Combination of Watermarking and Joint Watermarking-Decryption for Reliability Control and Traceability of Medical Images

D. Bouslimi, G. Coatrieux, *Member, IEEE,* M. Cozic and Ch. Roux, *Fellow, IEEE*

Abstract— **In this paper, we propose a novel crypto-watermarking system for the purpose of verifying the reliability of images and tracing them, i.e. identifying the person at the origin of an illegal distribution. 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 proof of traceability, is embedded during the decryption process. The scheme we propose makes interoperate such a combination of 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 physician at the origin of a disclosure even if the image has been modified.**

I. Introduction

The rapid evolution of multimedia and communication technologies offers new means of sharing and remote access to patient data [1-2]. In particular, medical imaging is already called to play important roles in applications like telesurgery, telediagnosis and so on. But at the same time, this ease of transmission and sharing of data increases security issues in terms of [3]:

- Confidentiality, which means that only authorized users can access to patient data.
- Availability, which guarantees access to medical information in the normal scheduled conditions of access and exercise.
- Reliability, which is based on: i) Integrity- a proof that the information has not been altered or modified by non-authorized persons; 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 one must nowadays consider is data traceability which aims at identifying the persons at the origin of an information disclosure.

Among available security mechanism, encryption is commonly used so as to ensure medical data confidentiality. However, once decrypted, one piece of information is no longer protected and it becomes hard to verify its integrity and its origin. From this point of view, encryption appears as an "*a priori*" protection mechanism. Watermarking has been proposed as a complementary mechanism to improve the security of medical images [4]. 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). Nowadays, different approaches combine encryption and watermarking in order to benefit of their complementarity 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, four categories of methods can be distinguished according to the way watermarking and encryption are merged:

- "*Watermarking Followed by Encryption*" (WFE) [5]- where watermark embedding is done before encryption.
- "*Encryption Followed by Watermarking*" (EFW) [6]- these methods take advantage of homomorphic encryption that allows inserting within an encrypted image an encrypted watermark. This operation conducted in the encrypted domain is like inserting the watermark in the clear text.
- "Joint Watermarking/Decryption" (JWD) [7]- where fingerprint embedding is conducted during the decryption process. This essentially reduces time computing and complexity on the server side.
- "*Joint Watermarking/Encryption*" (JWE) [8-9]- where a watermark is embedded during the encryption process. It allows verifying the image reliability in both encrypted and spatial domains.

In this work, in order to ensure both reliability and traceability of medical images, we propose to couple a common watermarking method based on QIM (Quantization Index Modulation) [10] with a JWD approach. Herein, among possible JWD algorithms, we decided to work with the LUT-based secure embedding scheme [7]. The system we propose allows inserting two messages or equivalently two watermarks: one before the encryption process containing reliability proof of the image and the second during the decryption process containing a traceability proof. Due to the fact the second watermark insertion can interfere with the first one, considering QIM, we establish the constraints to verify so as to be able to extract both watermarks and the security attributes they carry on.

The remainder of this paper is organized as follows. In section II, we independently present QIM and the JWD approach we use. We describe the proposed system in section III and evaluate its performance through some experimental on ultrasound images in section IV.

II. Cryptographic and Watermarking Primitives

A. Watermarking primitive: QIM

Quantization Index Modulation [10] relies on quantifying

D. Bouslimi, G. Coatrieux are with the Institut Mines - Telecom; Telecom Bretagne; Unite INSERM 1101 Latim, Technopole Brest-Iroise, CS 83818, 29238 Brest Cedex 3 France (e-mail: {dalel.bouslimi, gouenou.coatrieux}@telecom-bretagne.eu).

M. Cozic is with MEDECOM, Plougastel Daoulas 29470, France (e-mail: mcozic@wanadoo.fr).

Ch. Roux is with Institut Mines-TELECOM, Mines Saint-Étienne, CS 62362 - F-42023 Saint-Étienne, France

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

the components of one image according to a set of quantizers based on codebooks in order to insert a message. More clearly, to each message m_i issued from a finite set of possible messages $M = \{m_i\}_{i=0,\dots,q}$, QIM associates the elements of a codebook C_{m_i} such as:

$$C_{m_i} \cap C_{m_j} = \emptyset, i \neq j \ . \qquad (1)$$

Substituting one component of the image by its nearest element in the codebook C_{m_i} thus allows the insertion of m_i. Let us consider a vector of image pixels $X \in \mathbb{N}^n$ and a binary message, i.e. $m_i \in \{0,1\}$. In practice, to embed a binary message m_i within X, X is firstly projected on a unit normal vector $U \in \mathbb{R}^N, U = [u_1, \dots, u_i, \dots, u_N]$, randomly generated based on the watermarking key K_w. As depicted in Fig. 1, the line engendered by U is then divided into non overlapping intervals of equal size $\Delta/2$, where Δ represents the quantization step, a QIM parameter. To satisfy (1), each interval is associated to a codebook C_{m_i}. m_i is then embedded within X by moving it such that its projection on U corresponds to the center of the nearest interval that encodes m_i. In this context, the codebook C_{m_i} can be defined as follows:

$$C_{m_i} = \left\{c_{m_{i,k}}\right\} = \{(k + m_i/2)\Delta, k \in \mathbb{Z}\} \ . \qquad (2)$$

Consider p_x the projection of X on U, the watermarked version of X, i.e. X_w, is thus given by

$$X_w = X + (p_{xw} - p_x)U \ . \qquad (3)$$

where p_{xw} is the projection of X_w on U and which satisfies $p_{xw} = Q_{m_i}(p_x)$ with $Q_{m_i}(p_x)$ a function that determines the nearest element of p_x within C_{m_i}.
Let us consider $\hat{X}$ the received X_w, a possible attacked version of X_w. During the extraction step, the knowledge of the interval (or the codebook) to which belongs the projection $\widetilde{p_x}$ of $\hat{X}$ on U is enough to identify the embedded message. Thus, the detected m_i, $\widehat{m_\iota}$, is given by:

$$\widehat{m_\iota} = arg \min_{m_i \in \{0,1\}} \min_{c_{m_{i,k}} \in C_{m_i}} \left|c_{m_i,k} - \widetilde{p_x}\right| \ . \qquad (4)$$

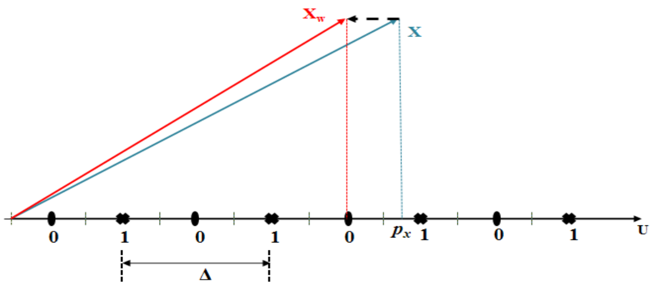

Figure 1. Insertion of a binary message with the QIM

B. ST- DM Secure Embedding Scheme

This solution is a generalization of the Chameleon cipher [11], the principle of which is as follows. Let us consider the plaintext or clear text is constituted of bytes (e.g. grayscale pixels). During the encryption operation, four entries of a random look up table (LUT) are selected and then xored with a plaintext component to obtain the ciphertext. The decryption process is similar to that of encryption except that the decryption LUT is different. This one is derived from the encryption LUT by injecting errors in some entries. As a consequence, the deciphered text will be slightly different from the original text. This difference forms a fingerprint that identifies the client or the receiver. The "ST- DM Secure Embedding Scheme" we use follows this principle [7]. It operates as follows. Let us consider a vector X of N components (e.g. pixels). The server generates an encryption LUT E of l entries: $[E[1], \dots, E[i], \dots, E[l]]$ where $E[i]$ is also of N components randomly generated following a Gaussian distribution $\mathcal{N}(0, \sigma_E)$. To encrypt X, s entries of E are secretly selected using a pseudo random generator initialized with the encryption key K_e. These entries are then summed and added to X to obtain its encrypted version X_e. If we consider t_j the selected entries, $1 \leq j \leq s$ and $1 \leq t_j \leq l$, from E to encrypt X, X_e is then given by:

$$X_e = X + \sum_{j=1}^{s} E[t_j] \ . \qquad (5)$$

The same encrypted version of X, X_e, is then sent to each client k, along with its decryption LUT D_k. To build D_k, the server generates a watermark W_k and combines it with the encryption LUT E. Indeed, the i^{th} entry of D_k. is given by:

$$D_k[i] = -E[i] + W_k[i] \ . \qquad (6)$$

The jointly watermarking/decryption process is similar to the encryption. Using K_e, the set of indices t_j is generated. The s entries of D_k are summed and the result is added to X_e for giving access to the watermarked vector X_w :

$$\begin{aligned} X_w &= X_e + \sum_{j=1}^{s} D_k[t_j] \ . \\ &= X + \sum_{j=1}^{S} W_k[t_j] = X + w^k \ . \qquad (7) \end{aligned}$$

where $w^k = \sum_{j=1}^{S} W_k[t_j]$. So, the difference between the encryption and the decryption LUTs allows the insertion into X of a fingerprint w^k that identifies the kth client.

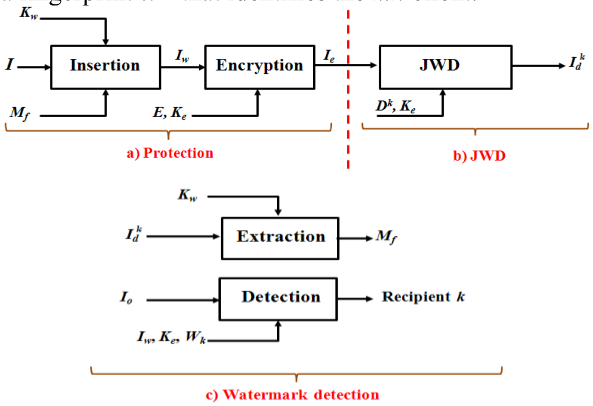

Figure 2. The different steps of our system (a) watermarking of the image; (b) joint watermarking decryption of received image; (c) reliability and traceability control.

III. Description of the Proposed System

The purpose of our system is to ensure the confidentiality of an image I through encryption and its reliability as well as to trace its final user by means of watermarking. Here, we assume that the practitioner "emitter" sends the same watermarked/encrypted image to K_u practitioners named as "receivers". As illustrated in Fig. 2, it relies on three main procedures: protection, Joint Watermarking/ Decryption

(JWD) and watermark detection. The first one allows us to insert into an image I a message M_f. This message contains security attributes assessing the image reliability. The obtained watermarked image I_w is then encrypted using the encryption LUT E and the encryption key K_e. The resulting image, i.e. I_e, is then sent to the k^{th} receiver along with its decryption LUT D_k. At the reception of I_e, the k^{th} receiver jointly decrypts and watermarks it into I_d^k with the help of JWD. I_e contains thus the watermark M_t^k that identifies the recipient. Based on M_f and M_t^k, it is thus possible to verify the reliability of I_d^k and to identify the physician who illegally redistributes it, respectively. This system is however operational if the insertion of M_t^k does not interfere with M_f.

A. Image Protection

As depicted in Fig. 2, this procedure is constituted of two main steps: insertion and encryption. The first step allows the insertion of a message M_f into I using K_w and QIM modulation (see section II.A). In the sequel, we consider M_f as a binary sequence: $M_f = \{b_1, \ldots, b_i, \ldots, b_m\}, b_i \in \{0,1\}$. To do so, I is firstly splitted into m non-overlapping blocks of N pixels $\{X_i\}_{i=1,\ldots,m}$, $X_i = \{P_1, \ldots, P_N\}$. One bit b_i is inserted into X_i with QIM. The watermarked version of X_i, i.e. X_{iw}, is thus given by (see Eq. 3)

$$X_{iw} = X_i + (p_{iw} - p_i)U \quad . \tag{8}$$

where $p_{iw} = Q_{b_i}(p_i)$ and $p_i = \langle X_i, U\rangle$.

X_{iw} is then encrypted using the encryption LUT E and K_e as described in section II.B. As the image is encoded on d bits, the encrypted version of X_{iw} (X_{ie}) is as follows (see Eq.5)

$$X_{ie} = \left(X_i + \textstyle\sum_{j=1}^{s} E[t_j]\right) mod\ 2^d \ . \tag{9}$$

The resulting watermarked encrypted image I_e is thus given by: $I_e = \{X_{ie}\}_{i=1,\ldots,m}$.

B. JWD procedure

As seen in section II.B, for each receiver k, the emitter generates a decryption LUT D_k by adding to the encryption LUT a watermark W_k, $W_k = \{W_k[j]\}_{j=1,\ldots,l}$, such as $W_k \sim \mathcal{N}(0, \sigma_w)$ (see Section II.B). The receiver k uses D_k and K_e to decrypt I_e. Each block X_{ie} is decrypted using Eq. 7:

$$\begin{aligned} X_{id}^k &= \left(X_{ie} + \textstyle\sum_{j=1}^{s} D_k[t_j]\right) mod\ 2^d \ . \\ &= \left(X_i + M_{ti}^k\right) mod\ 2^d \ . \end{aligned} \tag{10}$$

where X_{id}^k is the decrypted version of X_{ei} and $M_{ti}^k = \sum_{j=1}^{S} W_k[t_j]$. The watermark M_t^k is thus given by $M_t^k = \{M_{ti}^k\}_{i=1,\ldots,m}$. With this procedure, the main issue to be considered is to guarantee that the watermark embedded during the decryption process doesn't modify M_f. This imposes constraints on the generation of W_k.

***Choice of the watermark W_k* -** By definition, each bit b_i of M_f is QIM embedded in the projection of X_i on the vector U. To read this bit after the JWD process, it is necessary to ensure that the projection of X_{id}^k on U remains in the same interval that encodes b_i (see Fig. 3). As the projection of X_{di}^k on U is given by:

$$\langle X_{di}^k, U\rangle = \langle X_{iw}, U\rangle + \langle M_{ti}^k, U\rangle \quad . \tag{11}$$

the distortion on U due to the insertion of M_{ti}^k during the decryption process is $\langle M_{ti}^k, U\rangle$. So, to guarantee that M_f is not modified by JWD, M_{ti}^k must satisfy:

$$\langle M_{ti}^k, U\rangle < \Delta/4 \quad . \tag{12}$$

The ideal solution to minimize this distortion is to choose a value of M_{ti}^k such as $\langle M_{ti}^k, U\rangle = 0$, i.e. M_{ti}^k and U are orthogonal, i.e.

$$W_k[j] = \delta_j U_\perp \ \text{ where } j=1,\ldots,l \ . \tag{13}$$

where $U_\perp$ is a vector orthogonal to U and δ_i represents the magnitude of $W_k[j]$.

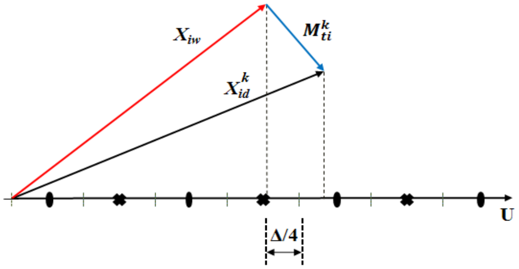

Figure 3. Projections on U of the watermarked block X_{iw} and of its decrypted version X_{id}^k.

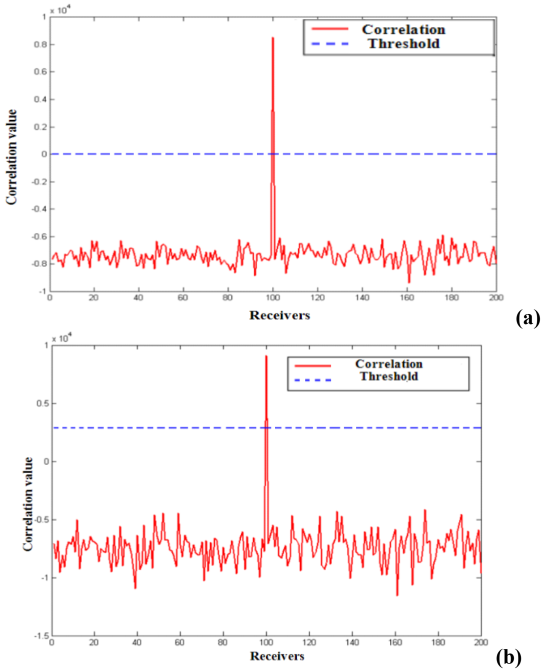

Figure 4. Detection of the watermark M_t^k of the k=100th receiver (the 100th over 200 possible receiver) at the origin of an illegal distribution in the presence of additive white Gaussian noise attack of standard deviation σ: a) $\sigma = 1$, b) $\sigma = 10$. Dotted line corresponds to a decision threshold computed as in [12]

C. Watermark detection procedure

***Extraction of M_f*-** I_d^k is firstly decomposed into blocks of N pixels, $\{X_{di}^k\}_{i=1,\ldots,m}$, from which bits of M_f are extracted.

***Identification of the receiver at the origin of a disclosure*-** As the detector, the emitter or a third party, has the original image and its watermarked version, we opted for the non-blind detection. This also offers a better tradeoff between robustness and image quality, due to the fact the original image knowledge allows better isolation of the watermark. The process we propose to follow consists firstly in calculating the watermarks of the K_u recipients, i.e. $\{M_t^k\}_{k=1,\dots,k_u}$, as follows (see Eq. 10)

$$M_t^k = \{M_{ti}^k\}_{i=1,\dots,m},\ M_{ti}^k = \sum_{j=1}^{S} W_k[t_j] \ . \tag{14}$$

and by next the correlation between each watermark M_t^k and the difference W between I_w and the observed image I_o, i.e. W=I_w-I_o. The receiver of which the watermark has the maximum correlation value and greater than a user defined fixed decision threshold is considered at the origin of the illegal distribution of I_o. The threshold detection is computed as in [12].

IV. Experimental Results and Discussion

Experiments were conducted on 100 8 bit depth ultrasound images of 576×688 pixels. The performance of our system are evaluated in terms of: capacity, watermark imperceptibility and robustness of the watermark M_t^k.

Capacity: As one bit of the message M_f is inserted per pixel block, the number of bits one can insert into an image depends on the size of the pixel blocks N and of the image. In this experiment, considering N=8 leads to a capacity of 1/8 bit per pixel or equivalently to a message of about 49536 bits. This capacity is sufficient enough for the insertion of some security attributes that will assess the image reliability. For instance, M_f may contain an authenticity code, which identifies the image origin (e.g. about 1000 bits by combining the French National Identifier with the DICOM Unique Identifier [13]), and an integrity proof like the digital signature of the image bit subset un-modified by the watermarking process [14].

Distortion: As our algorithm introduces on the average the same image distortion in each block of pixels, we decided to use the Peak Signal to Noise Ratio (PSNR) to measure the distortion between the image (I) and its watermarked and deciphered version (I_d^k) [8]. Considering Δ=8, the PSNR is greater than 44dB, an acceptable distortion in the case of ultrasound images (see [14] for more details). Notice that the image quality can be better preserved by only watermarking some of image blocks.

Robustness of the watermark M_t^k: Unlike the message M_f that can be fragile so as to assess image reliability, the watermark M_t^k identifies the receiver k and should be robust to malicious attempt to suppress it. Due to space limitations, we only evaluate the robustness of M_t^k against the Additive White Gaussian Noise attack (AWGN). In this experiment, we considered a scenario with K_u=200 receivers and where the 100th receiver rerouted its decrypted image I_d^{100} trying to remove M_t^{100} that identifies him with the AWGN attack. Figure 4 gives the detection results scheme obtained in the case of two attacks of standard deviation $\sigma = 1$ and $\sigma = 10$. It represents the correlation values in-between the extracted watermark and the watermarks of each of the 200 receivers, i.e. M_t^k, k=1...200 (see section III. C). As it can be seen, the maximum correlation value is obtained for the 100th receiver in conformity with our scenario. Notice that the distortion between the image I_d^{100} and its attacked version, measured in terms of PSNR is respectively of 48.14 dB and 28.12 dB for $\sigma = 1$ and $\sigma = 10$. In the latter case the image is strongly degraded but we can still identify the dishonest user.

V. Conclusion

In this paper, we have proposed a new crypto-watermarking system, which couples QIM based watermarking and the Joint Watermarking Decryption approach proposed in [7]. Its objective is to guarantee at the same time the confidentiality of the image by means of encryption and also to ensure the image reliability as well as its traceability allowing the identification of the physician at the origin of an illegal distribution of medical images. In order to achieve this goal, we have shown the constraints to satisfy in making interoperate watermarking and JWD and minimizing interferences between their respective watermarks. Experimental results show that the image distortion is acceptable for ultrasound images and that the JWD watermark used for user tracing is robust against AWGN attack. Future works will focus on enhancing the quality of the watermarked images.

References

[1] C. Quantin, D.O. Jaquet-Chiffelle, G. Coatrieux, E. Benzenine, F.A. Allaërt, "Medical record search engines, using pseudonymised patient identity: An alternative to centralised medical records", *International Journal of Medical Informatics*, vol. 80, n°2, pp.6-11, 2011.

[2] L. Kun, R. Beuscart, G. Coatrieux, C. Quantin, "Improving outcomes with interoperable EHRs and secure global health information infrastructure," *Engineering in Medicine and Biology Society, 29th Int. IEEE EMBS Conf.* pp.6158-6159, *2007*.

[3] G. Coatrieux, H. Maître, B. Sankur, Y. Rolland, R. Collorec, "Relevance of watermarking in medical imaging," *in proc. of Int. Conf. on IEEE EMBS ITAB*, USA, pp. 250-255, 2000.

[4] G. Coatrieux, H. Huang, H. Shu, L. Luo, Ch. Roux, "A Watermarking-Based Medical Image Integrity Control System and an Image Moment Signature for Tampering Characterization," *IEEE* Journal of *Biomed. and Health Inf.*, vol.17, no.6, pp.1057-1067, 2013..

[5] H. Seungwoo, K. Hakjae, L. Sungju, C. Yongwha, "Analyzing the Secure and Energy Efficient Transmissions of Compressed Fingerprint Images using Encryption and Watermarking," *Int. Conf. on Inf. Security and Assurance,* pp.316-320, 2008.

[6] N.D. Memon and P.W. Wong, "A buyer-seller watermarking protocol," *IEEE Trans. on Image Processing*, vol. 10, n° 4, pp.643-649, April 2001.

[7] M. Celik, A.N. Lemma, S. Katzenbeisser, M. van der Veen, "Secure Embedding of Spread Spectrum Watermarks Using Look-up-Tables," *ICASSP '07,* Hawaii, USA, vol. 2, pp. 153-156, 2007.

[8] D. Bouslimi, G. Coatrieux, M. Cozic, C. Roux, A joint encryption/watermarking system for verifying the reliability of medical images, *IEEE TITB* 16 (2012) 891–899.

[9] D. Bouslimi, G. Coatrieux, Ch. Roux, "A joint encryption/watermarking algorithm for verifying the reliability of medical images: Application to echographic images", *Computer Methods and Programs in Biomedicine*, vol. 106, pp. 47-54, 2011.

[10] B. Chen, G.W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital watermarking and information embedding," *IEEE Trans. on Inf. Theory*, vol. 47, n°. 4, pp. 1423-1443, 2001.

[11] R. J. Anderson and C. Manifavas, "Chameleon - a new kind of stream cipher," *in FSE '97*, London, UK: Springer-Verlag, pp. 107–113

[12] A. Piva, T. Bianchi, and A. De Rosa, "Secure client-side ST-DM watermark embedding," *IEEE Trans. Inform. Forensics Sec.*, vol. 5, no. 1, pp. 13–26, Mar. 2010.

[13] W. Pan, G. Coatrieux, N. Cuppens-Boulahia, F. Cuppens, Ch. Roux, "Medical Image Integrity Control Combining Digital Signature and Lossless Watermarking," *Lecture Notes in Computer Science*, vol. 5939, pp. 153-162, 2010.

[14] K. Chen, T.V. Ramabadran, "Near-lossless compression of medical images through entropy coded DPCM", *IEEE Transactions on Medical Imaging*, 1994.