Real-Time Watermarking of Medical Images and Secure Transmission Through Steganography

Ajay Biswas, Pabak Indu, and Souvik Bhattacharyya

Abstract Medical images and reports are highly confidential information and should not be leaked, or modified in any circumstances. Healthcare organizations constantly face cyber-attacks with the motive of stealing medical information for their personal gains, and during the COVID-19 pandemic, the attacks have almost doubled. Encrypting medical information and transmitting them over a secure channel ensures the confidentiality of information, however, can be stolen and decrypted later. Also, most of the health organization lacks in terms of security and privacy of their patient's information. A real-time watermarking and steganography protocol can mitigate the aftermath of a cyber-attack by hiding confidential reports into useless looking files such as log files and user manuals as soon the reports are generated. We propose a secure steganographic medical report system that can hide medical reports as well as identify fake ones. The system uses real-time watermarking during report generation which prevents any kind of modification; thereafter, these data are transmitted/stored using a secure steganography protocol. Finally, the data is retrieved by doctors or health workers after stego-extraction and watermark verification.

Keywords Features · Medical images · Steganography · Verification · Watermarking

1 Introduction

Digital watermarking is the process of marking digital multimedia such as images and videos by their owner to protect content and to claim ownership. Watermark is of two types: visible watermark and invisible watermark. A visible watermark is placed to warn users from stealing content and is directly placed over the multimedia in a suitable position with desirable transparency or visibility. Invisible watermarking on

A. Biswas (✉) · S. Bhattacharyya
Department of Computer Science and Engineering, University Institute of Technology, the University of Burdwan, Burdwan, WB 713104, India

P. Indu
Adamas University, Kolkata, WB 700126, India

M. Mukherjee et al. (eds.), *Advances in Medical Physics and Healthcare Engineering*, Lecture Notes in Bioengineering, https://doi.org/10.1007/978-981-33-6915-3_49

the other hand does not warn its users nor does it reveal its pretense. An invisible watermark is placed over multimedia using steganography. Steganography is the process of hiding information into a medium such as texts, images, videos, and images (Anderson and Petitcolas 1998).

A medical image of a person's organ or body part contains vital information about his/her health condition and shouldn't be disclosed against the patient's will. Securing medical images has always been a challenging task. They are stolen by hackers to get personal details of a person and later to dupe them by trapping them into fake medical treatments and insurance schemes. Not only medical images are susceptible to stealing but are also prone to modification and tampering (Cox et al. 1999).

Watermarking medical images to protect them from modification is quite new, and only a few researches have been conducted in this regard. Zain and Fauzi (2006) propose a medical watermarking system that provides tamper detection and recovery. Their proposed technique uses a secret key and a public chaotic mixing algorithm to embed and recover a tampered image. In another watermarking scheme, proposed by Puech and Rodrigues (2004) provides a crypto-watermarking method that uses private and public key ciphering. Eswaraiah and Reddy (2015) proposed a novel medical image watermarking technique that can detect tampering inside the region of interest (ROI) and recover the original region of interest. In another research by Parah et al. (2017), provided a discrete cosine transform (DCT)-based watermarking scheme for e-healthcare which can resist singular as well as hybrid attacks.

The rest of the paper is organized as below. In Sect. 2, we propose a client–server based real-time digital watermarking scheme. Section 3 provides results and analysis, and finally, Sect. 4 concludes our paper.

2 Proposed Method

Real-time watermarking images ensure better security and reliability over conventional watermarking. This is because a remote server is involved in the process, without its approval, the watermark verification or validation will not succeed. Our proposed architecture uses a remote server to communicate securely to its senders and store certain features in the database that will be later used in verification. The communication, image storage, and retrieval are done through steganography. Figure 1 provides the entire flowchart of the proposed system. Here, Alice has taken a medical image and wants to watermark it. He extracts features from the image and sends it to the remote server. The remote server verifies the sender and stores the features into the database. After storage, the server sends a unique identification number (ID) to the sender which will be later user in the watermark verification. Later, Alice watermarks the image with a unique ID. Now Alice can use this image as per his needs or sends it to someone else using steganography. Bob has received the stego image from Alice, and he performs stego-extraction to get back the watermarked image. He then sends the extracted unique ID and features to the remote

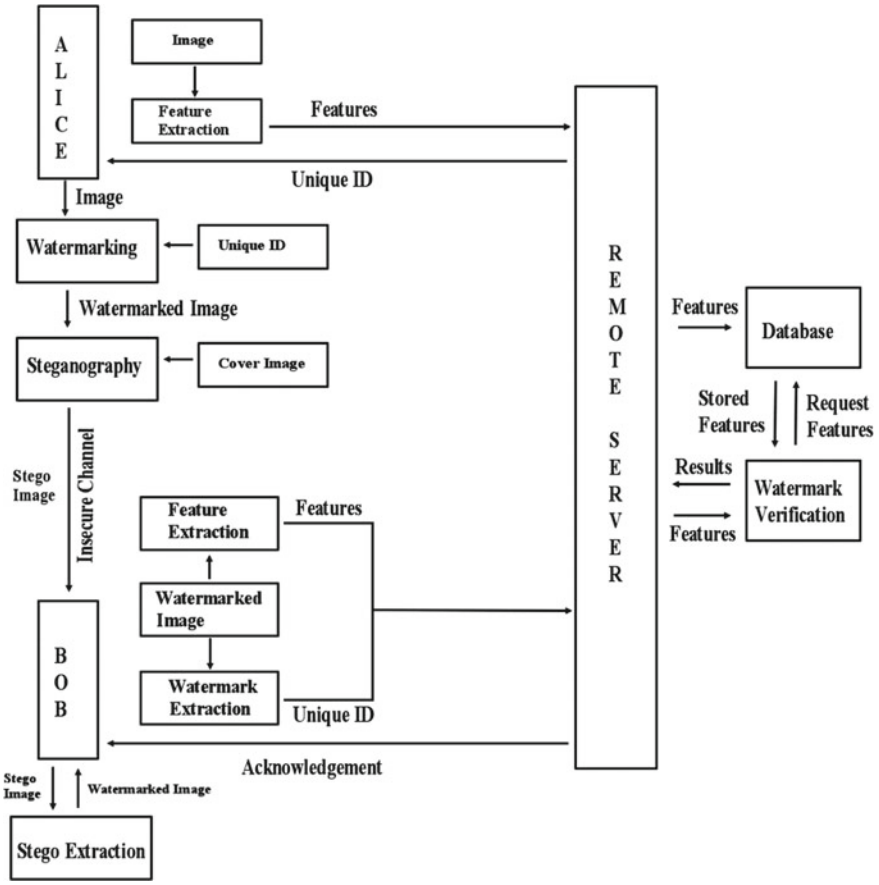

Fig. 1 Proposed real-time watermarking technique flowchart

server where it gets verified. If the verification is successful, the server sends back an acknowledgment. After receiving the acknowledgment Bob can safely assume that the image, he received is genuine.

The watermarking process is divided into two sub-processes, Sect. 2.1 describes the watermarking algorithm in detail, and Sect. 2.2 provides details of the watermark validation process.

2.1 The Watermarking Algorithm

The watermarking algorithm requires both server and client to be present. The key idea of this phase is to watermark a medical image so that gets to be treated as a genuine image. If an attacker tries to modify the image, he/she will have to incorporate

some change into the image and will result in the generation of a different feature set. As the server stores the feature set, the attacker's image will be rejected by the system during the verification. Algorithm 1 provides a detailed explanation of the watermarking process.

Algorithm 1 Watermarking Algorithm

	Input: Image *I* **Output**: Watermarked Image *W* **Data**: Feature Matrix *M*
1	Identify the watermarking region *b* in *I*
2	Extract features of *I* excluding *b* and store in matrix *M*
3	Use steganography to hide *M* and send it to the remote server using channel *C*
4	Server verifies the sender and stores M in the database
5	Server sends back a unique identifier *U* which can uniquely identify *M,* securely using steganography
6	Watermark *U* on the region *b*
7	Generate the resultant image *W* after watermarking
8	End

2.2 The Watermark Verification Algorithm

Similar to the watermarking algorithm, the watermark verification algorithm requires the presence of both client and server. Algorithm 2 describes the watermark verification phase in detail. During verification, sometimes the genuine image may not produce the same features as that generated in the verification phase. It may happen if the image undergoes compression or incorporates noise. To avoid this problem, the features have to be tested against machine learning models.

Algorithm 2 Watermark Verification Algorithm

	Input: Watermarked Image *W* **Output:** A Boolean determining the verification status **Data:** Feature Matrix *M,* Trained Dataset *T*
1	Identify the watermark region *b* in *W*
2	Extract features of *W* excluding *b* and store in the matrix *M'*
3	Extract the unique identifier *U* from *W*
4	Use steganography to hide M' and *U* and sends them to the remote server using channel C'
5	Server verifies the sender and retrieves *M* from the database using *U*

(continued)

(continued)

6	Server matches M' with M
7	If $M = M'$ then
8	Return *True*
9	Else
10	If $M \approx M'$ then
11	Test M' against trained Dataset T
12	If *Passed(M′)* then
13	Return *True*
14	Else
15	Return *False*
16	Else
17	Return *False*
18	End

3 Results and Analysis

The testing was performed using two machines, one of them acted as a server and the other as a client. The client machine contained the medical images which were later watermarked. Three different kinds of feature sets were used in the process, namely mean feature set, standard deviation feature set, and subtractive pixel adjacency matrix (SPAM) (Pevny et al. 2010) feature set. The mean and the standard deviation feature sets were generated by taking an 8 × 8-pixel block and applying mean and standard deviation, respectively. The SPAM686 feature set generates 686-second-order Markov-based features (Pevny and Fridrich 2007) of an image and is used to detect spatial domain steganography. The features generated from the SPAM686 can detect small changes in the spatial domain; therefore, it can be used to distinguish fake images from real ones. Figure 2 provides a comparison of "c4880h_s1" (Stegmann 2002) image with the images generated in three different scenarios.

After the watermarking stage, the images can be sent secretly using steganography. Table 1 provides the performance analysis of the least significant bit (LSB) (Bhattacharyya 2011; Chandramouli and Memon 2001) steganography with a hidden and visible watermark. The hidden watermarking was also performed using LSB steganography. Since the watermarking stage does not depend upon the steganography stage, the watermarked image should not affect the performance of steganography. Figure 3 shows the comparison of cover and stego image of "Lena 512 × 512" with "c4880h_s1 128 × 128" as cover image. The stego image was generated with peak signal-to-noise ratio (PSNR) of 54.16 dB.

To distinguish fake watermarked images from genuine watermarked images that were subjected to salt-and-pepper noise of density 0.01 and 0.001, the remote server

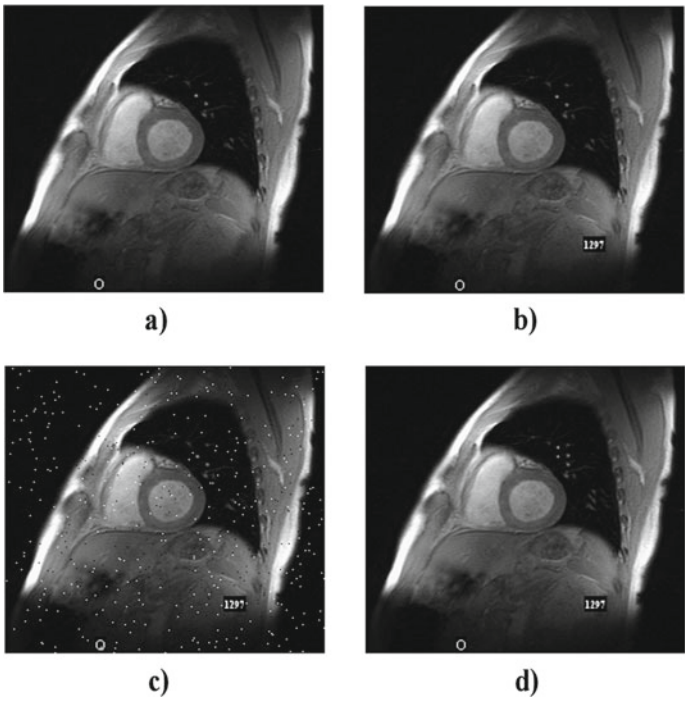

Fig. 2 **a** Original image; **b** watermarked image; **c** invisible watermarked image subjected to salt-and-pepper noise of density 0.01; **d** fake watermarked image

was trained and tested with 200 images from the Bossbase (Bas et al. 2011) using tenfolds cross-validated support vector machine (SVM) classifier (Ryu et al. 2008), 100 of each genuine and fake image. Figures 4, 5, and 6 provide receiver operating characteristics (ROC) curve for identifying fake images using mean, standard deviation, and SPAM features, respectively. Out of the three types of feature sets, the mean features provided better results. Both mean and standard deviation features provided better results when the noise was high; however, the SPAM features provided better results when the noise was low.

4 Conclusion

On analyzing results from Table 1, we can conclude that the proposed architecture does not affect the performance of steganography and kind of steganography can be incorporated until and unless the watermark is recoverable after stego-extraction. Also, we can see that the performance of both visible watermarked image and invisible watermarked image is similar; hence, any type of watermarking technique can be incorporated with the proposed architecture.

Table 1 Experimental results of LSB steganography over visible and invisible watermarked images

Cover image	Watermark type	Payload (bytes)	PSNR (dB)
Lena	No watermark	5174	59.12
	No watermark	17,462	53.87
	Invisible	5174	59.13
	Invisible	17,462	53.86
	Visible	5174	59.13
	Visible	17,462	53.87
Cameraman	No watermark	5174	59.14
	No watermark	17,462	53.87
	Invisible	5174	59.15
	Invisible	17,462	53.87
	Visible	5174	59.14
	Visible	17,462	53.87
Mandril	No watermark	5174	59.14
	No watermark	17,462	53.87
	Invisible	5174	59.14
	Invisible	17,462	53.87
	Visible	5174	59.14
	Visible	17,462	53.87

Fig. 3 **a** Invisible watermarked image; **b** cover Lena; **c** stego Lena

The proposed architecture provides a modular approach to validate and transmit medical images securely that give it an edge over current watermarking algorithms. In the future, testing will be performed with better steganography and watermarking algorithms with a variety of image formats. We will also work with other image distortion types.

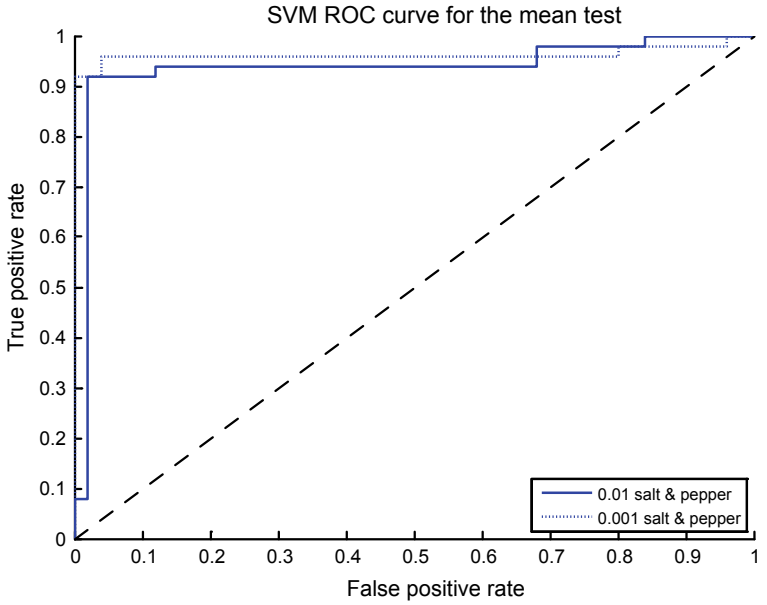

Fig. 4 ROC curve to distinguish fake images and noisy images using mean features and classification through SVM

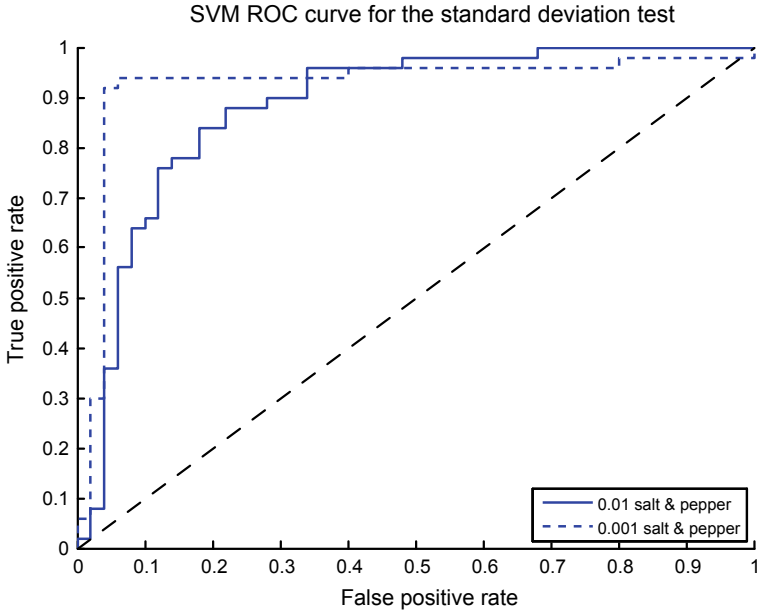

Fig. 5 ROC curve to distinguish fake images and noisy images using standard deviation features and classification through SVM

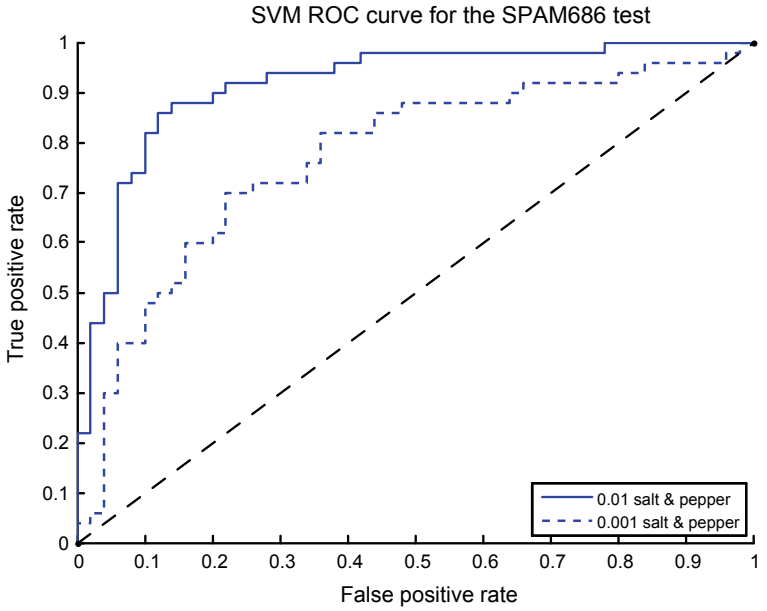

Fig. 6 ROC curve to distinguish fake images and noisy images using spam features and classification through SVM

References

Anderson RJ, Petitcolas FAP (1998) On the limits of steganography. IEEE J Sel Areas Commun 16(4):474–481

Bas P, Filler T, Pevný T (2011) "Break our steganographic system": the ins and outs of organizing BOSS. In: International workshop on information hiding. Springer, Berlin, Heidelberg, pp 59–70

Bhattacharyya S (2011) A survey of steganography and steganalysis technique in image, text, audio and video as cover carrier. J Global Res Comput Sci 2(4)

Chandramouli R, Memon N (2001) Analysis of LSB based image steganography techniques. In: Proceedings 2001 international conference on image processing (Cat. No. 01CH37205), vol 3. IEEE, pp 1019–1022

Cox IJ, Miller ML, Linnartz JMG, Kalker T (1999) A review of watermarking principles and practices. Digit Signal Process Multimed Syst 461–482

Eswaraiah R, Reddy ES (2015) Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest. IET Image Process 9(8):615–625

Parah SA, Sheikh JA, Ahad F, Loan NA, Bhat GM (2017) Information hiding in medical images: a robust medical image watermarking system for E-healthcare. Multimed Tools Appl 76(8):10599–10633

Pevny T, Fridrich J (2007) Merging Markov and DCT features for multi-class JPEG steganalysis. In: Security, steganography, and watermarking of multimedia contents IX, vol 6505. International Society for Optics and Photonics, p 650503

Pevny T, Bas P, Fridrich J (2010) Steganalysis by subtractive pixel adjacency matrix. IEEE Trans Inf Forens Secur 5(2):215–224

Puech W, Rodrigues JM (2004) A new crypto-watermarking method for medical images safe transfer. In: 2004 12th European signal processing conference. IEEE, pp 1481–1484

Ryu S-J, Lee H-Y, Cho I-W, Lee H-K (2008) Document forgery detection with SVM classifier and image quality measures. In: Pacific-Rim conference on multimedia. Springer, Berlin, Heidelberg, pp 486–495

Stegmann MB (2002) An annotated dataset of 14 cardiac MR images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark

Zain JM, Fauzi ARM (2006) Medical image watermarking with tamper detection and recovery. In: 2006 international conference of the IEEE engineering in medicine and biology society. IEEE, pp 3270–3273