Seamless Integration of Watermarks in DICOM Images

Óscar J Rubio, Álvaro Alesanco, José García

Communication Technologies Group (GTC), Aragón Institute of Engineering Research (I3A), University of Zaragoza, Spain

Abstract

This paper proposes a watermarking method that can be applied to DICOM images to strengthen its security and privacy. A robust watermark is used to improve the identification/authentication of the image, even if it has been seriously modified, semifragile watermarks are used to embed information intended to vanish if the image losses its clinical/research value, and fragile watermarks are used to detect and locate tampering. The watermarks embed selected DICOM fields, replacing the original values with the corresponding access keys, conveniently protected with Cryptographic Message Syntax. This watermarking does not distort the image, presents a good robustness-capacity tradeoff, operates fast and is highly compatible with the JPEG2000 compressor.

1. Introduction

The Digital Imaging and COmmunication in Medicine (DICOM [1]) standard is the most popular system for the storage, transmission, handling and impression of medical images regardless its origin. The vast majority of medical imaging equipment, PACS and RIS support its robust file format and network protocol. This standard implements a highly secure role-based access control to fulfill legal regulations, like the HIPAA [2] in United States. The security design is entirely based on cryptography, the most sensitive DICOM contents (fields and/or the image if it identifies the patient) are put into digital envelopes and sealed by means of Cryptographic Message Syntax (CMS [3]). Nonetheless, the answer of the system when the image is corrupted could be improved. With current security measures, a corrupted DICOM file can still be queried and the rest of its contents can be retrieved unless the corresponding digital envelopes are also tampered.

Watermarking [4] could be used to strengthen DICOM security in the aforementioned cases. These techniques permit embedding certain contents into the image by means of watermarks, which can be retrieved by means of access keys. An interesting feature is that the strength of the link between the image and the watermarks can be adjusted. Robust watermarks are intended to endure heavy image distortion, semifragile watermarks only resist mild image modifications (e.g. if they preserve its clinical/research value) and fragile watermarks are intended get totally distorted if the image suffers any small modification. As depicted in Fig. 1, in the medical context [5, 6] robust watermarks may be of use for the persistent identification and authentication of the image (by embedding the patient name -tag (0010,0010)-), semifragile watermarks to embed information regarding the image diagnosis, and fragile watermarks can implement integrity control and tamper location in the image. Since the fragile watermark must be a known reference element, this can be the patient name again, already link to the image by means of a robust watermark.

The main requirement to ensure the integration of watermarks within DICOM is that the embedding algorithm must be distortion-free. Maintaining the quality of these images is crucial since it may have already been reduced to a tolerable limit by some previous processing (e.g. compression). To meet this requirement it is not valid that the watermarks are embedded in regions-of-non-interest (RONI) of the image (e.g. black spaces), since they are easy to find by non-authorised users and they may be removed by specific medical image compressors. Also not acceptable using reversible watermarking techniques, since they need to remove all the watermarks to retrieve the original image, which implies unprotecting the image. Besides, those users without access to all the watermarks will work with a lower-quality version of the image. Second, robust and semifragile watermark must endure modifications of the image in the clinical context, namely brightness and contrast change, enhancement filtering, quality-preserving compression, cropping and insertion of visible watermarks. Robust watermarks must also resist modifications that downgrade the image quality. Third, the access key that links the image and the watermark must be signed with CMS (and encrypted if its content is confidential), to keep compliance with DICOM security. Finally, it is desirable that the watermark embedding-retrieval algorithms are simple, compatible with the JPEG2000 [7] compressor (supported by DICOM) and able to operate in real-time.

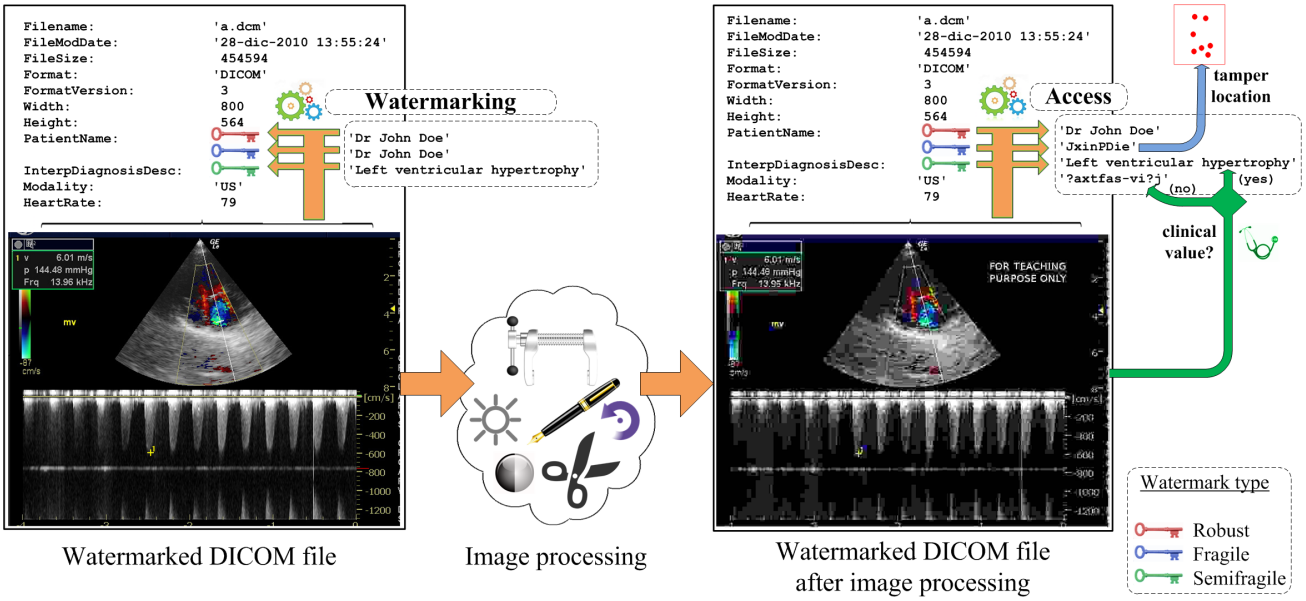

Figure 1. Proposed watermark embedding and access processes for DICOM files.

2. Distortion-free watermarking algorithm

The algorithm developed for the *embedding of watermarks* without distorting the image Im is described below:

1. The ROI of the image is segmented, by manual selection of an expert or by means of a specific software, and clipped from the background.
2. Color ROIs are transformed into grayscale by selection of its luma component, $Y = 16 + 65.481 \cdot R + 128.553 \cdot G + 24.966 \cdot B$. This ensures compliance with both color and grayscale images.
3. The maximum decomposition level of the CDF 9/7 [8] *wavelet* transform of the segmented grayscale ROI is calculated. The highest levels concentrate more energy and the lowest frequencies, being more resilient to most image modifications. Nonetheless, each level has only a 25% of the coefficients of the previous, that is to say, capacity to embed only shorter watermarks.
4. For robust watermarks, select those coefficients C belonging to *wavelet* decomposition levels $WL \geq 3$. For semifragile watermarks, those belonging to decomposition levels $1 - 2$. Fragile watermarks will be embedded in the HH subband of the first decomposition level. This multiplexing in the transform domain permits embedding several watermarks in the same image.
5. Select the $2.2 \cdot length(W)$ highest-magnitude coefficients in C, to obtain higher robustness. The average number of coefficients needed for the encoding of one W bit is $\sum_{k=1}^{\infty}(0.5)^k \cdot k = 2$, since the mean bit values of W and C will be forced to be 0.5.
6. Keep only the sign bit of the coefficients selected from C if the watermark is robust/semifragile, since it is the most robust against most image modifications. Keep the least significant bit if the watermark is fragile, for the opposite reason.
7. Reverse the sign of coefficients in pair columns of C, producing C^*, and the value of pair elements in W, producing W^*. The mean bit value of C^* and W^* is approximately 0.5, and the series of 0s and 1s are broken. This is consistent with step 5.
8. Encode the W^* bits as sign bits C_i^* in C^*. The resulting vector of positions is denoted as K. Each K_i element describes the distance between C_{i-1}^*, whose bit sign corresponds to W_{i-1}^*, and C_i^*, the first element (moving forward along C^*) whose sign bit matches the bit W_i^*. Moving backward or repeating positions in the encoding is prevented.
9. For a compact encoding of the key, $K := K - min(K)$ and each element is encoded with $\#bits = log_2(max(K))$. These two values, $min(K)$ and $\#bits$ are attached with K to enable the decoding.

The algorithm for the *retrieval of the embedded watermarks* begins by following steps 1-7. Next, the actual values of the key are calculated, $K := K + min(K)$. Then, each $\tilde{W}_i^*$ bit is retrieved by moving forward K_i^* places from the position on C_{i-1}^* and reading the sign bit of the corresponding coefficient. Finally $\tilde{W}$ is obtained by reversing the pair elements of $\tilde{W}^*$ and maintaining the rest.

Resynchronization step: the image may be subject to geometrical transformations, which in the image context are 90/180/270° rotation, vertical and horizontal flipping. A reference 64-bit robust watermark is embedded, and re-

trieved from the received position and each of the geometrical transformations (rotating/mirroring each subband in C^* and maintaining K). The retrieved watermark being more similar to the original reference corresponds to the transformation that returns the image to its original position. If it is observed that a retrieved watermark is more dissimilar (e.g. 63 wrong bits) than the most similar watermark (e.g. 40 correct bits), it means that the original position corresponds to the former, whose colors have been inverted. In that case, the values of C^* need to be inverted.

Protection of the watermarks

To prevent eavesdropping, the watermark embedding process ensures that no image coefficient encodes two different watermark bits (see step 8). Otherwise certain watermark bits could be derived without the image. Besides, the collusion and forgery attacks, that combine different watermarked images to produce faked watermarked versions, are also prevented. In this scheme the watermarked images are all exactly like the original, and the key/s to access the watermark/s content are suitably protected.

Every time a DICOM field is embedded into the medical image, the corresponding access key replaces the original value. Since the embedded fields often contain important information, they are usually contained in digital envelopes protected with CMS. If not, it is required to place the access keys into a protected envelope. This envelope must be digitally signed (with ECDSA if possible), and ciphered if the content of the key/s is confidential, preferably using AES for the symmetric ciphering and RSA 2048/4096 for the asymmetric.

3. Evaluation and conclusions

For the evaluation of this watermarking method, a variety of 40 different DICOM echocardiograms corresponding to modes B, M, color Doppler and continuous of 20 patients from Lozano Blesa Hospital (in Zaragoza) were used. These kind of tests were chosen instead of others (like magnetic resonances or CTs) because they contain very little energy, which makes watermarking more challenging. Random watermarks of different type and length were embedded in these echocardiograms. Different modifications were then performed on them, mild and aggressive compression, common image processing (e..g β correction, contrast stretching, inverting colors), local operations that remove or add details to the image, geometrical changes that can be inverted, clipping the ROI, inserting annotations and darkening private data. Finally the watermarks $\tilde{W}$ were retrieved from the modified echocardiogram and compared with the originals W, by means of the Normalized Hamming Distance, $NHD = \dfrac{\tilde{W} \oplus W}{length(W)}$. Where W and $\tilde{W}$ are binary vectors and $\oplus$ is the XOR logical operator.

In the second column of Tab. 1 it is measured the average distortion that each image modification $\tilde{ROI}$ produces in the ROI of the echocardiogram, by means of the Peak Signal-to-Noise Ratio, $PSNR = 20 \cdot log_{10}(\dfrac{MAX\{ROI\}}{\sqrt{\frac{1}{M \cdot N} \cdot \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} ||ROI(i,j) - \tilde{ROI}(i,j)||^2}})$. Among the non-invertible modifications, the local operations (except the edges sharpening) cause the highest distortion, making the echocardiogram lose its clinical/research value, since high and middle frequencies (the details) are removed. This is the reason why the watermarking always obtains worse results when tested against these modifications. On the other hand, darkening private data, clipping the ROI or adding visible watermarks cause little or no distortion since they usually modify only the RONI of the image, dismissed for the watermarking.

The results in Tab. 1 show the tradeoff between robustness (low NHD) and capacity (length of the watermark). *Robust watermarks* achieve low NHD values ($\leq$ 2.9%) for capacities up to 1024 bits. Moreover, the NHD can be reduced to 0 by encoding the watermarks with BCH(511,259,30) coding [9]. In authentication, the probability of false positives is given by $P_{fp} = \sum_{i=Th}^{n=length(W)} (0.5)^n \cdot \dfrac{n!}{i!(n-i)!}$. Therefore, it is enough to use a short 128-bit watermark with detection $Th = 126$ bits to obtain $P_{fp} = 2.5 \cdot 10^{-35}$, which ensures correct authentication. As intended, *semifragile watermarks* present good robustness-capacity ($NHD \leq 0.7\%$ up to 4096 bits) except for those attacks that make the echocardiograms lose detail: aggressive compression (e.g. JPEG2000 with $CR > 32$) and those local operations that remove middle and high frequencies. Finally, fragile watermarks ensure very little robustness against the modifications, even when using short 64-bit watermarks. The exceptions are the invertible modifications and those that not affect the ROI, since they do not change the clinical value of the echocardiogram. It is recommended to use fragile watermarks longer than 512 bits to obtain $NHD > 1\%$ when the modification is the addition of annotations. To locate tampered points in the echocardiogram, it is only necessary to pinpoint the wrongly detected watermark bits in their corresponding positions in C. The longer the fragile watermark, the higher the resolution in tamper location.

The overall embedding-retrieval delays, including protection and access to the key, are $< (160, 130)\ ms$. If the watermarking is integrated within the JPEG2000 compressor (results in Tab. 1 show high compatibility), the wavelet transformation is already performed, saving $\simeq$ 100 ms. This shows that the system is able to operate in real-time. Finally, the size of the access keys is

Table 1. Average endurance of variable-length robust, semifragile and fragile watermarks against common modifications in medical imaging.

	Image distortion -$PSNR$ (dB)-	Distortion of robust watermarks -NHD (%)-					Distortion of semifragile watermarks -NHD (%)-							Distortion of fragile watermarks -NHD (%)-		
#. Operation \ $Length(W)$	-	128/BCH	256/BCH	512/BCH	1024/BCH	2048	128	256	512	1024	2048	4096	8192	64	128	256
Compression																
1. JPEG QF=75%	38.1	0	0	0	0	0	0	0	0	0	0.1	0.4	2.2	50	50	50.8
2. JPEG QF=50%	33.2	0	0	0	0	0	0	0	0	0.3	1.2	2.9	7.6	48.4	48.8	50
3. JPEG QF=25%	29.9	0	0	0	0	0	0.4	1.2	1.1	2.4	3.7	7.5	13.2	46.9	50	49.8
4. JPEG QF=15%	27.9	0	0	0	0	0.5	1.6	2	2.9	4.5	7.1	12.5	18.7	48.4	49.6	48.8
5. JPEG2000 CR 4:1	55.1	0	0	0	0	0	0	0	0	0	0	0	0	53.1	51.6	50
6. JPEG2000 CR 8:1	53.1	0	0	0	0	0	0	0	0	0	0	0	0	50	50	49.6
7. JPEG2000 CR 16:1	45.7	0	0	0	0	0	0	0	0	0	0	0	0	48.4	51.6	50.4
8. JPEG2000 CR 32:1	38.1	0	0	0	0	0	0	0	0	0	0	0	1.4	50	49.6	50.6
9. JPEG2000 CR 64:1	32.1	0	0	0	0	0.8	0	0	0	0.8	3.3	9.5	18	50	48.4	50.2
Common image processing																
10. β correction -0.3	24.2	0	0	0	0.2/0	1	0	0	0	0	0	0	0.4	48.4	50	50
11. β correction +0.4	24.8	0	0	0	0	0	0	0	0	0	0	0	0	51.6	50	50.4
12. β correction +0.7	18.5	0	0	0.2/0	0.4/0	1.2	0	0	0	0	0.2	0.7	1.2	50	52	50.8
13. Contrast stretching 2%	26.6	0	0	0	0	0.1	0	0	0	0	0	0	0.1	51.6	52.3	49.6
14. Contrast stretching 10%	23.3	0	0	0	0.1/0	0.4	0	0	0	0	0	0.1	0.4	47.7	49.2	50
15. Invert colors	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16. Local hist. equal.	25.4	0	0	0	0	0.2	0	0	0	0	0	0	0.1	51.6	48.4	49.2
Local operation																
17. Edges sharpening	22.9	0	0	0	0.3/0	0.7	0	0	0	0	0.1	0.4	1	47.7	47.7	49
18. Median filter 5x5	19.0	0.4/0	0.8/0	1.6/0	2.9/0	4.8	42.6	40.4	43.1	40.1	37.5	37.7	38.2	49.2	50	49.6
19. Image averaging 5x5	20.8	0	0	0	0.3/0	0.9	57	59.4	59.4	58.9	57.7	54.9	54.2	51.6	50	50.4
20. Gaussian filtering 7x7	21.0	0	0	0.1/0	0.8/0	1.7	28.9	29.1	26.7	30.1	31.1	31.7	34.9	48.4	50	48.2
21. Gaussian filtering 11x11	20.8	0	0	0.3/0	1.2/0	2.6	23.4	25	22.7	23.5	25.5	27.4	31.9	49.2	49.6	49.2
22. Motion blur 7	21.7	0	0	0.2/0	0.9/0	1.6	32	29.7	28.4	28.9	30	30	29.3	49.2	50.8	50.6
Geometrical changes																
23. Clipping the ROI	Inf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24. Rotating 90°	12.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25. Rotating 180°	12.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26. Rotating 270°	12.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27. Horizontal flipping	13.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28. Vertical flipping	12.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29. Inserting annotations	28.7	0	0	0	0	0.4	0	0	0.1	0.2	0.4	0.7	0.9	0.2	0.4	0.6
30. Darken private data	Inf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$\simeq 7 \cdot length(watermark/s)$, which can be considered a moderate *overhead*.

The results demonstrate that the proposed watermarking method meets the requirements (discussed in Sec. 1) to be seamlessly integrated within DICOM.

Acknowledgements

This research work has been partially supported by project TIN-2011-23792/TSI from the Ministerio de Economía y Competitividad (MINECO), the European Regional Development Fund (ERDF) and the European Social Fund (ESF).

References

[1] Digital Imaging and Communications in Medicine (DICOM), National Electrical Manufacturers Association (NEMA). Accessed in March 2013, `http://bit.ly/ISaJRk`.

[2] The Health Insurance Portability and Accountability Act (P.L.104-191), 1996. Enacted by the U.S. Congress.

[3] R. Housley. Cryptographic Message Syntax (CMS), RFC 5652, IETF Network Working Group. Accessed in March 2013, `http://tools.ietf.org/html/rfc5652`, September 2009.

[4] Cox I, Kilian J, Leighton F, Shamoon T. Secure spread spectrum watermarking for multimedia. Image Processing IEEE Transactions on Dec. 1997;6(12):1673–1687.

[5] Coatrieux G, Maitre H, Sankur B, Rolland Y, Collorec R. Relevance of watermarking in medical imaging. In Information Technology Applications in Biomedicine, 2000. Proceedings. 2000 IEEE EMBS International Conference on. 2000; 250–255.

[6] Giakoumaki A, Pavlopoulos S, Koutsouris D. Secure and efficient health data management through multiple watermarking on medical images. Medical and Biological Engineering and Computing 2006;44(8):619–631.

[7] Taubman Author DS, Marcellin Editor MW, Rabbani Reviewer M. JPEG2000: Image Compression Fundamentals, Standards and Practice. Journal of Electronic Imaging 2002; 11(2):286–287.

[8] Cohen A, Daubechies I, Feauveau J. Biorthogonal bases of compactly supported wavelets. Comm Pure and Applied Math 1992;45:485–560.

[9] Bose R, Ray-Chaudhuri D. On a class of error correcting binary group codes. Information and Control 1960;3(1):68–79.