Web based Image Authentication Using Invisible Fragile Watermark

***Yusuk Lim, * **Changsheng Xu, * *** David Dagan Feng**

* Biomedical and Multimedia Information Technology (BMIT) Group,
Basser Department of Computer Science The University of Sydney NSW 2006 Australia
**Kent Ridge Digital Labs 21 Heng Mui Keng Terrace Singapore 119613
***Centre for Multimedia Signal Processing (CMSP)
Department of Electronic & Information Engineering, Hong Kong Polytechnic University

yslim@it.usyd.edu.au

Abstract

The demand of security is getting higher in these days due to easy reproduction of digitally created multimedia data. Digital watermark is the emerging technique to embed secret information into content for copyright protection and authentication. Watermark is embedded within an image that alteration and modification to the watermarked image can be detected in a fragile watermark system. Watermark detection is blind that does not require an original image and it is invisible to avoid revealing secret information to malicious attackers. A web based image authentication method based in digital watermarking is described in this paper. It can provide more controls to image owners and conveniences to clients who want to get authenticity of image by integrating benefits of using Internet to the watermark system.

Keywords: *Watermark, Authentication, and Client-server.*

1 Introduction

Headings should use the heading styles as shown. Numbering is automatic. The rapid development of computer and network technology has led to new era of digital multimedia. There are many advantages using advanced digital multimedia data such as easy creation, edition, reproduction and distribution. These advantages can facilitate unauthorized use as well, such as illegal copy and modification of the content. It is one of the biggest issues for content providers to protect their intellectual properties. There are two issues to be concerned; one is the protection of ownership such as copyright of artistic work and the other is authentication that content should be identical to the original when it is distributed. The watermark can provide a solution to the issues for copyright protection [1-3] and authentication [4,5].

There are many types of watermarking methods proposed and developed these days. The fragile watermark [6,7] is one of the watermarking methods for authentication that has a low robustness towards modifications where even small change of the content will destroy embedded information showing that there has been an attempt of attack. High robustness is a requirement for copyright protection to provide ownership in any kind of attacks. The fragile watermark method is useful to the area where content is so important that it needs to be verified for it being edited, damaged or altered such as medical images.

The traditional method used for authentication is cryptography. It hides data to unauthorized person so called preserving confidentiality [8]. The popular method of cryptography is public key encryption, which encrypts data using a private key, and an associated public key is used for decryption of secret message. The problems that might be arisen in this method are difficulty of maintenance and distribution of public key. It has great strength in confidentiality, but when the data is revealed to unauthorized personnel, there is no protection for content it-self, which is integrity control. There is a company that doing verification using digital signature technology [7] for text messages, but the proposed system will use digital watermark as an authentication method for images. The watermark has significant advantages compared digital signature systems [9] such as direct embedding of secret information to content it-self requires no additional information for verification while digital signature needs extra transmission of data.

Schneider and Chang [6] proposed a method to embed content-based signature using private key as a watermark. This authentication scheme also requires distribution of public key to verify the watermarked image. But the system proposed in this paper uses client-server model that server holds an watermark detection method internally and client can access to the server using Internet to verify the image, which does not requires distribution of public key that maybe the major problem of using public key encryption.

The web-based image authentication system using invisible fragile watermark will be discuss as follow. Section 2 introduces a web-based image authentication method based on digital watermarking including watermark embedding and server authentication. Section 3 shows the experiment results. Section 4 discusses more robust authentication methods in spatial and frequency domains using fragile watermark. Concluding remarks are given in Section 5.

2 Web-Based Authentication Method

The methods to be discussed in this section are Internet based client-server model and watermark embedding scheme. The web-based authentication system consists of two parts: one is a watermark embedding system and the

Copyright © 2002, Australian Computer Society, Inc. This paper appeared at the Pan-Sydney Area Workshop on Visual Information Processing (VIP2001), Sydney, Australia. Conferences in Research and Practice in Information Technology, Vol. 11. David Dagan Feng, Jesse Jin, Peter Eades, Hong Yan, Eds. Reproduction for academic, not-for profit purposes permitted provided this text is included.

other is authentication system. In case of watermark embedding system, it is installed in the server as application software that any authorized user, who has access to server, can generate watermarked image. The distribution can use any kind of network transmission such as FTP, e-mail etc. Once image is distributed to externally, client can access to authentication web page to get verification of image.

2.1 Watermark Embedding

The watermarking scheme used in this system is based on the work of Yeung and Mintzer [8]. The major difference is that it uses 7 most significant bits as an input for hash function where Yeung and Mintzer used color values. The hash function generates binary value of 0 or 1 using secret key that will be substituted to least significant bit (LSB) of each pixels.

The embedding process starts with a secret key that is used to generate a key dependent binary value function f, f:{7 MSBs}-> {0,1}. Since the chance of changing original color value is around 50% that the change would be minor and also it is fragile that any change of pixel can result in detection.

For a RGB image, the watermark method can be expressed as;

$$L(I,j) = f_R(R(i,j)) \oplus f_G(G(i,j)) \oplus f_B(B(i,j)) \quad \text{for each pixel } (i,j) \qquad (1)$$

where;

f:COLOUR(i,j): {7 MSBs} → {0,1} (hash function with a key)

f:f_{colour}(binary) → LSB substituted colour value

Colour = {R, G, B}

i,j : integer variables.

⊕ stands for binary addition.(each values are shifted then added).

The watermark is embedded according to this algorithm refer to Figure 1.The verification is easily checked using the relationship L(i,j)=f_{color}(color(i,j)) for each pixel (i,j).

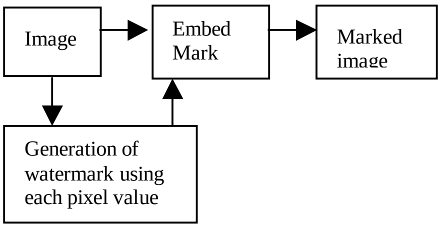


Figure 1. Watermark embedding

2.2 Server Authentication

The secure web server system has its own security policy to protect against unauthorized use of it, and also the firewall to add network security. The server assumed to be located in a safe location for physical security. The server should be secure enough to install authentication system inside of it.

The process of verification would be started with transferring watermarked image to the server. Once the image is uploaded, server uses its private key to detect watermark. There could be an attack while transmitting image file to the server, so the extra encryption can be used to hide image. There is a concept called SSL (Secure Socket Layer), which is the transport layer over TCP/IP network to provide authentication of server, client and encryption of message [10]. Programming API like JAVA provides its own interface of SSL that will be used in the development of the system. This fulfils four major security aspects, which are the confidentiality using encryption, the integrity using watermark, the access control using pass word and the physical security of using server [9].

The detection process is an inverse function of (1) to check each pixel's LSB. If there is a difference in any pixel, the server will generate warning message that the image may have been modified or damaged. The block containing false pixel will be displayed instead of specific pixel since attacker may use this information to find out the binary function. All these information will be generated in to HTML format. The diagram of figure 2 displays the over all procedure and the structure of the client-server for authentication.

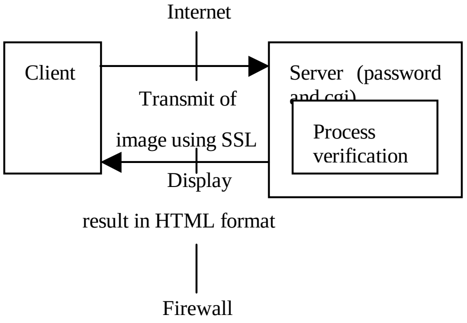

Internet
Client
Transmit of image using SSL
Server (password and cgi)
Process verification
Display result in HTML format
Firewall

Figure 2. Client-Server model

3 Experiment Results

The results showed that the implemented system detected major modifications applied for content of images. The tested modifications are geometrical modification, compression and rotation. The unaffected areas are shown in black whereas modified or damaged areas are shown in white. The average number of pixels have been changed for watermarking was 52% using 10 different images It is minor and invisible to human visual perception since it is proven that modification in LSB value does not affect on human visual perception by Johnson and Jajodia [18].

The watermarked images are modified and authenticated using the system. The original and watermarked images are shown in figure 3. In this case of image, 53% of the image's pixels have been change to embed watermark. The standard deviations of two images are same that has value of 52.67. The frequency of intensity is not affected by LSB substitution.

The geometry attack was applied to marked image in figure 4. The text is added using same colour with background which is black, the image sequence number in the top left hand side is changed and the colour modifications are applied some regions of the image. The regions with white colour are affected by modification in the right hand side image. The 95% JPEG compression is applied to watermarked image in Figure 5. Most of the region with color other than black turned out to be white informing of modification.

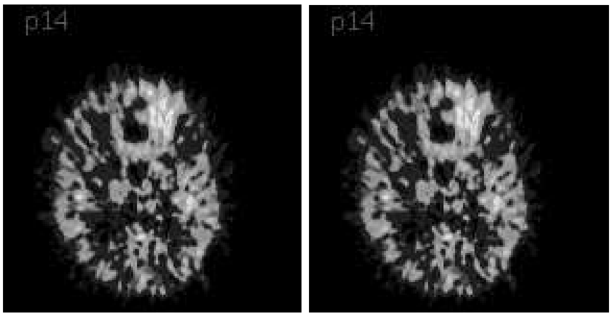


Figure 3. Original and marked image

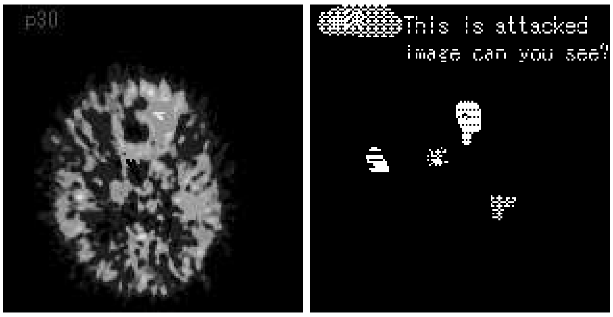


Figure 4. Geometry attack and detection

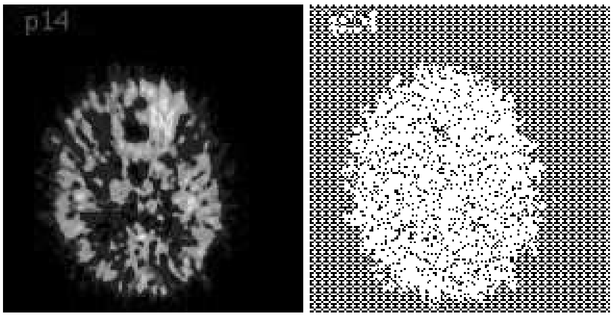


Figure 5. Compression attack and detection

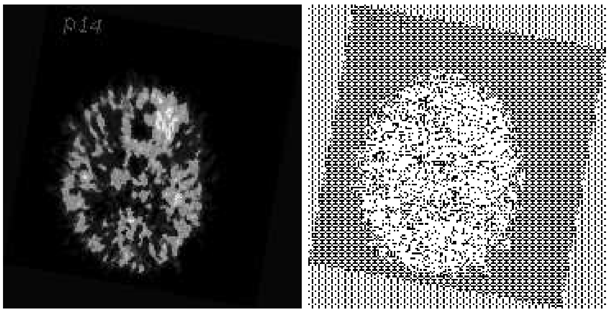
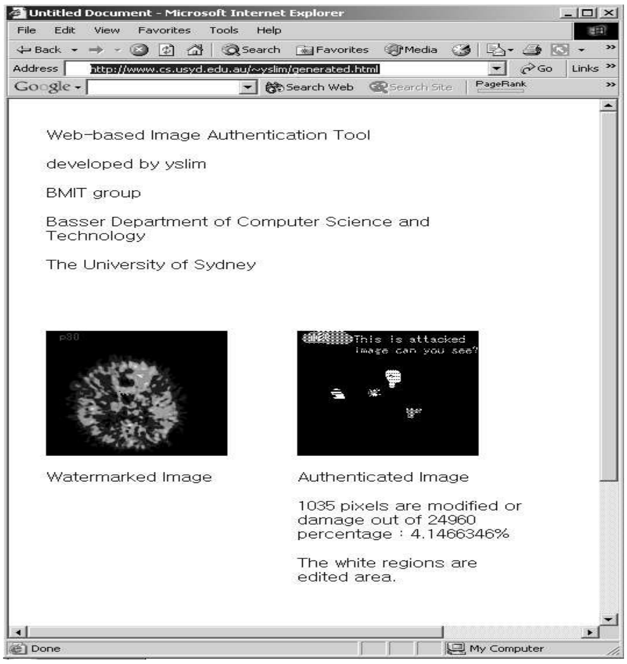


Figure 6. Rotation attack and detection

In Figure 6, geometrical rotation was applied to marked image. The black and white vertical lines shows that the image is modified. Above images are generated for authentication system reside inside server, a report of those results are generated in HTML format. The generated report in HTML format contains affected regions and text information; refer to figure 7. As a result, the implemented system can detect major attacks.



http://www.cs.usyd.edu.au/~yslim/generated.html

Web-based Image Authentication Tool

developed by yslim

BMIT group

Basser Department of Computer Science and Technology

The University of Sydney

Watermarked Image

Authenticated Image

1035 pixels are modified or damage out of 24960 percentage : 4.1466346%

The white regions are edited area.

Figure 7. Web generated authentication report

4 Discussion

The proposed system can be used where the content of image is valuable, which requires it to be ensured in the distribution that the copy is identical to original. For example, medical image requires great integrity of content since any change in image might affect the diagnosis even it is small amount. This could be the further research area to provide efficient image authentication tool for medical images. The thread hold of diagnostically acceptable distortion level is the key issue in medical images. Using invertible watermark and web-based authentication tool like the developed system may be used without affecting diagnosis. Cosman et al. [19] has revealed that lossy compression has little effect on diagnosis of CT images, which provides a feasibility of using watermark in medical image.

As it is mentioned above, the fragile watermark method used in this system has some weak aspects towards some attacks like lossy-compression, interfering secret binary logo proposed by Frederich et al. [12] and also if the binary function is used for long period of time, it gives more chance to attackers to crack the function by gathering information. There are other methods can be used in fragile watermarking system. The major categories are spatial domain method and frequency domain method. These two categories will be discussed in the following subsections.

4.1 Spatial Domain Methods

The spatial domain method is about embedding watermark information directly into images pixel such as a method proposed by Bender and Gruhl [11], and the method used in this implemented system by Yeung and Mintzer [8]. The LSB substitution is the main structure of this method. It is relatively easy to implement but there are disadvantages that can not be condoned such as it is not robust enough to protect watermark information against lossy compression and collage-attack introduced by Fridrich, Goljan and Memon [12]. The enhanced spatial domain methods were proposed in the aspect of robustness, such as obtaining digest from hash function [13] and adding a bipolar M-sequence in the spatial domain by Wolfgang and Delp [14]. The objective of fragile watermark is embedding breakable mark into the image. In this aspect, the spatial domain methods are applicable for fragile watermarking scheme if it can prevent revealing of imbedding functions.

4.2 Frequency Domain Methods

The embedding watermark in the frequency domain of a signal can provide more robustness than spatial domain. It is strong against attack like compression, cropping where spatial domain is not. The methods used for transform to frequency domain are discrete cosine transformation (DCT) [15-16] and wavelet transformation [17]. Those methods have high robustness and it is more applicable for copyright protection since fragile watermark is appropriate for authentication.

5 Conclusion

Digital watermarking is a new technology in multimedia and signal processing fields. It provides great help in copyright protection and authentication. In this paper, web based image authentication was investigated. There are advantages of using server-based authenticator since it can cope with distribution of public key and using extra security provided within server. The fragile watermark is efficient for authentication of content whether it is altered or not. The result showed that this system could detect most of modification to content of image. The more secure and accurate system can be achieved by using enhanced watermarking algorithms.

6 Acknowledgement

This research is supported by the ARC, UGC and CMSP Grants.

7 Reference

MINTZER, F., BRAUDAWAY. W. G., YEUNG. M .M. (1997): Effective and ineffective digital watermarking. ICIP.

SCHYNDEL, R. G., TIRKEL. A. Z., OSBORNE, C. F. (1994): A digital watermark. ICIP.

WOLFGANG, R. B., DELP, E. J. (1996): A watermark for digital images, ICIP.

YEUNG, M. M., MINTZER, F. (1997): An invisible watermarking technique for image verification. ICIP.

STORCK, D. (1996): A new approach to integrity of digital images. IFIP conf. On Mobile Communication.

SCHNEIDER, M., CHANG, S-F.: A Robust Content based Digital Signature for Image Authentication. in Proceeding IEEE International Conference on Image Processing, 1996, Lausanne, Switzerland.

XIE, L., ARCE, G. R. (Sep. 1998): A Blind Wavelet based Digital Signature for Image Authentication. in Proceeding of the European Signal Processing Conference, Rhodes, Greece.

YEUNG, M., MINTZER, F. (1997): An Invisible Watermarking Technique for Image Verification. Proc. ICIP'97, Santa Barbara, California.

WILLIAM STALLINGS (1999): Cryptography and Network Security 2nd edition, p 23-24, Prentice-Hall, Inc. ISBN 0-13-869017-0.

Netscape information web site, "Introduction to SSL",http://developer.netscape.com/docs/manuals/security/sslin/contents.htm.

BENDER, W., GRUHL, D., MORIMOTO, N. (1996): Techniques for data hiding, pp. 131-336, IBM Systems Journal, vol. 35, no. ¾.

FRIDDRICH, J., GOLJAN, M., MEMON, N. (2000): Further attack on Yeung-Mintzer Fragile watermark Scheme, *Proc. SPIE Electronic Imaging 2000*, San Jose, January 24-26.

WONG, P. (April 1999): A Watermark for image integrity and ownership verification, pp. 374-379, *Final Program and Proceeding of the IS&T PICS 99*, Savanna, Georgia.

WOLFGANG, R., DELP, E. (1996):A Watermark for digital images", pp. 219-222, *Proceedings of the IEEE International conference on image processing*, vol 3.

COX, I. , ET AL. (1996): A Secure, Robust Watermark for Multiemdia, *in Information Hiding: First International Workshop proceeding, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 185-206*

KOCH, E., ZHAO, J. (Jun. 1995): Towards Robust and Hidden Image Copyright Labeling, pp .452-455, *in IEEE Workshop on Nonlinear Signal and Image Processing.*

XIA, X., BONCELET, C. G., ARCE , G. R. (Feb. 1996): A Multiresolution Watermark for Digital images, *in Proceedings of the IEEE International Conference on Image Processing.*

JOHNSON, N. F., AND S. JAJODIA (1998): Exploring Steganography: seeing the Unseen, pp. 26-34, IEEE Computer, vol. 31, no.2.

COSMAN, P. C., ET AL. (Feb. 1994): Thoracic CT Images: Effect of Lossy Image Compression on Diagnostic Agency, pp. 517-524, Radiology, vol. 190, no.2.