Binary Image Watermarking with Tampering Localization in Binary Wavelet Transform Domain

Oh-Jin Kwon[1], Won-Gyu Kim[1] and Youngseop Kim[2]

[1]*Department of Electronics Engineering, Sejong University, Seoul 143-747, Korea*
ojkwon@sejong.ac.kr

[2]*Department of Electrical and Electronics Engineering, Dankook University, Yongin 448-701, Korea*
wangcho@dankook.ac.kr

Abstract

We propose a watermarking scheme for the authentication of binary images. We transform the image using the binary wavelet transform and partition the resulting coefficients into tree-structured blocks. The size of each block is decided by the number of flippable coefficients in the block. The flippable coefficients are found in the frequency domain by investigating high frequency coefficients. A random bit sequence generated by the block identifying key is embedded in the flippable coefficients in each block. Block-wise localization of tampered regions is accomplished with imperceptible visual degradations. Counter-measures against the vector quantization attack to which most existing watermarking schemes are vulnerable may also be implemented in our scheme

Keywords: *Binary Image, Fragile Watermarking, Binary Wavelet Transform, Image Authentication, Vector Quantization Attack*

1. Introduction

Image watermarking is the technology used for embedding information in digital images invisibly and is known to be an efficient method for image authentication. Binary image watermarking is increasingly required for important scanned documents. Authentication is of particular importance due to the ease of editing them fraudulently [1, 2, 3, 4, 5, 6]. References [1, 2, 3] proposed watermarking algorithms for binary images which could be used to test whether the image had been corrupted. Such tests give only a positive or negative result. The authors are aware of only a few papers for localizing tampered regions of binary images [4, 5, 6]. References [4, 5] proposed empirical algorithms localizing tampered regions in a block-wise fashion. The disadvantage of these algorithms is that they use fixed-sized blocks, each of which has a separately and independently embedded watermark. These methods normally suffer from block exchange or block copy attacks. Reference [6] avoids this problem by employing the block chaining process, embedding the position information of the previous block involved in watermarking in the block being currently watermarked. However, their experimental results demonstrate that false positive and negative errors may occur if sequential blocks do not have enough flippable pixels.

Most algorithms proposed so far perform the watermarking block-wise. The flaw in these approaches is that they use fixed-sized blocks. Whenever they encounter a block including an insufficient number of flippable pixels, such as a block with pixels that are mostly white or

black, this block must be skipped because embedding bits in the block inevitably causes distortions noticeable to the human eye. In this paper, we propose a binary wavelet transform

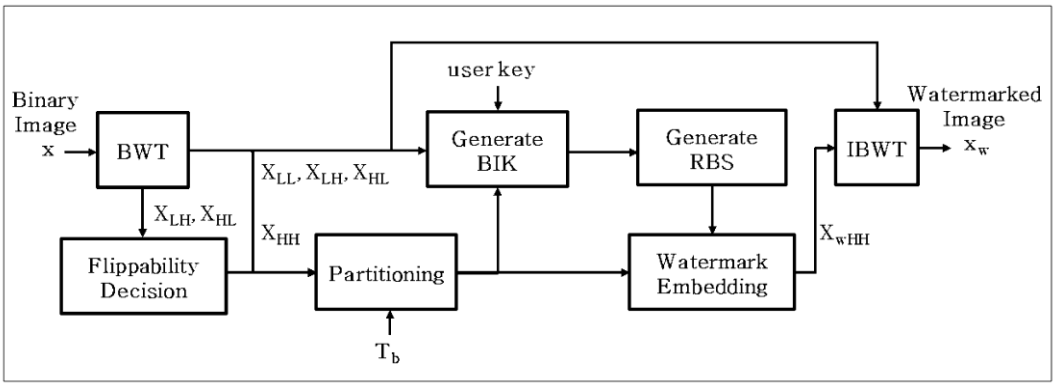


Figure 1. Block Diagram of Watermark Embedding

(BWT) based algorithm to improve tampering localization. Flippability of a pixel is decided in the BWT domain instead of in the traditional spatial domain. Whereas previous algorithms use blocks of fixed size, the proposed method firstly utilizes tree-structured blocks of flexible size for binary images. Based on the number of flippable BWT coefficients, it adaptively changes the size of blocks and prevents false positive and negative errors caused by the lack of flippable pixels.

In Section 2, our proposed algorithm is described. Experimental results are given in Section 3.

2. Proposed Algorithm

2.1. Watermark Embedding and Tampered Region Localizing

The proposed watermark embedding process is shown in Figure 1. The binary image, x, is first transformed by BWT. The BWT decomposes the input image into subbands, namely the LL, LH, HL, and HH bands. Flippability of a BWT coefficient is determined by considering the HL and LH band coefficients. Denoting the BWT coefficient from the kth row and the lth column of the LL, LH, HL, and HH bands by $X_{LL}[k,l]$, $X_{LH}[k,l]$, $X_{HL}[k,l]$, and $X_{HH}[k,l]$, respectively, this may be summarized as follows.

Method 1: If ($X_{LL}[k,l]$, $X_{LH}[k,l]$, $X_{HL}[k,l]$, $X_{HH}[k,l]$) = (#, 1, 0, #) and {($X_{LL}[k,l\text{-}1]$, $X_{LH}[k,l\text{-}1]$, $X_{HL}[k,l\text{-}1]$, $X_{HH}[k,l\text{-}1]$) $\neq$ (#, 1, 0, #) or ($X_{LL}[k,l\text{+}1]$, $X_{LH}[k,l\text{+}1]$, $X_{HL}[k,l\text{+}1]$, $X_{HH}[k,l\text{+}1]$) $\neq$ (#, 1, 0, #)}, then the position $[k,l]$ is the end point of the horizontal edge line and $X_{HH}[k,l]$ is flippable.

Method 2: If ($X_{LL}[k,l]$, $X_{LH}[k,l]$, $X_{HL}[k,l]$, $X_{HH}[k,l]$) = (#, 0, 1, #) and {($X_{LL}[k\text{-}1,l]$, $X_{LH}[k\text{-}1,l]$, $X_{HL}[k\text{-}1,l]$, $X_{HH}[k\text{-}1,l]$) $\neq$ (#, 0, 1, #) or ($X_{LL}[k\text{+}1,l]$, $X_{LH}[k\text{+}1,l]$, $X_{HL}[k\text{+}1,l]$, $X_{HH}[k\text{+}1,l]$) $\neq$ (#, 0, 1, #)}, then the position $[k,l]$ is the end point of the vertical edge line and $X_{HH}[k,l]$ is flippable.

In the above, the '#' indicates that this value is not considered in the decision. It will be shown in Section 3 that, compared to the methods that find flippable pixels by 3×3 block pattern matching in the spatial domain [2,4,5,6], this method has better capacity with less visual distortion.

X_{HH} is divided, using a tree structure, into the smallest possible blocks with the constraint that each block has at least T_b flippable coefficients. When the partitioning is complete for a block, the branch chain from the root to the current block is recorded. For each partitioned block, we concatenate 1) the branch chain; 2) the BWT coefficients of the LL, LH, and HL bands, which are unchanged by the watermark embedding process; and 3) the nonflippable coefficients of the HH band. We hash the result by a cryptographically secure hashing function and generate the so-called block identifying key (BIK) by encrypting the hash with a user key. A random bit sequence is generated using the BIK and is embedded in the flippable HH band coefficients as a watermark, where the number of embedded bits is variable for each partitioned block. Finally, we combine the flipped HH band (X_{wHH}) with the other bands (X_{LL}, X_{LH}, X_{HL}), perform the inverse BWT (IBWT), and obtain the watermarked image, x_w.

2.2. Improvement for Countermeasure against Vector Quantization Attack

Despite the performance of our method localizing content targeting attacks, it should be noted that our method is vulnerable to the vector quantization (VQ) attack which is also known as the Holliman-Memon attack, the birthday attack, or the collage attack [7]. The VQ attack assumes that the attacker knows the authentication scheme and has the database of images authenticated by the same user key. If we watermark and authenticate each block independently (as done in Section 2.1), the attacker may find the authenticated blocks in the same position in all database images and replace the original block with the closest match. Consequently, most watermarking schemes verifying each block independently fail to localize this VQ type of replacement.

Several possible methods for countermeasure against the VQ counterfeiting attack may be considered. The use of neighborhood-dependent blocks may be a choice and is of particular interest to this paper. This method eliminates the block-wise independence of the watermark by including information from neighboring blocks in the watermark of a block. Using this method, the VQ type of replacement is no longer authenticated by the watermark extracting process because the larger support covering the neighboring blocks is not preserved. When a group of blocks is counterfeited by the VQ attack, the boundary blocks enclosing the attacked area are detected as counterfeited blocks. In Section 2.1, we have constructed the BIK of a block from the BWT coefficients of the LL, LH, and HL bands and the nonflippable coefficients of the HH band. In order to implement the neighboring block dependency in our scheme, we also select the above-mentioned coefficients enclosing the block from neighboring blocks and include those boundary coefficients in the calculation of a block's BIK.

3. Experimental Results

In our scheme, the threshold value T_b for the block partitioning is a tradeoff between security and localization accuracy. If this value is larger, the size of the partitioned blocks becomes larger and the false decision rate decreases. We have chosen the value to be 15 for this paper's presentation. In this case, the maximum false decision rate for a block is 2^{-15} = 0.0000305 which is almost negligible. As for the hashing function, the MD5 and SHA algorithms have been most commonly used [3,4,5,6]. Here, we use the SHA-512 algorithm whose output length is 512 bits since its maximum input message size is (2^{128}-1) bits which suffices for our usage.

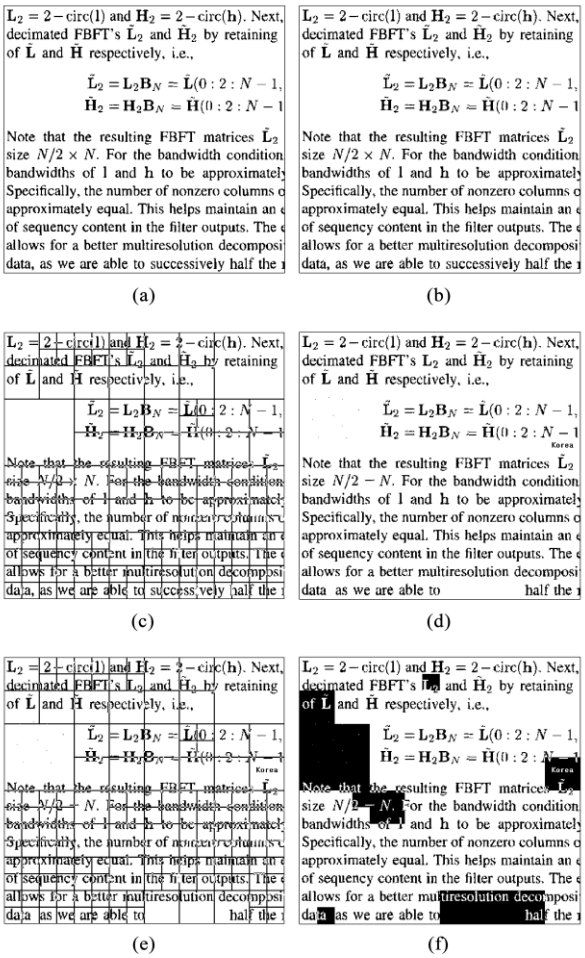


Figure 2. Experimental results: (a) an original binary image, (b) the watermarked image, (c) the result of partitioning (a), (d) the corrupted image, (e) the result of partitioning (d), and (f) the result of localizing tampered regions.

The experimental results on a sample image are shown in Figure 2. Figure 2(a) shows an original binary image. Figure 2(b) shows the watermarked image. As shown, the visual degradation of the watermarked image is almost negligible. Figure 2(c) displays the result of partitioning in the spatial domain. Figure 2(d) shows the corrupted image modified by erasing, replacing, and adding. The result of block partitioning of Figure 2(d) is shown in Figure 2(e). The result of localizing tampered regions is shown in Figure 2(f). As shown, all instances of tampering, including adding and erasing a few dots and erasing sequential blocks, have been detected.

In order to compare the localization accuracy for images of different resolutions, we also show the results of block partitioning for images of different resolutions. Detailed information about the sample images and the results of block partitioning with threshold value T_b=15 are listed in Table 1. It is found that the size of partitioned blocks ranges from 32×32 to 512×512. The number of partitioned blocks for each size is listed in Table 1.

Table 1. Experimental Results of Block Partitioning, Capacity, and Distortion

Sample image		Figure 2(a)	Image1024	Image2048
Scanning resolution (dpi)		300	300	600
Image size (pixels)		512×512	1024×1024	2048×2048
Number of black pixels		32,602	168,588	760,321
Number of partitioned blocks (T_b=15)	32×32	128	96	12
	64×64	24	156	641
	128×128	2	19	31
	256×256	0	0	12
	512×512	0	0	1
Capacity (bits)	Proposed	4,722	13,777	31,048
	Wu and Liu	2,603	8,630	20,292
Distortion (ELD_{total}/N)	Proposed	2.5450	2.4897	2.4095
	Wu and Liu	2.5656	2.5673	2.5571

Comparisons of our method with those proposed by [4] and [5] are made. References [4] and [5] basically use the 3×3 block pattern-matching algorithm, proposed by Wu and Liu [2], for finding flippable pixels in the spatial domain, and they partition an image into blocks of fixed size. In their recommended implementations, the localization accuracy achieved by [4] and [5] is 128×128 and 33×33, respectively. In our method, the size of partitioned blocks is variable and dependent on the threshold value T_b. Our experiment with threshold value T_b=10–20 shows that the smallest size of partitioned blocks is 32×32 which is smaller than in the compared methods. Furthermore, by reducing the threshold value T_b, we can achieve more finely partitioned blocks.

Comparisons of capacity (number of flippable pixels) and visual distortion are also performed. We use the edge line distortion measure (ELDM) denoted by "ELD_{total}/N" in [8] for the distortion comparison. Experimental results on various test images show that the "ELD_{total}/N" value almost linearly approximates the subjective rankings with a score of 1 for the least distortion and 4 for the most distortion and visual degradations begin to be perceived by humans around a score of 2.75 [8]. More details on ELDM are preferably referred from [8]. Wu and Liu's method for finding flippable pixels divides an image into non-overlapping 3×3 blocks and decides the center pixel's flippability score based on the lookup table reflecting the smoothness and the connectivity of the block. The flippability score of each block is assigned as one of the values: 0, 0.01, 0.125, 0.25, 0.375, and 0.625. We have counted the blocks whose flippable score is higher than 0.01. The number of flippable pixels of Wu and Liu's method and the number of flippable BWT coefficients of our method for each sample image are listed in Table 1 as capacity values. The calculated "ELD_{total}/N" values of Wu and Liu's method and our method for each sample image are listed in Table 1 as distortion values. When we calculate these distortion values, we have flipped a full capacity number of bits for each method. It is shown that our method has better capacity with less visual distortion for all sample images. It is also noted that all the distortion values in Table 1 are around 2.5 (less than 2.75), which means that all the modified images with a full capacity number of bits for each method are almost imperceptible to the human eye.

References

[1] Y. Lee, H. Kim and Y. Park, "A New Data Hiding Scheme for Binary Image Authentication with Small Image Distortion", Info. Sci. vol. 179, pp.3866-3884 **(2009)**.

[2] M. Wu and B. Liu, "Data Hiding in Binary Image for Authentication and Annotation", IEEE Trans. Mul. vol. 6, pp.528-538 **(2006)**.

[3] H. Yang, A. C. Kot and S. Rahardja, "Orthogonal Data Embedding for Binary Images in Morphological Transform Domain – a High–Capacity Approach", IEEE Trans. Mul., vol. 10, pp. 339-351 **(2008)**.

[4] H. Y. Kim and R. L. Queiroz, "Alteration–Locating Authentication Watermarking for Binary Images", In: Proc. Int. Workshop on Digital Watermarking, pp. 125-136. Seoul **(2004)**.

[5] C. H. Tzeng and W. H. Tsai, "A New Approach to Authentication of Binary Images for Multimedia Communication with Distortion Reduction and Security Enhancement", IEEE Commun. Lett. vol 7, pp. 443-445 **(2003)**.

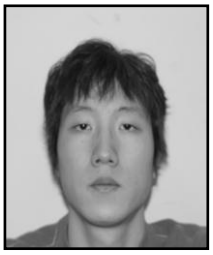
[6] H. Yang and A. C. Kot, "Binary Image Authentication with Tampering Localization by Embedding Cryptographic Signature and Block Identifier", IEEE Sig. Proc. Lett. vol. 13, pp.741-744 **(2006)**.

[7] M. Holliman and N. Memon, "Counterfeiting Attacks on Oblivious Block–Wise Independent Invisible Watermarking Schemes", IEEE Trans. Image Proc. vol. 9, pp.432-441 **(2000)**.

[8] J. Cheng and A. C. Kot, "Objective Distortion Measure for Binary Text Image Based on Edge Line Segment Similarity", IEEE Trans. Image Proc. vol. 16, pp.1691-1695 **(2007)**.

Authors

Oh-Jin Kwon received a B.S. degree from Hanyang University, Seoul, Korea, in 1984, an M.S. degree from the University of Southern California, Los Angeles, 1991, and a Ph.D. degree from the University of Maryland, College Park, in 1994, all in electrical engineering. From 1984 to 1989, he was a research staff member at the Agency for Defense Development, Korea, and from 1995 to 1999, he was the head of the Media Laboratory at Samsung SDS, Seoul. Since 1999, he has been a faculty member at Sejong University, Seoul, Korea, where he is currently an Associate Professor. His research interests are image and video coding, watermarking, analyzing, and processing.



Won-Gyu Kim received the B.S. degree in information & communication engineering and the M.S. degree in electronics engineering, both from Sejong University, Seoul, Korea, in 2007 and 2009, respectively. Currently, he is a research assistant at the Information and Telecommunication Research Institute in Sejong University. His research interests are image and video watermarking and analyzing.

Youngseop Kim received the M.S in Computer Engineering from the University of Southern California in 1991, and the Ph.D. in Electronic Systems from Rensselaer Polytechnic Institute in 2001. He was a manager at Samsung SDI until 2003. Currently he is an Associate Professor at Dankook University in Korea. He is the resolution member and the Editor of JPsearch part 2 in JPEG, the co-Chair of JPXML in JPEG, and Head of Director (HOD) of Korea. He is also Editor-in-chief of the Korea Semiconductor and Technology Society. His research interests are in the areas of image/video compression, pattern recognition, communications, stereoscopic codecs, and augment reality.