

Medical Image Authentication and Restoration Using Block-Wise Fragile Watermarking and Clustering Approach

Ajala Funmilola A.¹, Ojebamigbe Victoria I.², Adegoke Benjamin O.³

^{1,2,3}Department of Computer Science and Engineering, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomosho
(¹faajala@lautech.edu.ng, ²ojebamigbe500@gmail.com, ³adegokebo@yahoo.com)

Abstract-Many medical image watermarks have been proposed to protect the authenticity and integrity of medical images in an age where digital images can be easily modified and perfectly reproduced. In a fragile marking system, a signal (watermark) is embedded within an image such that subsequent alterations to the watermarked image can be detected with high probability. The insertion of the watermark is perceptually invisible under normal human observation. These types of marks have found applicability in image authentication systems.

This paper focuses on the study of medical image watermarking methods for protecting and authenticating medical data and also proposes a fragile block based medical image watermarking technique for embedding data of patient into medical image.

Keywords- Watermarking, ROI, RONI, Tamper Detection, Recovery

I. INTRODUCTION

Most hospitals and health care systems involve a large amount of data storage and transmission such as administrative documents, patient information, medical images, and graphs. Among these data, the patient information and medical images need to be organized in an appropriate manner in order to facilitate using and retrieving such data and to avoid mishandling and loss of data when sharing.

The evolution of information technology and communication, offer to the medical sector many opportunities to practice medicine at a distance in order to enhance the quality of life and increase the quality of medical service which makes special safety and confidentiality of medical images an essential requirement, because critical judgment is done on medical images, therefore it must not be changed in an illegitimate way otherwise, an undesirable outcome may results due to loss of decisive information. Also since exchange of medical images is done through open unsecured network there is a need to provide strict security and authentication of medical images to ensure only occurrence of legitimate changes. (Zain and Clarke, 2007)

Digital watermarking is a technique use to ensure authentication and copyright protection, which includes the

embedding and extraction process. In embedding process some secret information is embedded into medical images. Extraction process deals with the extraction of secret message, which is embedded in the medical image. If failure occurs in extraction process the physician would come to know that there has been some kind of tampering with that image, and he would take precaution of not making diagnosis based on that image. However, if the extraction process extracts the correct watermark, which generally consumes a few seconds, physician can continue with diagnosis.

A digital watermark can be visible or invisible. A visible watermark typically consists of a conspicuously visible message or a company logo indicating the ownership of the image, when a television broadcaster adds its logo to the corner of transmitted video, the objective is to attach ownership or other descriptive information to the signal in a way that is difficult to remove. An invisible watermarked image appears very similar to the original, but with a level of secrecy instilled. The existence of an invisible watermark can only be determined using an appropriate watermark extraction or detection algorithm. An invisible watermarking technique consists of an encoding process and a decoding process.

Watermarking is broadly categorized into three classes namely fragile, semi fragile and robust watermarking. Robust watermarks are often used to prove ownership claims and so are generally designed to withstand common image processing tasks such as compression, cropping, scaling, filtering, contrast enhancement, printing, scanning. Fragile watermarking came in existence because of ensuring the legitimacy and data integrity especially when it is utilized as evidence in court or in medical diagnosis, A fragile watermark is a mark that is readily altered or destroyed when the host image is modified through a linear or nonlinear transformation. Also it is used in image authentication because of its sensitivity against alteration. Fragile watermarking again can be classified into two categories namely Block wise fragile Watermarking and Pixel wise fragile watermarking. In block-wise fragile watermarking, host image is divided into small blocks and watermark information is derived from the vital content of block of the host image. In case of image alteration, the tampered block and watermark contained in that block will mismatch and by this inequality one can easily identify the tampered block.

According to embedding and extraction criteria, fragile watermarking techniques for image authentication can be

divided into two categories: spatial domain and transform domain.

A. Spatial Domain

Spatial domain watermarking uses block by block watermarking e.g. they embed the watermarks on a randomly selected 8x8 blocks of pixels of the image.

B. Frequency Domain

To embed a watermark, a frequency transformation is applied to the host data. Then, modifications are made to the transform coefficients. Possible image transformations include Discrete Fourier Transform.

II. LITERATURE REVIEW

A block wise fragile watermarking proposed by Rey and Dugelay (2002), is standard technique which is based on scramble encryption. According to their approach the watermark, derived from a block is randomly distributed on to the LSB of the whole image. This scheme is suitable to identify the tampered block but lacks image recovery.

Fridrich and Goljan (1999) have also proposed the Block wise technique which is based on self-embedding in, they propose two methods to keep the image integrity. The first technique is based on quantization of block wise Discrete Cosine Transform (DCT) coefficient and is represented as 64 or 128 bits. These bits are used to replace one or two Least Significant bit (LSB) of another block. In the second method a new image is produced by reducing gray level of the original image. These reduced gray levels are cyclic shifted and embedded in to pixel difference. If some modification is done on watermarked image then the quantized DCT coefficient and the new reduced gray level image can be used to reconstruct the principal content of the tampered area.

Block-wise fragile watermarking has some limitations like within a particular block some pixels are really altered and some are not which is undesirable. Hence notion of pixel-wise fragile watermarking came into picture in which watermark information is derived from gray value of pixels and further embedded into image itself. Any alteration in gray value of pixel will be responsible for wrong value of watermark in further calculation at receiver side hence one can easily recognize altered pixel with high precision. A pixel wise fragile watermarking scheme is suggested by Y Lim *et al.* In this technique seven Most Significant Bit (MSB) of a gray value is given as an input to the hash function.

Using a secret key and hash value we calculate a single value either 0 or 1 for each pixel and this value is embedded in the first LSB of corresponding pixel. Any change in gray value of pixel will return wrong hash value and altered pixel can be identified easily. Pasunuri Nagarju et al (2013)

X Zhang and S Wang have proposed a statistical watermarking technique for accurately localizing tampered pixel in .They have calculated a set of tailor-made authentication data for every pixel with some additional test data and embedded into the host image. On the receiver side by

examining the pixels and their respective authentication data, one can reveal the exact pattern of the content alteration. All these proposals are good enough to identify the exact position for tamper but lacks in restoration of alteration. Hence some more proposals are also there which are used to recover the image content. X Zhang and S Wang have proposed fragile watermarking with error free restoration capability.

III. METHODOLOGY

Proposed algorithm is based on k-Medoids clustering scheme which is representative object based technique, also the algorithm is purely in spatial domain and watermark is invisible. The method involves embedding process which starts with the generation of watermark and later on the watermark is embedded in NROI, and extraction process. The block diagram of the proposed method is given below.

A. Watermark Embedding

Watermark embedding process which starts with the generation of watermark and later on the watermark is embedded in NROI. So first we describe the procedure for generation of watermark.

1) Algorithm to Generate Watermark

Step1. Generate a fixed hexadecimal number message by using a HASH function for a particular message defined by the sender, which is the Message Authentication Code (MAC)

Step2. Put MAC into a file in the sequence according to secret message

Step3. Read the text file containing the patient information

Step4. Convert character data into integer values.

Step5. Concatenate the data and MAC into a single line array having length say's', such that Table= (i) which is element of $[0,1]$, $1 \leq i \leq t$.

Step6. Convert data into corresponding binary code and form the vector which may have length of M bits such that vector = $\{w_2(i)|w_2(i) \in [0, 1], 1 \leq i \leq M \}$.

2) Watermark Embedding Algorithm

Watermark embedding process can be classified into four different phases viz. clustering of pixels, Recovery bit generation, Authentication bit generation and Block mapping. (Antony et al, 2007)

a) Clustering of Pixels algorithm

Step1: Start

Step2: Remove first 3 LSBs of all pixels.

Step3: Represent 5 binary bits each with P_i .

Step4: Divide the image I into number of blocks having dimension 4×4 .

Note: Each block will contain 16 gray scale values. So the total number of blocks will be $N/16$, every block which contains 16 gray values will be inputted to k-medoids clustering algorithm to make three different clusters.

Step5: Randomly choose k gray values in a block as the initial representative seeds.

Step6: Repeat step 5

Step7: Assign every remaining gray value to the cluster with nearest representative seed.

Step8: Arbitrarily choose a non-representative gray value.

Step9: Find the difference between representative gray value and non-representative gray value which is the total cost S

Step10: Compute the total cost S of representative gray value with non-representative gray value.

Step11: If $S < 0$ then swap the representative gray value with non-representative gray value.

Step12: Repeat until no change.

Step13: Calculate the mean for each cluster as a round integer.

Step14: Represent the mean as m1, m2, m3 in descending order.

Step15: Stop

b) Recovery bit generation algorithm

For each block 45 recovery bits is generated, these bits are formed as a vector V. The three means for a block is $m3 > m2 > m1$ in descending order.

Step1: Start

Step2: Map the mean values with their corresponding two bit pattern.

Step3: Convert m3 into 5 bit binary form and put on first five indexes of vector V.

Step4: Calculate

$$D1 = m3 - m2$$

$$D2 = m3 - m1$$

$$D3 = D2 - D1$$

Step5: Convert D1 and D3 into 4 bit binary form and put them into next eight indexes of vector V.

Step6: Map the all 16 gray level values with the two bit mapping binary bits for their corresponding mean values of those clusters in which they belong.

Step7: Stop.

c) Authentication bit Generation Algorithm

For each block, we are using three bits for alteration detection for more accuracy. These three bits are called Union bit, Affiliation bit and Spectrum bit. These names depend on their generation procedure.

Step 1: Start

Step2: Represent P_i as any pixel of BlockB.

Step3: Represent 5 MSBs of P_i as b_a where $a \in (3...7)$ for union bit calculation.

Note: b_a^r and b_a^c are binary value of corresponding row and column value in spatial image plane.

Step4: calculate

$$A^{s1} = \text{Ex-OR}(b_{a-3}^r, b_a^r) \tag{1}$$

$a=7,6,3$

$$A^{s2} = \text{Ex-OR}(b_{a-3}^c, b_a^c) \tag{2}$$

Note: Where, A^{s1} represents bitwise Ex-OR operation between row value and pixel value whereas A^{s2} represents bitwise Ex-OR operation between column value and pixel value.

Step 5: Union bit is given as

$$\sum_{t=1.16} (\sum_{f=1.2.5} (A_{ij}^{s1} \wedge A_{ij}^{s2}) \text{ mod } 2) \text{ mod } 2 \tag{3}$$

The second LSB of P_i is called as Affiliation bit. It shows the relation among all MSBs of P_i .

Step6: Affiliation bit is given as

$$\sum_{t=1.16} (\sum_{v=7.6.4} (b_{tv} + b_{tv-1}) \text{ mod } 2) \text{ mod } 2 \tag{4}$$

Step7: Calculate spectrum bit by using a binary matrix of size $\frac{N}{16} \times \frac{N}{16}$ which is generated pseudo-randomly with the help of a secret key.

Step8: Insert the 3 authentication bit for each pixel on the last 3 indexes of vector V.

Step9: Create a matrix M for all blocks having size 16 x 3 using forty eight bits and permute it using a secret key.

d) Block Mapping Algorithm

We cannot simply put forty eight bit information of one block into same block because, after any alteration we will not be able to recover the extensive pixel values if essential bit pattern from forty eight bit information which carries main information (till 13th) is lost.

Step1: Start

Step2: Using secret key pseudo randomly exchanges the content of the matrix M_i for one block with other matrix M_j .

Note: This step will be done for each matrix M_i where $i = [1, 2...16]$. The forty eight bits must be inserted on the first three Least Significant Bit positions of 16 pixels of mapping block. The pixel's gray level range will be increased from $[0...31]$ to $[0...255]$, whatever image is gotten from matrix M_i will be the watermarked image. Assuming that the original distribution of 3 LSBs is uniform, the average energy of distribution caused by watermarking on each pixel can be calculated as:

$$\text{MSE} = \frac{1}{mn} \sum_{t=0}^{m-1} \sum_{f=0}^{n-1} (I(I, j) - K(I, j))^2 \tag{5}$$

Where MSE is mean square value which is form $m \times n$ two monochrome image I and K in which one of the image is

original host image and another one is watermarked image. Now the PSNR (peak signal to noise ratio) is defined as

$$\text{PSNR} = 10 \log_{10} \frac{(MAX)^2}{MSE} \quad (6)$$

B. Watermark Extraction Algorithm

The Proposed algorithm is using blind approach for image alteration detection and recovery. It means at the receiver end there will be only one altered image and on the basis of our extraction algorithm we find the tamper location as well as restore it with good imperceptibility. Suppose any attacker has altered some pixel values without changing image size. Then at receiver end it will be desirable to detect tampered location. Hence alteration detection algorithm is as follows.

Step1: Start

Step2: Generates the pseudo random matrix of size $\frac{N}{16} \times \frac{N}{16}$ using same secret key which was used at the time of watermark embedding.

Step3: Extract forty eight bit stream from each block and using the secret key, match the forty eight bit stream with its corresponding block which was permuted at the time of embedding.

Step4: Rearrange the matrix M_i of size 16×3 for each block using the secret key and make it, as it was the time of embedding.

Step5: Calculate the Union bit and Affiliation bit for each block by the help of equation 3 and 4. Compare the calculated Union, Affiliation and spectrum bits with the corresponding extracted Union, Affiliation and Spectrum bit, if mismatch found then mark that block as altered one.

Once we locate the altered block we need to restore it in such a way so that proper imperceptibility is maintained. So for restoring the block, algorithm is as follows.

Step6: For each altered block, extract first fifteen rows from the corresponding matrix M_i and convert it into the form of row vector V .

Step7: The decimal of the first five bits from the vector V is highest mean value m_3 . Decimal of another four bit is D_1 and next four bits is D_3 . Calculate m_2 and m_1 using following way

$$\begin{aligned} m_2 &= m_3 - D_1 \\ D_2 &= D_3 - D_1 \\ m_1 &= m_3 - D_2 \end{aligned}$$

Step8: Replace all two consecutive binary bits from 14th to 45th position for each V , with their corresponding mean value. After that we get 16 gray level value which ranges from 0 to 31.

Step9: Appends three 0s as first three LSB at the end of each gray value to make the range from 0 to 255. Then make a matrix M_r of size 4×4 from that 16 value.

Step 10: Replace the altered block by its corresponding M_r matrix. This procedure will be done for all altered blocks.

Step11: Stop

Note: using equation 6 and 7 we can check the effectiveness of restoration.

C. Parameters for Evaluation of Performance of Authenticity of Image

The similarity between the original image and the watermarked image defines the authenticity of the watermark. The most common parameter used in the evaluation of the authenticity of the watermark is the Peak Signal-to-Noise Ratio (PSNR).

Peak Signal-to-Noise Ratio (PSNR): PSNR is a measure to indicate how close an image is to another and to analyze the watermark embedding distortions on images. PSNR can be calculated as:

$$\text{PSNR} = 10 \log_{10} \frac{(MAX)^2}{MSE}$$

Mean Square Estimate: This is given as:

$$\text{MSE} = \frac{1}{mn} \sum_{t=0}^{m-1} \sum_{j=0}^{n-1} (I(I, j) - K(I, j))^2$$

Where;

MSE is mean square value which is form $m \times n$ two monochrome image I and K in which one of the image is original host image and another one is watermarked image.

Payload: This is the actual data or information that is to be embedded in the image

Correlation factor: decides the similarity or mutuality between the original image and the watermarked image which defines the integrity of the watermark image.

Scaling Factor: This is the ratio of measurement of the watermarked image compared to the original image.

IV. RESULT

The system user interface is designed and developed in on MatLab 7.10a, which is a multi-paradigm numerical computing environment and a fourth generation programming language.

The proposed technique is tested on a brain dicom image with different payload and the scaling factor as shown in the table below. The watermarked image shows high embedding capacity up to 13K and good visual quality in terms of PSNR (Peak Signal to Noise Ratio).

The start page (fig 2) is the page that pops up that when the code is runned, it contains the title of the software, a button that when click on shows the interface where the image will be watermark and another button that where clicked on shows the interface to authenticate or dewatermark the watermarked image, when authentication is done on the image the Exit button is clicked.

When "Click to watermark" button is clicked on which is fig 3, it requires that a dicom image is loaded, when the image is loaded the metadata (information or details) of the dicom image is provided at the middle of the interface, it provides a button to watermark image and save watermarked image.

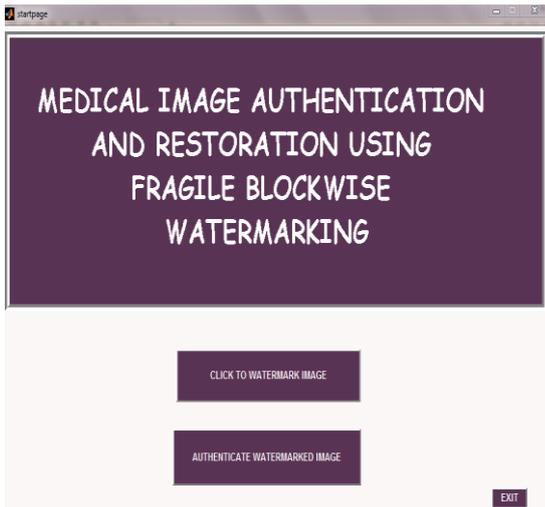


Figure 1. Start page

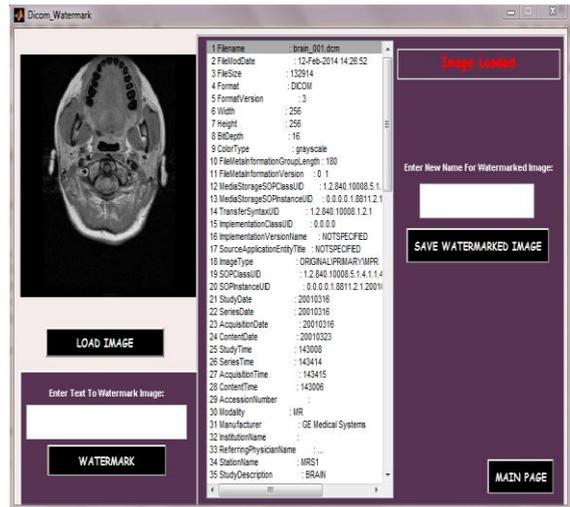


Figure 3. Image Loaded

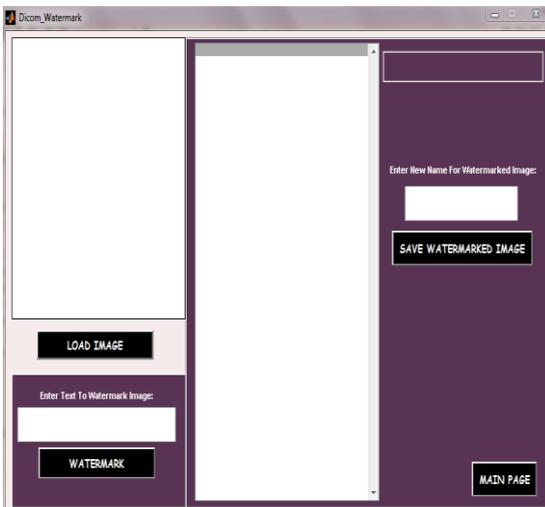


Figure 2. Click to watermark

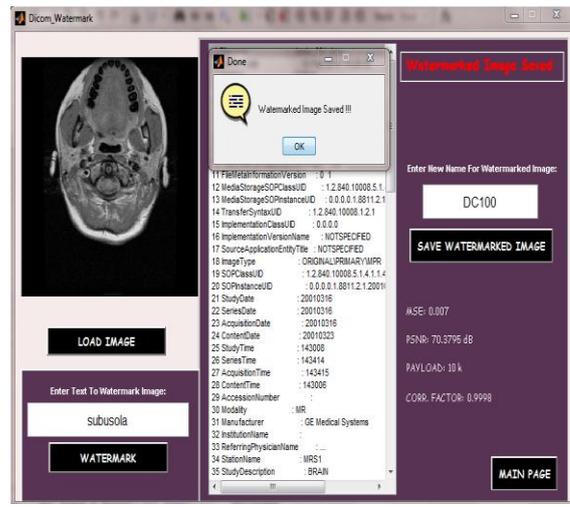


Figure 4. Saved watermark Image

In fig 4 a text to watermark image is entered which is case sensitive and the watermark button is clicked on to watermark image. After watermarking, it is required that the watermark image is saved with a name which enables the image to be easily retrieved when authenticating, also when the image is saved it shows the benchmark use for the evaluation of the image which are the MSE, PSNR, Payload, Correlation factor as shown in fig 5 Main page is clicked on to go back to the start page.

Fig 7 shows the interface to authenticate image. To authenticate the watermark image it is required that the saved watermarked image is loaded, after loading the metadata (information of the watermarked image) will not display except the text use for watermarking is provided (remember the text is case sensitive) as shown in fig 8, This is to ensure that the user that wants to access the patient information is an authorized user and has the correct text to dewatermark the image, else the patient information will not be displayed.

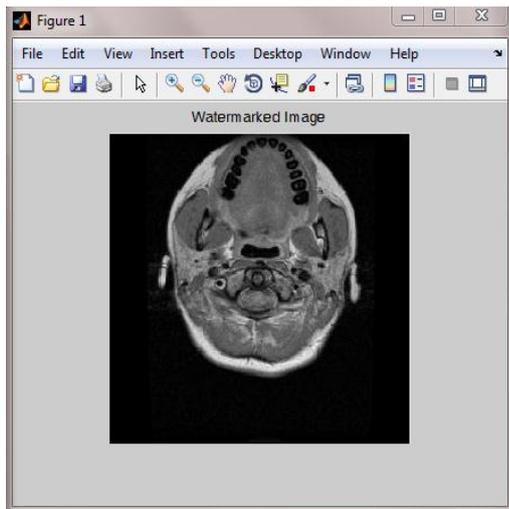


Figure 5. Watermark Image

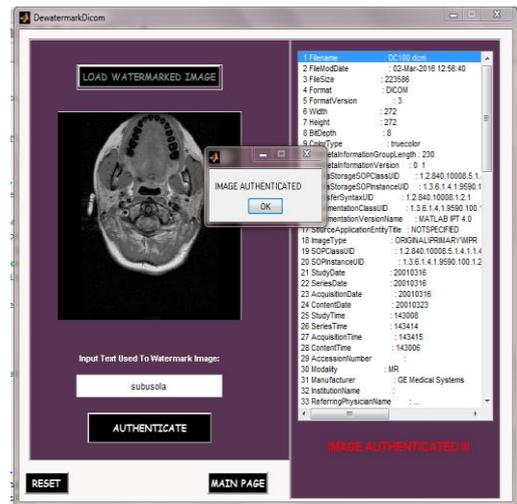


Figure 8. Image Authenticated

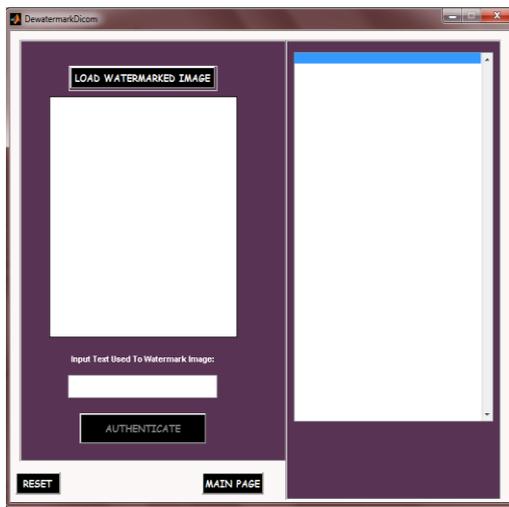


Figure 6. Authentication interface

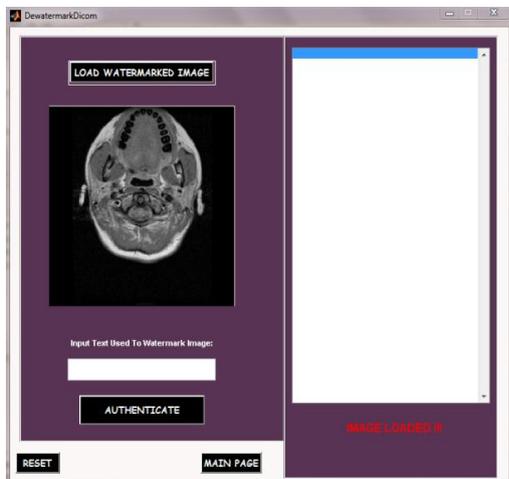


Figure 7. Interface of Image Authentication

Table 1 shows the result gotten on an x-ray brain dicom image at constant scaling factor, constant scaling factor means that the size of the dicom image has not been tampered with, also it has not been sized in any form to allow the different payload ranging from 8byte to 13byte (amount of information) that is to be embedded in it.

Table 2 shows the result gotten on an x-ray brain dicom image at constant payload byte, constant payload byte means that the amount of information that is to be embedded in the image is constant. Also the image can be resize to allow for a large amount of information (payload byte) to be embedded in it.

TABLE I. X-RAY IMAGE OF BRAIN AT CONSTANT SCALING FACTOR

Test image (Brain Dicom image)	Payload (Bytes)	MSE	PSNR (dB)	Correlation factor	Scaling factor
Image 1	8K	0.0036	73.1254	1	0%
Image 2	10K	0.0070	70.3795	0.9998	0%
Image 3	13K	66.9874	66.9874	0.9996	0%

TABLE II. X-RAY IMAGE OF BRAIN AT CONSTANT PAYLOAD BYTES

Test image (Brain Dicom image)	Payload (Bytes)	MSE	PSNR (dB)	Correlation factor	Scaling factor
Image 1	13K	0.0158	66.4536	0.9995	-10%
Image 2	13K	0.01222	67.5944	0.9997	0%
Image 3	13K	0.000694	69.4274	0.9998	+10%

From Table 1 above the watermarked image is embedded with a payload up to 13kbyte at constant scaling factor, when the payload is increase above 12kbytes there is change in the correlation factor which decides the mismatching from the original image, which means that when the amount of

information that is to be embedded in the image is more than 12kbytes there will be a change in the correlation factor (i.e. there will be a change in the similarity between the original image and the watermarked image) which affects the authenticity of the watermarked image and the PSNR of the image, at a low embedding factor there will be a good PSNR but if the embedding factor is high the PSNR will be bad. Therefore it is essential to monitor the embedding factor so that the quality of the image will not be affected by the watermark embedding.

From Table 2 the embedding capacity of the image can be increased by increasing the scaling factor but this leads to degradation in the required signal due to the degradation in the quality of the image. By increasing the scaling factor we can increase the embedding capacity of informatory data at a constant payload which allows for a constant integrity of image.

Therefore from the above it has been shown that the watermark embedding is invisible and has a good PSNR if the embedding factor is low, so the quality of the image will not be affected by the watermarking embedding. The watermark embedding is very sensible to any distortion. While embedding the data, ROI of medical image is avoided to ensure the integrity of ROI.

V. CONCLUSION

In this work, block based fragile watermark is used which divides the original image into small blocks and watermark is embedded in such a way that any attempt to alter the content of the image will also alter the watermark itself, also K-medoids is used which is a representative object based clustering technique which ensures that the gray level value which is used to replace other gray value within a block belongs to the same block. Clustering for each block is done by using a common characteristic hence for each element within a single cluster has same characteristics.

The experimental results gotten demonstrate that the watermark embedding is invisible and has a good peak signal-to-noise ratio (PSNR) if the embedding factor is low. The watermark embedding is very sensible to any distortion it generates a completely different watermark only when the image content is modified, this scheme is good at authentication which allows for the storing and transmission of electronic patient record along with image authentication codes or secret keys, which can be extracted at the receiving while the original image can be recovered perfectly.

REFERENCES

- [1] Anthony T. S. Ho, Xunzhan Zhu, lilian h. tang, (2007): Digital watermarking authentication and restoration for Chinese calligraphy image, IEEE, 1-4244-0882-2/07,
- [2] C. Rey, and J. L. Dugelay (2002), "A survey of watermarking algorithms for image authentication", EURASIP Journal on Applied Signal Processing Vol. 6, pp. 613-621. Jatinderkumar, Shailendra Singh, Dept. of computer science, Punjab engineering college chandigarh
- [3] J.M. Zain and M. Clarke (2007) "Reversible Region of Non-Interest (RONI) Watermarking for Authentication of DICOM Images," International Journal of Computer Science and Network Security, vol. 7, pp. 19-28,.
- [4] Jiri Fridrich and Miroslav Goljan (1999.): Images with Self-Correcting Capabilities. IEEE, 0-7803-546 2/99.
- [5] L. M. Vargas and E. Vera, "An Implementation of Reversible Watermarking for Still Images" IEEE volume 11, feb 2013.
- [6] M.E. Yalçin and J. Vanderwalle, "Fragile Watermarking and Unkeyed Hash Function Implementation for Image Authentication on CNN-UM", KatholiekeUniversiteit Leuven, Department of Electrical Engineering (ESAT), April 2002.
- [7] Mi-Ae Kim and Won-Hyung Lee. A Content-Based Fragile Watermarking Scheme for Image Authentication Springer-Verlag Berlin Heidelberg, AWCC 2004, LNCS 3309, pp. 258-265, 2004.
- [8] PasunuriNagarju, RuchiraNaskar and RajatSubhra Chakraborty,(2013) "Improved Histogram Bin Shifting based Reversible Watermarking", International Conference of Intelligence systems & Signal processing (ISSP).IEEE.
- [9] Y. Liu, W. Gao, H.Yao and S. Liu (2004), "A Texture-based Tamper Detection Scheme by Fragile Watermarking", Computer Science and Technology. Department of Harbin Institute of Technology.