Embedding Biometric Watermark on Text Document using Flipping

Ioi-Tun Lam and Chi-Man Pun
Department of Computer and Information Science
University of Macau
Av. Padre Tomás Pereira, Taipa
Macau, China
victorlit@gmail.com, cmpun@umac.mo

Abstract: - In respect to the issues on privacy, security and legal significance of a document, some sorts of security protections should be put on a document to ensure its genuineness and integrity. In this research, signatories’ encrypted digital biometric fingerprint in binary format will be embedded into a binary text document by Flipping – one of the methods in spatial domain. During the embedding process, document will be adaptively partitioned into blocks with fixed size of pixels according to the number of bits in the watermark message. Each watermark bit is embedded into each block by Flipping. Based on the odd or even number of pixels in each block on the embedded document, the fingerprint watermark message is extracted after decryption. Experimental results from our prototype system show that the proposed method is successfully tested for embedding and extracting a fingerprint watermark message in a document no matter it is written in hieroglyph or in alphabetic character.

Key-Words: - biometric fingerprint, spatial domain, watermark message, decryption, hieroglyph

1 Introduction
Traditionally, authentication is done by simply matching the signatures signed on the document against those in identity documents. However, this type of authentication is not highly guaranteed since signatures can be imitated and the authentication is not done by signature experts in general. Therefore, other technical methods in enhancing document security and authentication[1] have been developed in order to protect a document from being modified or counterfeited. Password validation, encryption, assign file permission[2], embedding visible security feature in a paper mass document[3] are some popular techniques used to protect a document from being unauthorized accessed and modified. In respect to authenticate the genuineness of a document, hiding some private information as an authentication message into it is a widely known method. Covert Channels, Steganography, Anonymity and Copyright Marking are some of the information hiding[4] techniques. Digital watermarking is one of the types from Copyright Marking and is a process of embedding information into a digital media such as text document (binary, gray scale or color), audio and video. If the media is copied, then the hidden information is also carried in the copy.

Due to the high uniqueness of human fingerprint, a thought of embedding a digital fingerprint[5] as the hiding watermark authentication message in a document emerges. In fact, this idea is worthyly promoted since fingerprint signature is hard to be imitated. By this way, document genuineness authentication will rely on the matching between the digital fingerprint extracted from the watermarked document and the digital fingerprint data previously archived in the system or captured immediately during the process. Can this idea be implemented? Which digital watermarking methods should be adequate to apply? How is it got done? What will be the performance? Are there any limitations? This research will focus on answering all of these queries.

2 Proposed Method
In general, digital data watermarking techniques[4] can be grouped into two classes: Spatial Domain and Transform Domain. Transform Domain[6] [7] such as[8] [9] [10] [11] transforms the original text document into frequency components and then embeds message into particular frequency regions. Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT) are two common methods in this domain. Whereas, Spatial Domain[6] [7] such as [12] [13] [14] [15] [16] [17] modifies pixel value directly and Flipping method is one of its implementing methods. Theoretically, Flipping, DWT and DCT methods can be applied to embed a watermark message in a grey or color digital document. But in practice, due to the discreteness features, Flipping is less favorable since its discrete flipping values may not be capable for applying to the gradual changing color spectrum compared with the continuous frequency transformation values in DWT and DCT. Moreover, the embedded document from DWT and DCT method is more robust and has higher visual quality than that from Flipping. Therefore, DWT and DCT methods are normally selected for watermarking a message to a grey and color document even though their complicated computation algorithm on frequency transformation involves larger overhead than in Flipping. However, DWT and DCT fail in embedding a watermark message to a binary document but Flipping can. In addition to Flipping’s higher data hiding volume, simplicity and efficiency advantages, it is popularly...
chosen for binary documents watermarking, and thus will be chosen as the proposed method. This method had been studied by a few scholars:

In 1999, Tomio AMANO and Daigo MISAKI [12] proposed a method in which each bit of a watermark was encoded as a positive or negative displacement of the difference between the average feature values of the two symmetrical partition sets; Min Wu and Bede Liu [13] proposed a method known as Shuffling which number of black points were averagely dispersed in every partition block; In 2000, Min Wu, Edward Tang and Bede Liu [14] proposed a method which one bit was hidden in each block by flipping pixel; In 2001, Q. Mei, E. K. Wong, and N. Memon [15] proposed a technique that information bits were embedded by modifying pixels directly along the 8-connected boundary of the character; In 2003 Haiping, Alex and Jun [16] proposed a method based on the Distance-Reciprocal Distortion Measure that provides an efficient way to select the pixels to flip in embedding to secure data-hiding algorithm for binary document images; In 2004, Min Wu and Bede Liu [17] proposed a method by calculating high scope (best point) and odd-even corresponding relation point to embed message by flipping pixels.

Flipping requires that both the original text document and the watermark message are in binary format which can be decomposed by numerous pixels, where each pixel represents one bit having either black (1) or white (0) state. A volume test between the text document and the watermark message (a fingerprint) will be performed beforehand to ensure a sufficient volume for embedding the watermark message. The volume test involves three steps: partitioning text document into pixels; determining block size of pixels grouping; and computing flipping block volume. (Fig.4.) below shows the changes of flipping block volumes against various partitioning figures in 2x2 and 3x3 block size. (Fig.10 vs Fig.12 and Fig.11 vs Fig.13) in experiment section below show that the visible difference on the watermarked images between 2x2 and 3x3 block size is almost imperceptible. For the sake of simplicity, a 2x2 block size will be selected. Moreover, (Fig.7) in experiment section shows that the fingerprint watermark message in resolution 128x128 is acceptable for the embedment. Based on the above findings and (Fig.4) results, 16384 blocks is the minimum flipping block volume required for the embedment and thus the text document partition resolution must be at least in 1024x1024. But in seeking for a better difference invisible between the text document and the watermarked text document, partition resolution 2048x2048 with block size =2x2 will be applied in this research.

The text document and the watermark message will first be partitioned into 2048x2048 (Fig.1) and 128x128 (Fig.7) pixels respectively. Pixels in the text document are grouped into blocks in size 2x2. And then, a volume test will be applied to check if there are sufficient flipping blocks which contain both black (1) and white (0) state pixels to embed the watermark message pixels. The message will be embedded pixel by pixel into the flipping blocks of the text document according to the following flipping rules: If the message pixel is in black (1) state, we must make the flipping block to contain odd number of black (1) state pixels (Fig.2) by flipping a white (0) state pixel into a black (1) state pixel in case that the block contains even number of black (1) state pixels. On the contrary, if the message pixel is in white (0) state, we must make the flipping block to contain even number of black (1) state pixels (Fig.3) by flipping a black (1) state pixel into a white (0) state pixel in case that the block contains three black (1) state pixels or by flipping a white (0) state pixel into a black (1) state pixel if the block contains only one black (1) state pixel. In all cases, we should select a pixel located on the boundary of a character for flipping so that the appearance of visible artifacts can be avoided and the flipping block must still contain both black and white state pixels after flipping.

![Fig. 1. Partitioned Text Document](image1)

![Fig. 2. Flipping to have an odd number black (1) state pixels in the block](image2)

![Fig. 3. Flipping to have an even number black (1) state pixels in the block](image3)

### 2.1 Embed Fingerprint watermark message to text document using Flipping

The following is the steps to complete the message embedment (Fig.5):

1. Partition the text document into 1048576 (2048x2048) small blocks in a way that each block contains 2x2 pixel patterns equally.
2. Encrypt the message using secret key.
3. Partition the fingerprint watermark message into 16384 (128x128) small pixels equally.
4. Embed the message pixel by pixel according to the flipping rule mentioned above.

### 2.2 Extract the message from a watermarked Document using Flipping

The following is the steps to extract the embedded fingerprint watermark message from a watermarked document (Fig.10):

1. Partition the watermarked document into 1048576 (2048x2048) small blocks in a way that each block contains 2x2 pixel patterns equally.
2. Check each block. If the block contains pixels all having the same state (either white (0) or black (1)), go to next block.
3. If the block contains odd number of black (1) state pixel, add a black (1) state pixel into the composing watermark message.
4. If the block contains even number of black (1) state pixel, add a white (0) state pixel into the composing watermark message.
5. Combine the message.
6. Decrypt the message using secret key.

Fig. 4. Volume Test: flipping block volume in different block size vs document partition resolution

Fig. 5. Embed a fingerprint watermark message to text document using flipping

Fig. 6. Extract an embedded fingerprint watermark message using flipping

3 Experimental Results

Experiments include: embedding a fingerprint watermark message into a Chinese (hieroglyph character) text document and a Portuguese (alphabetic character) text document; extracting an embedded fingerprint watermark message from a watermarked text document; and extracting an embedded fingerprint watermark message from a scanning image of a watermarked document.

3.1 Watermark embedment into a Chinese document and a Portuguese document

In this experiment, a Chinese and a Portuguese text documents (Fig.8 and Fig.9 respectively) both partitioned in 2048x2048 and a 128x128 fingerprint watermark message (Fig.7) are used for the embedment. After applying the above mentioned Flipping procedure with block size=2x2 to embed the fingerprint watermark message, their watermarked text documents (Fig.10 and Fig.11) are produced. In comparing the original text documents with their corresponding watermarked text documents, it can be seen that the difference between them is almost invisible. Repeating the same experiment with block size=3x3 and comparing the original text documents with their corresponding watermarked text documents (Fig.12 and Fig.13), it can be seen that the difference between them is also invisible. The requirement of invisibility after embedding a fingerprint watermark image into a text document is successfully proven. Moreover, two more findings can be concluded from this experiment: First, message embedment into a Chinese text document yields a bit better effect on difference invisible than embedment into a Portuguese text document (+27.03 dB vs +26.75 dB). In general term, Flipping method for watermark
message embedment works a bit better in hieroglyph character text document than in alphabetic character text document (see Fig.10 vs Fig.11, Fig.12 vs Fig.13). Second, block size does not produce significant difference on the watermarked text documents (see Fig.10 vs Fig.12, Fig.11 vs Fig.13). Thus, Spatial Domain’s Flipping method can be applied for document watermarking.

Fig. 7. Fingerprint Watermark Message (128x128). This text document is from http://forensictifact.wordpress.com

Fig. 8. Original Chinese text document (2048x2048). This document is from http://www.dsaj.gov.mo

Fig. 9. Original Portuguese text document (2048x2048). This document is from http://www.dsaj.gov.mo
3.2 Message extraction from a watermarked text document

In this experiment, the watermarked text document (Fig.10) is used for the extraction. After applying the above mentioned procedure for message extraction, an embedded fingerprint watermark message (Fig.14) will be extracted from the watermarked text document. In comparing the original 128x128 fingerprint watermark message (Fig.7) with the extracted fingerprint watermark message (Fig.14), it can be seen that the difference between them is almost invisible. Therefore, Spatial Domain’s Flipping method can be applied for extracting an embedded fingerprint watermark message which can be used for authentication against the digital fingerprint data previously archived in the system or the fingerprint captured immediately during the process.

3.3 Message extraction from a scanning image of a watermarked document

In this experiment, the scanning image (Fig.15) of the watermarked document (Fig.10) is used for the extraction. After applying the message extraction procedure, an embedded message (Fig.16) will be extracted. The extracted message fails to indicate a visible fingerprint image. Further study shows that position shift, size shrinkage and amplification of the scanning image cause the document partition to generate flipping blocks having black (1) and white (0) state pixel configuration different from those blocks in the watermarked text document even if they both are partitioned with the same resolution like 2048x2048. Once the state pixel configuration has been changed, the embedded message will equally be modified. Thus, Flipping method will extract a different watermark message.

Theoretically, having the assumption that shrinkage and amplification happened on the scanning image are equally applied in magnitude to the entire image, its state pixel configuration of the partitioned flipping blocks should be kept unchanged as soon as we can line out the partition area exactly the same as the watermarked text document while applying the same partition resolution. However, in practice, we have not found an application which can exactly line out the same partition area by tilting and shifting the entire image even if we surround the content of the original text document by a square composed by 4 very thin lines. Therefore, Spatial Domain’s Flipping method fails to extract the embedded fingerprint watermark message from a scanning image. And thus, it cannot be applied for a submitted document authentication purposes.

4 Conclusion

Spatial Domain’s Flipping method can successfully embed a fingerprint watermark message into a binary text...
document. The watermarked text document is proven to be difference invisible against the original text document. It can also extract an embedded fingerprint watermark message from a watermarked text document for authentication. Both message embedment and extraction will not produce significant difference in applying Flipping method with different flipping block size and in different character formats.

Based on the research result, authentication can be implemented by comparing signatory’s fingerprint captured immediately during the process against the fingerprint watermark message extracted from the watermarked text document archived in data centre. However, the authentication can merely successfully identify the genuine against signatories but not the document itself. In other words, the signatory could submit a counterfeit or forged document for an illegal purpose. In coping with this flaw, an ideal solution could be the one scanning the submitted watermarked document and extracting the fingerprint watermark message from it. And then, the authentication would be based on the comparison of the signatory’s fingerprint captured immediately during the process against the embedded fingerprint watermark message extracted from the scanning image. However, an experiment for extracting an embedded fingerprint watermark message from a scanning image is proven to be unsuccessful since the position shift, size shrinkage and amplification of the scanning image generate flipping blocks having black (1) and white (0) state pixel configuration different from those in the watermarked text document. Further studies on controlling these affecting factors might improve Flipping method’s watermark extraction capability against scanning image. Therefore, until this moment, Flipping method could not be fully considered as a method capable for authenticating document, particularly for a submitted one.

References


