Tracing Medical Images Using Multi-Band Watermarks

Mingyan Li, Sreeram Narayanan and Radha Poovendran
Department of Electrical Engineering, University of Washington
email: radha@ee.washington.edu

Abstract— **The enforcement of Health Insurance Portability and Accountability Act (HIPAA) in April 2003 highlights the determination of federal government to protect the privacy of patients' records. However, we identify that there is a gap between current security solutions for the privacy protection of medical images and the HIPAA security guidelines: there is no measure to prevent and trace the authorized users who distribute medical images illegally. We evaluate the suitability of some most widely used watermarking techniques for tracing medical images, and demonstrate a need of developing new watermarking schemes that withstand standard medical image processing while maintaining high image quality for diagnostic purpose. We propose a multi-band watermarking scheme, that is robust to low pass filtering and high pass filtering, and yields high perceptual quality. We also present open research problems.**

Keywords— **medical image security, watermarking, privacy, HIPAA**

I. Introduction

Privacy protection of medical images has always been an important issue in the management of patients' medical records. As part of Health Insurance Portability and Accountability Act (HIPAA), the standards to protect health data privacy issued by the federal government took effect on April 14th 2003 [1]. The HIPAA mandates hospitals, doctors, and other health providers to ensure "confidentiality and integrity of individually identifiable health information, past, present or future."

As digital technology pervades our society, a vast amount of medical images now exists in electronic format for easy storage and maintenance. Ubiquitous wired and wireless networks make it possible to access and share data among professional personnel from anywhere, to promote high quality care for patients. The convenience of data access and distribution also poses a great challenge on privacy protection for patients' information. Constant efforts have been made to provide security solutions [2], [3], [4], [5] to ensure (i) medical image transmission cannot be accessed by unauthorized parties (confidentiality), (ii) received images are not modified during transmission (integrity), and (iii) images are from correct sources to the claimed receivers (authentication). Continuously updated Digital Imaging and COMmunication in medicine (DICOM) standards provide guidelines to ensure authentication, integrity and confidentiality of medical images [2].

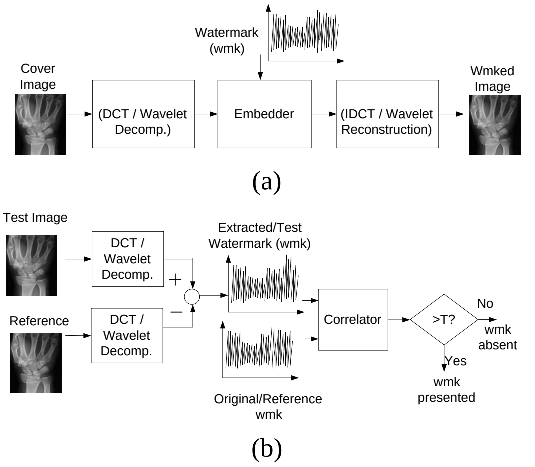
However, security measures in DICOM and the research on medical image security [3], [4], [5] do not guarantee privacy of patient data at the recipient end. At present, even if the original data is privacy preserving, the recipient can violate it by giving the images away to unentitled parties. Thus the current privacy mechanisms are not adequate to meet all the HIPAA requirements [1]. In this paper, we make the first attempt to study the problem of tracing illegal distribution of medical images. We use watermarking techniques to trace the authorized person who first distributed the image, on retrieval of images from unauthorized parties. A weaker version of the tracing problem is identifying the ownership of the image itself.

Spread Spectrum (SS) scheme [6], Image-Adaptive DCT (IA-DCT) scheme [7], IA Wavelet (IA-W) scheme [7] are among the most widely used watermarking schemes. These techniques survive processing common to multimedia. However, we find that they are unsuitable for medical image, where processing is performed in order to extract information specific to the physician. These processing may be in the form of frequency selective operations like, High Pass Filtering (HPF) and Low Pass Filtering (LPF). Therefore, *new watermarking schemes that can withstand these processes and preserve high perceptual quality for diagnostics are required.*

In this paper, we propose a novel Multi-Band Embedding (MBE) scheme. The algorithm embeds watermarks at different frequency bands in the original image, such that if the processing removes some components, the rest is detected. The MBE survives HPF and LPF, while preserving high perceptual quality. The rest of this paper is organized as follows: In the next section, we explain how watermarking is performed and detected, and discuss a few common watermarking techniques; Suitability of watermarking to medical images will be explored in Section III and a new watermarking scheme will be proposed; Section IV gives the results of various attacks and their effects on the watermarking schemes; We conclude with possible future directions in Section V.

II. Watermarking

Watermarking is a technique to embed identification codes, called watermarks, into *cover media* or *host media*. Watermarking is used for the protection of intellectual property, data integrity, and data authentication [6]. Under the assumption of a unique watermark per image per user, watermarking can be used as fingerprinting. A watermarking system consists of two components: a watermark embedder, and a watermark detector. Watermark embedding is performed in either spatial domain or transform domain. Embedding watermarks in transform domain is shown to be more robust than that in spatial domain [6]. Fig. 1 illustrates a transform domain watermarking system, where a watermark is inserted into the transformed image at positions determined by the embedder. At the detection module, original images as well as the watermarks are assumed to be available. To examine the presence of the watermark the detector extracts a watermark and calculates the



(a)

(b)

Fig. 1. A transform domain based watermarking system, (a) a watermark embedder, (b) a watermark detector.

correlation between the reference and the test watermark. If the correlation is above a prespecified threshold T, the watermark is determined to be valid.

Cox *et al.* provided a comprehensive study on watermarking techniques [8]. In this paper, we focus on the most widely used watermarking schemes due to their high robustness.

A. Watermarking techniques studied

Spread spectrum watermarking: Secure *Spread Spectrum* (SS) watermarking proposed in [6] is a seminal work and still remains the most widely used watermarking scheme [9]. Cox *et al.* [6] viewed the channel between a watermark embedder and a detector as a communication channel, and extended the idea of spread spectrum communication to watermarking: a watermark, as a narrow banded signal, is transmitted in a wide band signal (cover image) such that the signal energy presented in any single frequency is imperceptible. They discovered that watermarks should be embedded into perceptually dominant spectral components of the cover, making them imperceptible and robust to distortions. Chosen watermarks are sequences of independent zero mean unit variance Gaussian random variables, and thus orthogonal to each other. Watermarks of length n are embedded to the n largest (in magnitude) discrete cosine transformation (DCT) coefficients excluding the DC component. Let x_i denote the i^{th} component of a watermark, $i = 1, ..., n$, v_i and v'_i denote the coefficients of the i^{th} frequency of the original and the watermarked image, respectively. v'_i can be generated from v_i and x_i using *Additive Embedding* (AE) or *Multiplicative Embedding* (ME) as shown below:

$$\begin{aligned} v'_i &= v_i + \alpha x_i \qquad x_i \sim N(0,1) \qquad i = 1, ..., n, && (1) \\ v'_i &= v_i(1 + \alpha x_i), \qquad i = 1, ..., n && (2) \end{aligned}$$

where α is a scaling factor. The smaller the α, less the distortion between the original and the watermarked image. AE (1) is image independent, and hence suitable for the cases where watermarks need to be forcibly embedded and the magnitudes of v_is are very small. The ME (2) better exploits the embedding capacity of frequency components and is suitable when the magnitudes of v_is vary widely. The SS technique employs multiplicative embedding [6].

Image Adaptive schemes: Podilchuk *et al.* introduced visual models [7] to improve the robustness of SS without introducing perceptual degradation. In the *Image Adaptive* (IA) schemes [7], the strength and embedding positions of watermarks are adaptively chosen based on the perceptual sensitivity of the cover image in terms of *Just Noticeable Differences* (JND), which is image dependent.[1] There are two IA schemes: IA-DCT based on 8×8 block DCT transformation, and IA-W based on wavelet decomposition. Let J_i be the calculated JND values, the embedding in IA schemes is described as:

$$v'_i = \begin{cases} v_i + J_i \cdot x_i & \text{if } v_i > J_i \\ v_i, & \text{otherwise.} \end{cases} \qquad (3)$$

Note that instead of fixed length watermarks used in SS, the IA schemes explore the potential embedding capacity of every frequency component in the host image and thus in general, have a longer watermark sequence than SS.

B. Performance metrics

A watermarking scheme is evaluated based on two critical yet conflicting performance metrics: (i) imperceptibility (ii) robustness to distortion and attacks to eliminate watermarks.

Imperceptibility measures: Peak Signal to Noise Ratio (PSNR) is a widely used measure of fidelity (similarity between the original and the distorted image). PSNR is defined as:

$$\text{PSNR} = 10 \log \left(\sum_i \frac{255^2}{(y_i - y'_i)^2} \right), \qquad (4)$$

where y_i and y'_i are the values of the i^{th} pixel in original image and watermarked image, respectively.

Subjective inspection of an image by medical experts is another approach but the subjective evaluation varies from one expert to another. We will use PSNR as a performance metric in this paper.

Robustness measure: Correlation measures the statistical similarity of sequences. To evaluate robustness, the watermark is extracted from a test image that underwent modifications, and correlation between the test watermark and the reference watermark is calculated. Normalized correlation, *sim* in [6], is used in this paper and is defined as:

$$sim(X, X^*) = \frac{X \cdot X^*}{\sqrt{X^* \cdot X^*}}, \qquad (5)$$

where X and X^* are the original and reconstructed watermark sequence, respectively.

III. Watermarking for medical image

A. Design consideration

In medical imaging, distortion introduced by watermarks or compression should be kept to a minimum so that diagnostic value is not compromised. The high fidelity requirement

[1] JND is the concept in visual modeling defining a level of distortion that is perceptual in 50% of experimental trials [8] and is used to derive perceptually based quantizers in compression applications [7].

renders watermarking in medical images a harder task than multimedia applications where a much higher perceptual distortion is tolerated. Based on studies involving suitability of lossy compression for medical images, a minimum PSNR of 40-50db is considered acceptable [10], while 20-30db PSNR is tolerable for multimedia applications. The stringent requirement of high PSNR limits the resiliency of watermarks to various attacks. On the other hand, high PSNR requirement also limits the types of attacks that can be utilized in order to erase watermarks of medical images.

More importantly, watermarks should survive the standard image processing like LPF and HPF that improve diagnostic quality of medical images. Unlike multimedia, HPF which functions as an edge detector enhances the information content. These processes introduce distortion on images, which can destroy watermarks, and thus are interpreted as attacks on watermark. The watermarks must also survive other common image processing such as: JPEG compression, cropping, and intentional attacks, such as averaging attack, where several watermarked copies are averaged in order to remove or reduce the strength of a watermark. Two critical design requirements for medical watermarking design are:

- **Req1:** high image fidelity to meet diagnostic requirements
- **Req2:** robustness to filtering operations while withstanding conventional attacks.

B. *Fidelity Requirement of SS, IA-DCT and IA-W*

To make SS meet Req1, we choose a small value of α in (2), such that it maintains high fidelity at the cost of robustness. The possibility of a weaker watermark for high fidelity being resistant to distortion is a topic for our future study. The original IA-DCT and IA-W embedding algorithms (3) do not provide a parameter to control the strength of the watermarks. To fulfill Req1, we employ the modified IA schemes [8] as:

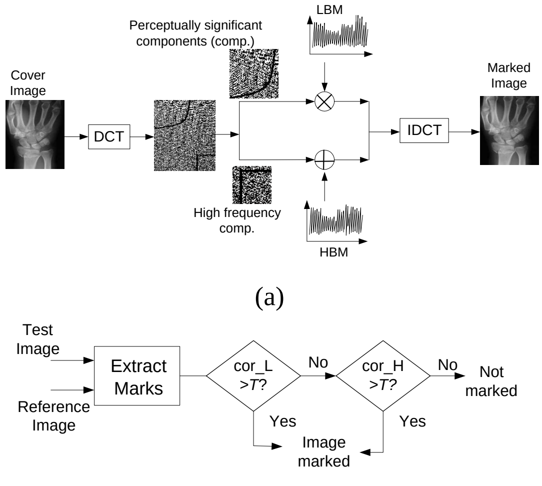
$$v_i' = \begin{cases} v_i + \alpha \cdot J_i \cdot x_i & \text{if } v_i > J_i \\ v_i, & \text{otherwise.} \end{cases} \qquad (6)$$

However, as will be shown in Section IV, SS and IA schemes cannot withstand Req2. The failure demonstrates a need for new development in watermarking for medical images. We present a novel technique below.

C. *New Multi-Band Embedding (MBE) scheme*

We propose a new watermarking scheme, called Multi-Band Embedding (MBE), that is robust to frequency selective filtering. As the name suggests, we embed the watermark in chosen frequency bands. While our scheme is generalizable, we embed only on two frequency bands in this paper for illustration.

MBE is built on SS with an additional mark to survive HPF. Fig. 2(a) illustrates the MBE embedder. The embedding algorithm consists of two parts: multiplicative embedding of a low band mark (LBM) in the frequencies corresponding to highest magnitudes except DC, and additive embedding of a high band mark (HBM) in the high frequencies. While embedding in the high frequencies the watermark is inserted in the 32×32 block corresponding to highest frequencies. Since high frequency coefficients have very low magnitude, additive embedding is chosen so that the watermarks are not scaled to very low values. The α in additive embedding is set to be the average of all the 32×32 high frequency coefficients, in order to keep magnitude of the embedded watermark within the range of the coefficients.



(a)

(b)

Fig. 2. A Multi-Band Embedding (MBE) watermarking system (a) Embedding algorithm: a low band mark (LBM) is multiplicatively embedded into perceptually significant components and a high band mark (HBM) is additively embedded to high frequencies (the right lower 32×32 block of the transformed image), (b) Detection algorithm: the watermark is detected when either the correlation between LBM and its reference (cor-L) or that between HBM and its reference (cor-H) is above the prespecified threshold T.

Fig. 2(b) presents watermark detection in MBE. To confirm whether a watermark, as a combination of LBM and HBM, is present in a test image, the detector calculates correlation between the extracted LBM and the reference LBM, and that between the extracted and reference HBM. If one of the correlations is above a prespecified threshold, the watermark is assumed to be present.

Compared to the standard SS [6], the additional mark in the MBE scheme will lower the PSNR after watermark embedding. However, due to their low magnitude, the mark in high frequencies has a much lower impact on PSNR than the mark in perceptually significant frequencies. Indeed, our simulation study shows that MBE and SS have similar PSNR at a given scaling factor α.

IV. Simulation and Discussion

We evaluate the performance of four candidate watermarking schemes: SS, IA-DCT, IA-W and MBE on medical images in terms of PSNR and robustness to various image distortions. We use 23 medical images, 20 640x480 images and three 490x490 images. We perform the simulation using Matlab. Fig. 3 shows the effect of scaling factor α on the PSNR of the four schemes. α is varied from 0.01 and 1. It is evident from Fig. 3, that PSNR of MBE is almost identical to that of SS, except a negligible 0.5 db difference when $\alpha = 0.01$.

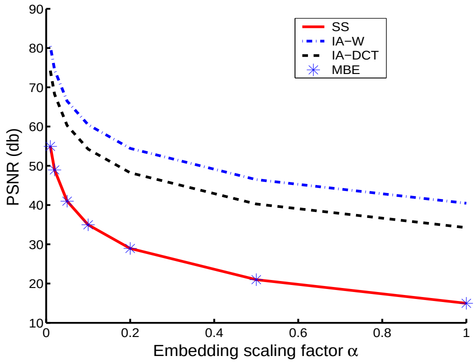


Fig. 3. PSNR vs. embedding scaling factor α when $0.01 \leq \alpha \leq 1$, as the average of 23 images. Note that PSNR of MBE and SS are almost the same.

For robustness test under all the schemes, we choose the value of α that guarantees 40 db PSNR to meet high fidelity requirement. The threshold for sim is set at $T = 5$, which is well above the similarity between two independent gaussian random sequences [6]. In this simulation, we consider the following processing and attacks, and their combinations. (i) LPF: image is operated upon by a 3×3 smoothing window. LPF is performed for image smoothing and noise reduction. (ii) HPF: image is operated by a standard 3×3 high pass kernel. HPF is performed for edge detection. (iii) Averaging: an average image corresponding to m colluders is generated. For illustration, we use $m = 10$. To detect colluders, the extracted watermark is tested against all the candidate watermarks. (iv) JPEG compression: compressed images are written in JPEG format using `imwrite` function in Matlab, with quality factor being 75. (v) Cropping: the central quarter of the watermarked image is cropped to mimic cropping attack. For the purpose of watermark detection under cropping attack, the missing part is replaced by the original cover image. Table I presents the results of robustness test.

1) Performance of SS, IA-DCT, and IA-W: From Column 3 in Table I, it is obvious watermarking using SS, IA-DCT and IA-W do not survive the HPF. In the SS scheme, a watermarks is inserted in the 1000 highest magnitude frequency coefficients. Usually, most of the energy in an image is concentrated at low frequencies, and thus the low frequency components are of high magnitude. Hence, in SS scheme, a watermark is embedded mostly in the low frequencies. Thus the watermark is not recovered after a HPF operation.

TABLE I

Robustness Comparison of Schemes under Various Distortions

The results are based on 23 medical images. For columns 2, 4, and 7, where the distortion involves an averaging attack of 10 colluders, the average number of detected colluders is given. No false negative in colluder detection is found. For columns 1, 3, 5 , and 6, the percentage of survived watermarks is presented.

Wmk schem.	LPF	Avg.+ LPF	HPF	Avg.+ HPF	JPEG	Crop	Avg.+ JPEG +Crop
SS	70%	0	0%	0	100%	100%	5.8
IA-W	87%	0	0%	0	100%	100%	2.5
IA-DCT	65%	0	0%	0	100%	100%	10
MBE	74%	0	100%	10	100%	100%	5.8

Image adaptive schemes embed watermarks only if the magnitude of the coefficient of frequency is larger than JND. Most high frequency coefficients - after 8×8 block DCT for IA-DCT and four level decomposition for IA-W - are close to zero, and consequently lower in magnitude than JND. A very short watermark sequence is inserted into high frequency region in IA-DCT and IA-W, which barely survive HPF.

2) Performance of our MBE scheme: Under MBE scheme, watermarks are embedded in the 1000 highest magnitude coefficients. Hence if a smoothing operation is performed on the watermarked image, these watermarks, which as mentioned before mostly belong to low-frequencies, survive and therefore are detected. In the case of MBE, since watermarks were forcibly embedded in the high frequencies, they survive the HPF and thus are recoverable.

From Column 2 in Table I, we notice that none of the four schemes can detect a single colluder, if averaging of 10 watermarked images and LPF are performed. The failure calls for more robust schemes.

V. Conclusions

Tracing illegal distribution of medical images is a problem that has not been addressed so far. By showing that SS, IA-DCT and IA-W fail simple HPF, we demonstrated that they are unsuitable for watermarking medical images. Frequency selective operations are often used in medical images. We then proposed a multi-band watermark embedding technique (MBE) that survives HPF and LPF operations and preserves high quality in terms of PSNR. The MBE technique can be generalized to survive any given frequency selective filtering. It remains an open question whether it is possible to embed watermarks to survive all possible image processing operations on medical images while preserving high diagnostic quality.

References

[1] United Stated Department of Health and Human Services, “HIPAA: Medical Privacy - National Standards to Protect the Privacy of Personal Health Information,” [Online], Available: http://www.hhs.gov/ocr/hipaa/.

[2] HEMA, “DICOM: Digital imaging and communication in medicine,” [Online], Available: http://medical.nema.org/

[3] F. Cao, H. K. Huang, X. Q. Zhou, “Medical image security in a HIPAA mandated PACS environment,” *Computerized Medical Imaging and Graphics*, vol. 27, pp. 185-196, 2003.

[4] T. J. Owens, S. Tachakra, K. A. Banitsas, and R. S. H. Istepanian, “Securing a medical wireless LAN system,” *Proc. 23rd Annual Conf. IEEE Engineering in EMBS*, Istanbul, Turkey, 2001.

[5] G. Coatrieux, H. Maitre, B. Sankur, Y. Rolland, and R. Collorec, “Relevance of watermarking in medical imaging,” *Proc. IEEE EMBS Conf. on Inform. Tech. Appl. in Biomedicine*, Arlington, VA, Nov. 2000.

[6] I. J. Cox, J. Kilian, F. T. leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Processing*, vol. 6, No. 12, pp. 1673-1687, 1997.

[7] C. I. Podilchuk, and W. Zeng, “Image-adaptive watermarking using visual models,” *IEEE J. Selected Areas in Commun.*, Vol. 16, No. 4, pp. 525-539, 1998.

[8] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, 2002.

[9] J. J. Eggers, R. Buml, R. Tzschoppe, and B. Girod, “Scalar Costa scheme for information embedding,” *IEEE Trans. Signal Processing*, Vol. 51, No. 4, pp. 1003-1019, 2003

[10] K. Chen and T. V. Ranmabadran, “Near-Lossless Compression of Medical Images Through Entropy-Coded DPCM,” *IEEE Trans. Medical Imaging*, Vol. 3, No. 3, pp. 538-548, 1994.