BEST- LOCATION TECHNIQUE FOR MAMMOGRAM AUTHENTICATION BY LSB BASED WATERMARKING

CIK FERESA BINTI MOHD FOOZY

UNIVERSITI TEKNOLOGI MALAYSIA

BEST- LOCATION TECHNIQUE FOR MAMMOGRAM AUTHENTICATION
BY LSB BASED WATERMARKING

CIK FERESA BINTI MOHD FOOZY

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Computer Science (Information Security)

Faculty of Computer Science and Information System
Universiti Teknologi Malaysia

OCTOBER 2009

Special Dedicated to My Beloved Parents,
Mohd Foozy Bin Ghazali and Noraity Binti Nordin
My Brothers and Sisters,
Cik Farah Hani, Cik Fazleen, Mohd Farith and Fazly
My Supportive Friends,
Batch 9 Infosec, Shamala, ChaiWen, Sha, Zana, Huda, Ain, Ruby, Yus and Roslan.

ACKNOWLEDGEMENT

IN THE NAME OF ALLAH, MOST GRACIOUS, MOST COMPASSIONATE

Alhamdullilah, thank you to Allah for giving me the blessing for health, strength and earestness to accomplish and fulfill this project report.

I would like to take this opportunity to say the highest appreciation to Prof. Dr. Azizah Abdul Manaf as my supervisor who always gives full sopport and faithfulness in all guidance, advice and commitment upon on the effort for this project. To Dr. Akram M. Zeki, who also gives full support and help me in this project.

This appreciation also goes to my beloved family, lectures and all my friends who always encourage and give me full support to finish this project. Indeed, the courage and support is highly appreciated. Finally to Ministry of Higher Education, University Tun Hussein Onn Malaysia (UTHM) and the family of Faculty of Information Technology and Multimedia (FTMM) for both morale and financial support for this course. May Allah Bless all of you.

Thank you.

ABSTRACT

This project has been developed to embed the patient information into 25 samples of mammogram images in a grayscale format. MATLAB Software has been used as a tool to develop the algorithm where the algorithm has a capability to embed and extract the hidden data. Since the algorithm will identify the breast area before embedding process, the patient's information will be embedded at the best location which is at the background area without interrupting the breast area. Little distortion at the background images are seen when embedding is using LSB 1 and LSB 2. Meanwhile, few images are distorted when using LSB 3. However, there is no distortion identified at all samples when using LSB 4 until LSB 8. Finally, based on the PSNR result, it shows that the quality of watermarked mammogram images is in high value and the watermarked mammogram images are in a good quality.

ABSTRAK

Projek ini telah dibangunkan untuk memasukkan maklumat pesakit di dalam 25 jenis sampel imej mammogram yang berformat *grayscale*. Algoritma ini dibangunkan dengan menggunakan perisian MATLAB untuk memasukkan maklumat di dalam imej dan membaca kembali maklumat yang telah disembunyikan. Oleh kerana, algoritma ini akan mengenalpasti kawasan dada sebelum prosess memasukkan maklumat pesakit dilakukan, maka maklumat tersebut akan dimasukkan di kawasan yang terbaik iaitu di kawasan latar belakang mammogram di mana proses ini tidak meganggu kawasan dada pada imej mammogram. Sedikit kerosakkan pada bahagian latar belakang imej boleh dilihat jika menggunakan LSB 1 dan LSB 2. Manakala beberapa imej mammogram akan mengalami kerosakkan jika menggunakan LSB 3. Tetapi tiada kerosakkan dikenal pasti pada sampel apabila menggunakan LSB 4 sehingga LSB 8. Akhir sekali, berdasarkan keputusan PSNR yang diperolehi, ia menunjukkan bahawa kualiti imej mammogram yang telah dimasukkan maklumat pesakit adalah tinggi dan kualiti mammogram tersebut adalah baik.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NAME		**DESCRIPTION**
BMP	-	Bitmap
DWT	-	Wavelet Transform
DCT	-	Cosine Transform
JPEG	-	Joint Photographic Experts Group
LSB	-	Least Significant Bit
MSE	-	Mean Squared Error
PSNR	-	Peak Signal to Noise Ratio

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1   Introduction

As the technology drastically change especially in the medical environment, the medical imaging technology such as mammogram, Magnetic Resonance Imaging (MRI) and others are widely used in the medical domain for clinical purpose such as medical procedure to diagnosis the patients' internal problem without having the surgery operation. For example, mammogram is used to identify the breast cancer and it helps to detect the breast abnormality. Based on Pam Stephan (2007), all women are recommended to have a baseline mammogram at the age of 40 and then continued by a mammogram every couple of years until 50. Then after 50, a woman should have a mammogram every year. Below are several types of mammogram views that will be taken for screening:

i.   Cranio-caudal (CC) is taken from above a horizontally compressed breast.

ii. Medio Lateral Oblique (MLO) is taken from the side and at an angle of a diagonally compressed breast.

iii. Latero Medial (LM) is taken from the side towards the center of the chest.

iv. Mediolateral (ML) is taken from the center of the chest out. Spot compression is compression on only a small area this is to get more detail of the breast structure.

v.   Cleavage view is a view from both breast compressed. This is to see the tissue nearest center of chest.

vi. Magnification is to see borders of structures and calcifications.
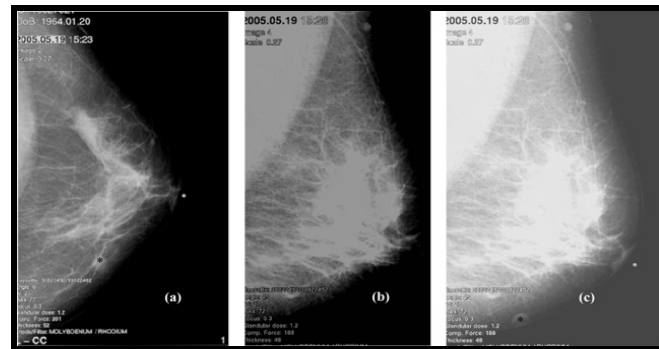
Figure 1.1: Cranio-Caudal (CC) View
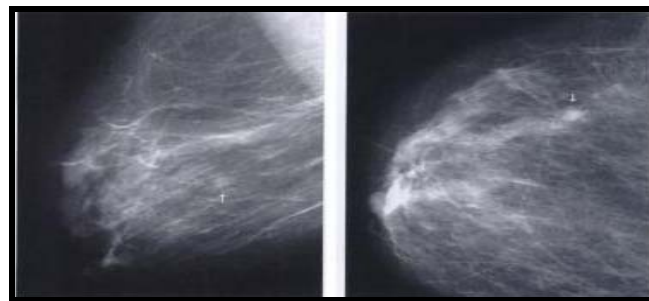
Figure 1.2: Mediolateral Oblique (MLO) Mammographic View

Figure 1.3: Medio Lateral (ML) Mammographic View

Since there are many advantages of medical images are discovered and it is frequently used in the medical domain, most hospitals are facing with issues to manage large amount of data storage such as administrative document, patient's information and medical images. Therefore, it is important to handle those data accurately to avoid problem of lost, tampering and mishandling record at the hospital (J. Nayak et al., 2008).

There are few ways to manage the patient's information and the medical images and one of them is using the watermarking technique in medical images. The purpose of watermarking is to hide a message into the host of document or multimedia format. As one to many communications technique, when sending the watermarking image, only authorized user can read the hidden message. Thus, by inserting watermarking in the medical image, it will help to authenticate the owner of medical image. This is to ensure the medical image is issued from the right source and to solve the missing medical document problem (G.Coatrieux et al.2008).

Moreover, there are many researches of medical watermarking images that has been done such as C.S.Woo et al. (2006) proposed a multiple watermarking method to store the medical images in a digital form and in a secure way in order to avoid the data from being exposed to the unauthorized person. A. Giakoumaki et al. (2006) proposed a wavelet based multiple watermarking schemes in the medical images for secure and efficient health data management. D.C. Lou et al. (2009), embedded large amount of data, maintained the quality of medical image and restored the original image after extracted by using multiple layer data hiding in spatial domain and Least Significant Bits (LSBs) technique.

Additionally, several advantages of watermarking in the medical image have identified by S. Boucherkha and M. Benmohamed (2005) such as:

i. By embedding watermarking in medical image, the patient can protect their information such as diagnosis result or personal details from being viewed by unauthorized person.

ii. Medical image watermarking can help to authenticate the patient, if the connection between the image and patient is lost.

iii. In addition, medical image watermarking will help the staff to identify or search the old medical image in the Hospital Information System (HIS) collection because there are some cases that medical images and patient's records need to be verified the integrity before use.

iv. Images may help to discover new findings in medical case. Thus, it is needed to protect the copyright and integrity of the medical image by digital watermarking.

v. Embedding the authentication code in the image will make it less sensitive to attack than appending the information on medical image.

To get a clear scenario of medical imaging processes see **Figure 1.4** for the radiology department work flow. First of all, a patient will do a medical checkup with the doctor or the medical assistant. If a patient has problem that needs to do a medical imaging, the doctor will refer the patient to the consultant by writing a request to him. Then, the consultant will accept the request and the patient can make an appointment. If the consultant feels that patient needs to do medical imaging, the patient will be referred to the physician by sending a request form to apply imaging examination. The examination will be done by hospital personnel (nurse) under the supervision of the radiologist. The result of the imaging examination is a set of films that will be viewed by the radiologist. The radiologist will state his findings in a written format and sent it to the referring physician. Finally, the referring physician will perform the overall diagnosis and discuss the diagnosis treatment with the patient (G.Muller, 2008).


Figure 1.4: Radiology Department (G.Muller, 2008).

As we know, hospitals are the place for the patients to get treatment and all the medical records, patients' information and others need to be managed properly. Because of that, there are many systems are developed for the above purpose such as:

i.   Picture Archiving & Communications Systems (PACS)
ii. Clinical Information Systems (CIS)
iii. Radiology Information Systems (RIS)
iv. Cardiology Information Systems (CIS)
v.   Radiation Therapy Systems
vi. Patient Monitoring Systems

Picture Archiving and Communication Systems or PACS is a system for storage, retrieval, distribution and presentation of images. PACS uses Digital Imaging and Communication in Medicine (DICOM) format. The main purpose of PACS is to manage hard copy of medical images like film archives. Next purpose is to remote access for viewing and reporting or accessing the information from the difference physical location. However, the medical images that stored on PACS have been accessed every year thus the security and privacy of patients still in doubt.

Clinical Information system (CIS) is a system that built by IBM for patient's data management which is based on the patient's information age. The system was developed to replace the Medical Records Department of a medical institution, to supporting the storage, manipulation, and distribution of clinical information throughout the organization.

Moreover, Radiology Information Systems (RIS) is a solution which has function of radiology billing services, comprehensive financial tracking and audit trail, appointment scheduling, patient's managed care data, powerful reporting and patient's database storage. Cardiology Information system (CIS) is a multi-modality system that can be accessed by patients to access echo and nuclear cardiology, wave forms, report and patient's information. Radiation Therapy Systems is for radiation treatment such as breast conservation therapy. Finally, for Patient Monitoring

system, the purpose of this system is to monitor the critical patient. The example of the patient monitoring system is the heart monitor.

Most the applications that mention above are already implemented in some of the hospitals and clinics. For authentication of medical image, watermarking technique is proposed since there are few advantages that have been identified.

1.2   Background of the Problem

Currently, PACS has been used to handle the medical images and this system integrates with several systems in the hospital but the result of diagnosis is written in a piece of paper and the possibilities to lose the medical image or to be tampered by unauthorized users that enter into the system may occur.

Since, the growing trends in the maintenance of medical records give an impact to the security issues in the management of medical information. It creates the interest among the researchers to explore the authentication of medical images. Mohamed Kallelet et al. (2007) developed an algorithm for embedding multi signatures using a multiple watermarking scheme to preserve the image history in the medical field by using hash function and LSB technique to embed only the patient diagnosis in medical images. Y.Li et al. (2007) did watermarking method using the Discrete Cosine Transform Technique (DCT) to hide the information into the mammogram background.

For patient authentication, A.W.T. Goh et al. (2008) studied about an authentication of altering the image data and Rongrong Ni (2008) proposed new algorithm to locate the tampered regions. Experiments show the result of watermark medical image is PSNR=51.21dB. S.Boucherkha and M. Benmohamed, (2005), examine 3 medical images of ultrasound in difference sizes and the result shows the difference in PSNR (Peak Signal to Noise Ratio). G.Coatrieux et al. (2008), proposed

a method to ensure the authenticity and maintainability of data. From the findings this method allows medical image managers to gather only the data of the same patient without knowing the true identity of that patient. Since, patient identification methods are different from country to country in Europe, the data that has been embedded in medical images are the combination national health numbers and pivot Id such as a family based identifier referring to family medical records.

In the research community, the watermarking technology is recognized by them because of the medical confidentiality protection, origin and data authentication that potentiality contributes to the medical information management systems (A. Giakoumaki et al., 2006). Since, the medical data is stored in the system, the access control to the system will be given only to the authorized user and when the access is given, there is possibility that the patient's information does not remain in one place. Modification and lost can also be happen (G.Coatrieux et al., 2008). Thus, by inserting the watermark in the medical image, the problem of lost, untraceable, unintentional distortion and malicious modifications on the quality of medical images can be prevented. Along with that, the medical image watermarking will be able to trace the source or the owner of medical image (authenticate) and can detect the changes that have been made in medical image (integrity). However, by embedding the watermark on the medical image there will be possibilities that the image will be affected when extracting the watermark is done. When the quality of medical images is interrupted it can influence the diagnosis result.

There are few types of watermarking techniques such as Least Significant Bit Hiding (LSB), Direct Cosine Transformation (DCT) and Wavelet Transformation. LSB is the simplest method and S-Tools 4, HideBSeek, Steganos and StegoDos are based on the LSB replacement (S. Huliiv et al., 2004). Implementation random LSB technique hiddes the data in the least and second to least bits without being noticed easily. A study on the LSB embedding technique by D.Neeta and K.Snehal (2005), stated the result that LSB can hide the images in 24-bit, 8-bit or gray scale format of a .png file and a .bmp file. Moreover, this paper mentions that the data can be hidden in the least significant bits of the cover image and the human eyes would not be able to notice the hidden image in the cover file. D.C. Lou et al. P (2009) also used the

LSB technique to embed a large amount of data and maintained the quality of medical image.

1.3   Problem Statements

Based on the reviews on journals and papers that have done, it shows that the authentication is one of the relevant elements that needs to be implemented in watermarking of medical image application (G. Coatrieux et al., 2002) and the quality of medical images are really important when embedding and extracting the watermark. Below are the problem statements:

i.   Embedding the watermarking randomly will affect the quality of mammogram image thus it will affect the diagnosis.

ii.   The breast area which contains important medical information should not be modified during the watermarking.

iii.   Embedding the patient's information in the best location would not affect the critical areas or potion of mammogram.

1.4   Project Aim

The project aim for this research is to identify the best location on the mammogram to embed the patient's information without affecting the quality of the image and to develop an authentication technique of watermarking mammogram using the LSB technique.

1.5   Project Objectives

Below are the objectives of the project:

i.   To study and analyze the current technology of medical image authentication.

ii.   To investigate the issues and problems in authentication technique in mammogram.

iii. To develop an authentication technique of watermarking mammogram using the LSB technique.

iv.   To identify the best location on the mammogram to embed the patient's information without affecting the quality of the image.

1.6   Project Scope

Below are the project scopes of authentication medical image watermarking project:

i.   Twenty five (25) samples of mammogram images will be used to test the algorithm for quality and authentication testing purpose at the extraction process. The mammogram images samples are taken from Breast Cancer Cases Molson Informatics Project, McGill Faculty of Medicine website.

ii.   The mammogram images will be in grayscale format.

1.7   Summary

The advances in multimedia and communication technology nowadays have provided new ways to store, access and distribute medical data in a digital format. Hence, by identifying best location on the mammogram to embed the patient's information without affecting the quality of the image will be studied. The development of authentication techniques for mammogram watermarking can be as one of the alternatives to overcome the weaknesses of losing and mishandling of medical images that will risk the patients. Moreover, comparison with old mammograms is very important for breast cancer diagnosis (Diagnostic and Interventional Radiology, 2008) and by embedding watermarking in medical image will be easier to identify the medical image by authenticating the patient to prove that the medical images belong to the correct person and the medical image is issued from the trusted source.