$