A Novel Data Hiding Method for Two-Color Images*

Gang Pan, Yijun Wu, and Zhaohui Wu

Department of Computer Science and Engineering
Zhejiang University, Hangzhou, 310027, P. R. China
gpan, wzh@cs.zju.edu.cn

Abstract. Binary images have only two colors, which makes the embedding of invisible data difficult. In this paper, we propose a new data hiding method that can hide a moderate amount of data in a host binary image, such as binary cartoon images, scanned texts, signatures, without introducing noticeable artifacts. The proposed method employs subblock pattern classification to maintain visualization effect and mechanics of multilevel supblock to improve the capacity. Extracting of the hidden data does not require the knowledge of the original image. The experiments demonstrate that the proposed method can provide excellent perceptual quality of the marked image. The potential applications include invisible annotation, alteration detection and covert communication.

1 Introduction

With the huge success of the Internet, digitization of various kinds of media is getting wider popularity for the transmission, wide distribution and storage. The advantages of digital media include convenient transmission, effortless access, lossless copy, facile edit and reliable storage. However, they also introduce a new set of challenging problems regarding security, that are not able to be achieved only by encryption. The problems have generated a flurry of recent research activities in the area of digital watermark and data hiding.

The study on digital watermark and data hiding has received great achievements over last several years. Many different methods have been proposed for still image. They can be classified into two categories based on the casting domain: 1) luminance intensity in the spatial domain [1,2], for instance, LSB (Least Significant Bit) approach and Patchwork method, and 2) transform coefficient magnitude in the frequency domain, which modify frequency coefficients after applying a proper transform [3,4,5,6], e.g. DWT, DCT, FFT, etc. The new technique has a variety of potential applications involving digital media, including copyright protection, annotation, covert communication, and alteration detection.

* This work supported by Zhejiang Provincial Natural Science Foundation of China under Grant 699035.

S. Qing, T. Okamoto, and J. Zhou (Eds.): ICICS 2001, LNCS 2229, pp. 261–270, 2001.

However, objectives of most of these methods are color images and grayscale images and will fail to apply to two-color images. As an important class of images, digital binary images are widely used in Internet. There is a significant difference between binary image and other natural images. That is, the binary images are only two colors images without complicated color and texture variation, and change a pixel can be easily detected. This peculiar characteristic makes it more difficult to embed invisible digital information in them. The only solutions known to us dealing with binary image are [7,8]. Wu[7] presented a data hiding scheme for binary images for the first time. It partitions the image into blocks, then tries to embed as many as one data bit in each block via AND operation with a secret key matrix whose size is the same as the blocks. However, the perceived quality is poor because it does not take into account the visualization effect. It introduces many isolated points near the boundaries, which causing noticeable artifacts. An improved method for higher security and capacity proposed by [8], but the visibility of marked image is still a problem, even poorer than [7] in some cases, since it may introduce isolated point in any location of the host image.

This paper addresses to the visibility of marked image. We propose a multi-level supblock based data hiding method that can hide a moderate amount of data in the binary images, e.g. binary cartoon, scanned text, and signatures. The hidden data can be extracted without the original host image. The proposed method greatly outperforms the previous approaches in capability of transparency. The potential applications include invisible annotation, changes detection and covert communication.

The paper is organized as follows. The description of the proposed scheme is presented in Sect. 2. Experimental results and analysis are given in Sect. 3. Finally, concluding remarks are provided in Sect. 4.

2 Description of the Proposed Method

As we mentioned, it is more difficult to embed a piece of critical data in binary images under constraint of the visibility, because there are only two elements in the pixel-value space of in binary images. How to maintain good perceptual quality comes to be the major problem. We have observed that whether a revised pixel is noticeable strongly depends on its neighbors. For instance, if a pixel in homochromous region changes to another color, the difference will be quite noticeable. And it will be hard to detect comparatively if some neighbors of the pixel have the same color as the pixel after modification. Therefore, during embedding, what kinds of pixels to alter should be determined according to conditions of its neighbors.

Our method is motivated by the above observation. Its fundamental thought is summarized below. Firstly it partitions the host image into many blocks, then classify these blocks into different level. The high rank (or level number) represents good visual performance. Given a piece of critical data, our scheme will manage to embed the data in the blocks with highest rank. The "0" and

"1" are respectively represented by a pair of blocks with only difference of the central pixel.

To improve the capacity and insensibility, we introduce the concept of "supblock", which is composed of several base blocks overlaid with each other. Correspondingly, those base blocks with smaller size are called "subblock". With mechanics of supblock, we can examine a local region by means of different combinations of pixels for more subblock patterns.

The block diagram of embedding and extracting is shown in Fig. 1.

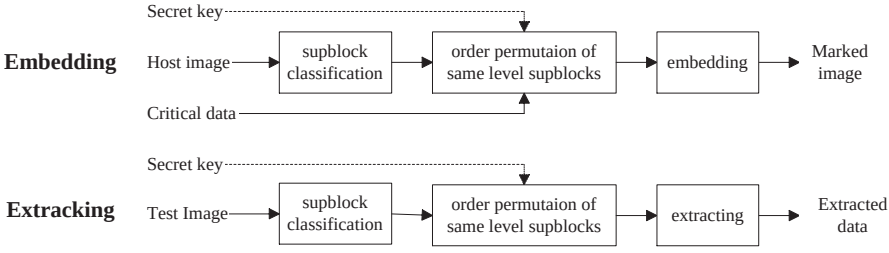


Fig. 1. The block diagram of embedding and extracting procedures

2.1 Subblock Classification

The objective of subblock classification is to overcome the perceptual quality reduction by pixel modification. In our approach, each subblock is connected with a level number (*rank*) according to its pattern, indicating influence on visibility by assumed change of the central pixel in the subblock. In other words, supposed that the central pixel in a subblock is changed, we consider the variation in connectivity and smoothness and investigate how the visibility effect reduces by such a change, then the rank is determined. The higher rank implies that change of central pixel in subblock reduces visual quality less and should has a higher priority for embedding.

We take the 3-by-3 subblock as a sample, shown in Fig. 2. Whatever is the central pixel in subblocks, change of central pixel of subblock in Fig. 2(b) will obviously got less attention than that in Fig. 2(a). So that pattern of subblock in Fig. 2(b) has higher rank than Fig. 2(a). In this manner, we classify all the patterns into different level. There are 256 (2^8) subblock patterns totally, exclusive of the central pixel. We denote the collection of patterns of rank n by A_n.

There are several approaches for embedding one bit data h in a subblock $\boldsymbol{B}$. We listed two of them below, where the central pixel of the subblock is denoted by c.

a) Let $c = h$.

b) Let $c = (SUM(\boldsymbol{B} \oplus \boldsymbol{K}) + h) \bmod 2$.

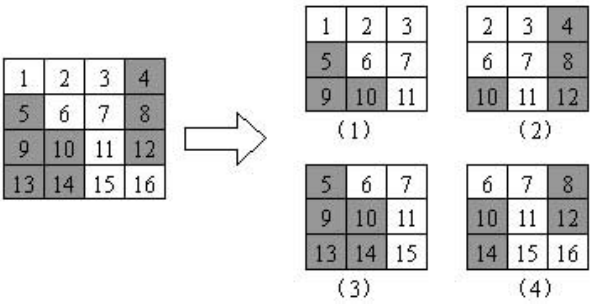
(a) (b)

Fig. 2. Two patterns of 3-by-3 subblock with different rank.

Where $\boldsymbol{K}$ is a given matrix whose size is the same as the subblock $\boldsymbol{B}$, "$\oplus$" is the bitwise exclusive-OR operator. And $SUM(\boldsymbol{X})$ is the sum of all elements in matrix $\boldsymbol{X}$.

2.2 Mechanics of Supblock

If we directly employ the simple scheme of partition of the image followed by examining level of subblock, subblock number of a certain level will be quite restricted, since the pixels of a pattern are often scattered in different subblocks. However, the simple scheme has never taken into account these patterns. To take advantage of the patterns whose pixels are distributed in multi-subblock, it is necessary to mend the simple scheme. In this paper we employ the mechanics of supblock. The supblock is larger than subblock, hence it contains more than one subblock, e.g. a 4-by-4 supblock contains four 3-by-3 subblocks, shown in Fig. 3, and the labeled numbers represent the pixel correspondence between the subblock and the supblock.



4×4 supblock 3×3 subblock

Fig. 3. Illustration of supblock mechanics for top level.

Suppose that we have a 4-by-4 supblock S, we denote the i^{th} subblock of the supblock by $B_i(S)$ $(i = 1, 2, 3, 4)$. The *subblock embeddable function* $E_n(S, k)$ of rank n is recursively defined as follows:

i) For $k = 1$,

$$E_n(S,1) = \begin{cases} 1, & \text{if } B_1(S) \in A_n \\ 0, & \text{else} \end{cases} \tag{1}$$

ii) For $k > 1$,

$$E_n(S,k) = \begin{cases} 1, & \text{if } B_k(S) \in A_n \text{ and } \forall i < k, m \geq n, E_m(S',i) \neq 1 \\ 0, & \text{else} \end{cases} \tag{2}$$

Where S' is the new supblock, supposed that central pixel in $B_k(S)$ is always changed.

The subblock embeddable function $E_n(S,k)$ depicts whether the k^{th} subblock of S is suitable for embedding at rank n. $E_n(S,k)$=1 or 0 respectively means "yes" or "no". The subblock $B_k(S)$ is *n-level-embeddable* if $E_n(S,k) = 1$.

There may be more than one n-level-embeddable subblock in a supblock. For visibility effect, only one of them is picked for embedding. We define the *supblock embeddable indicator function* $\xi(S,n)$ as

$$\xi(S,n) = \begin{cases} \inf\{\, k | E_n(S,k) = 1\}, & \text{if } \sum_k E_n(S,k) \geq 1 \\ 0 & \text{else} \end{cases} \tag{3}$$

The function $\xi(S,n)$ describes that at rank n,whether the supblock is embeddable, and if so, which subblock is selected for embedding. The supblock S is *n-level-embeddable* if $\xi(S,n) > 0$.

We consider the sample shown in Fig. 3 for top level. In addition, we should define A_N, the pattern set of top level N. Here we let it be the collection of patterns similar to the pattern of Fig. 2(b). For the supblock shown in Fig. 3, we can obtain that $E_N(S,k) = 0$ for $k = 1, 2, 4$ and $E_N(S,3) = 1$,further,$\xi(S,N) = 3$, meaning the 3^{rd} subblock (Fig. 3(3)) is suggested for embedding.

The efficiency of supblock mechanics will be demonstrated in Sect. 3.

2.3 Embedding and Extracting

During the embedding procedure, the method will try to embed the critical data in the supblocks with high rank. At the same level, all the embeddable supblocks are permuted randomly based on a secret key before the embedding. The permutation has two advantages. First, it avoids selected supblocks cluster. Second, it also improves security. To summarize, the full procedure of embedding is as follows:

1. Partition the binary image into supblocks with the same size, e.g. 4×4.
2. Set *Level*=MAXRANK
3. While *Level* > 0 and NOT finishing all critical data
   a) For each supblock S_i that has not been marked, compute $\xi(S_i, Level)$
   b) Perform the random permutation of the supblocks with $\xi(S_i, Level) > 0$
   c) Embedding the critical data in the permutated supblocks, one supblock for one bit

d) If finishing the embedding, quit.
e) $Level \leftarrow Level - 1$

The embedding of critical data is inverse procedure. It is not difficult to deduce the procedure of extraction. It is ignored here.

3 Experimental Results and Analysis

The experiments summarized below were all conducted with 4-by-4 supblock size and 3-by-3 subblock size, unless otherwise specified. Our experimental results presented are composed of three parts. The first is tests on our method. The comparison with other schemes is performed in the second part. And the last part is a demonstration on application of tampering detection.

3.1 Tests on Our Methods

We have conducted many tests on our method. It really achieves the excellent performance because of the unnoticed changes after embedding. Meanwhile, it has the moderate capacity. Some results are presented in Fig. 4. To embed 200 bits in the host image with size 166×198, 105 pixels are changed, but the marked image differs very little from the original host image. The difference map between Fig. 4(a) and Fig. 4(b) is shown in Fig. 4(c).

Fig. 4. Data hiding by our method. (a) the original host image with size 166×198, (b) the marked image after embedding 200 bits, (c) the difference map, indicated by black pixel (totally 105 pixels).

As we mentioned earlier, the supblock mechanics can improves number of subblocks efficiently. We have performed the test on the supblock mechanics using 100 binary images of different sizes and different content. Some of the

statistical data are shown in Fig. 5. It indicates that after carrying out the supblock mechanics, it achieves an increase of number of the embeddable subblocks with 20%–45% percent at different level, compared with using 3-by-3 subblocks directly. Figure 5 shows the comparison of four of the highest ranks.

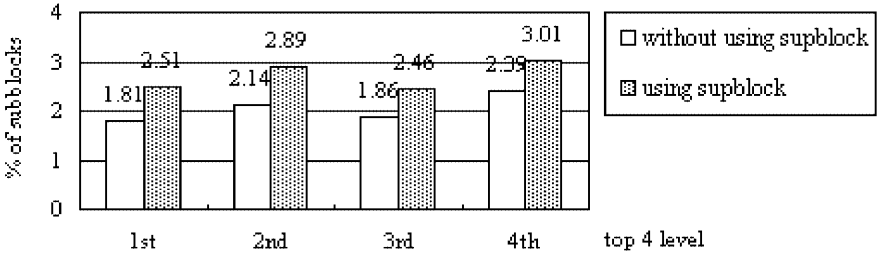


Fig. 5. Efficiency of supblock mechanics.

3.2 Comparison with Other Methods

To compare with other methods, we have implemented the WU98 scheme [7] and the PAN00 scheme [8]. We use the same blocks size as implementation in our scheme (that is 4-by-4) for WU98 scheme, and use 16×16 of block size for PAN00 scheme since its advantage is exposed only when the block size is large. For fair comparison of images' perceptual quality, size and content of the critical data are same for all of the three methods.

One of the results is shown in Fig. 6. The host image size is 293-by-384 and the amount of the critical data is 480 bits. Obviously, PAN00 scheme introduces image-wide "visual noise", and WU98 also introduces "visual noise" near the boundaries.

In both WU98 scheme and PAN00 scheme, each pixel in a block is changeable. Without considering perceptual loss, they always introduce some isolated pixels, whose neighbors' color is all opposite to its color. It reduces the image quality seriously. Contrarily, our approach achieves the superior performance compared with WU98 and PAN00 schemes.

3.3 Tampering Detection

Because of ease to edit digital images, the authentication of these documents is becoming a great concern during recent years. The proposed scheme can be used for the purpose of tampering detection.

Figure 7 shows a sample of the scanned text for alteration detection. The host image is the first page of the paper by Fabien A. P. Petitcolas etc. After embedding another binary image with 84×125 in the host image, we remove a

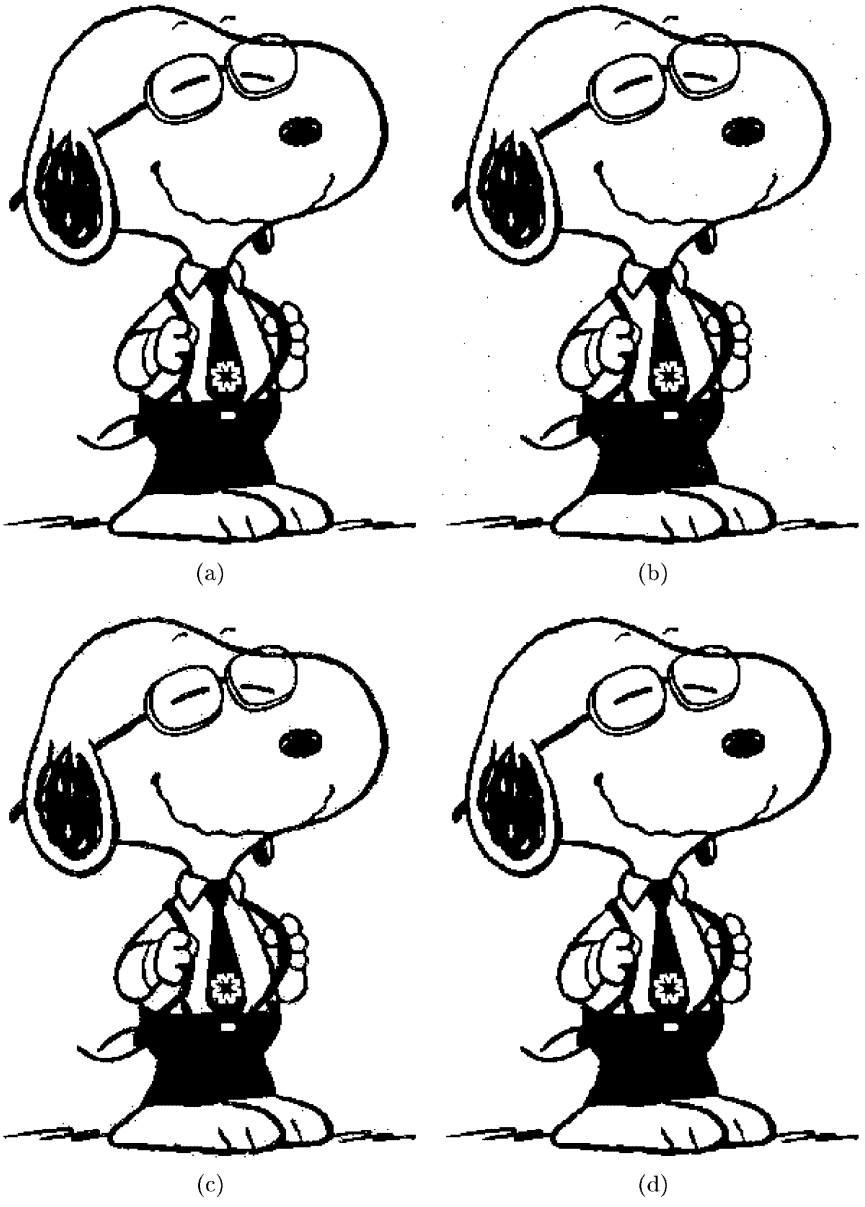


Fig. 6. Comparison with other methods. The amount of hidden data is 480 bits. (a) the original host image with size 293×384, (b) the marked image by PAN00 scheme with block size 16×16, (c) the marked image by WU98 scheme with block size 4×4, (d) the marked image by our method with supblock size 4×4 and subblock size 3×3.

radix point in the marked image. Figure 7(c) is the extracted data after alteration. Obviously, the extracted binary image is significantly different from the original binary image of panda logo.

Information Hiding—A Survey

FABIEN A. P. PETITCOLAS, ROSS J. ANDERSON, AND MARKUS G. KUHN

Information-hiding techniques have recently become important in a number of application areas. Digital audio, video, and pictures are increasingly furnished with distinguishing but imperceptible marks, which may contain a hidden copyright notice or serial number or even help to prevent unauthorized copying directly. Military communications systems make increasing use of traffic security techniques which, rather than merely concealing the content of a message using encryption, seek to conceal its sender, its receiver, or its very existence. Similar techniques are used in some mobile phone systems and schemes proposed for digital elections. Criminals try to use whatever traffic security properties are provided intentionally or otherwise in the available communications systems, and police forces try to restrict their use. However, many of the techniques proposed in this young and rapidly evolving field can trace their history back to antiquity, and many of them are surprisingly easy to circumvent. In this article, we try to give an overview of the field, of what we know, what works, what does not, and what are the interesting topics for research.

Keywords—Copyright marking, information hiding, steganography.

I. Introduction

It is often thought that communications may be secured by encrypting the traffic, but this has rarely been adequate in practice. Æneas the Tactician, and other classical writers, concentrated on methods for hiding messages rather than for enciphering them [1], although modern cryptographic techniques started to develop during the Renaissance, we find in 1641 that Wilkins still preferred hiding over ciphering [2, ch. IX, p. 67] because it arouses less suspicion. This preference persists in many operational contexts to this day. For example, an encrypted e-mail message between a known drug dealer and somebody not yet under suspicion, or between an employee of a defence contractor and the embassy of a hostile power, has obvious implications.

So the study of communications security includes not just encryption but also traffic security, whose essence lies in hiding information. This discipline includes such technologies as spread spectrum radio, which is widely used in tactical military systems to prevent transmitters

Manuscript received February 2, 1998; revised December 1, 1998. The work of F. A. P. Petitcolas was supported by Intel Corporation under the grant "Robustness of Information Hiding Systems." The work of M. G. Kuhn was supported by the European Commission under a Marie Curie Training Grant.

The authors are with the University of Cambridge Computer Laboratory, Security Group, Cambridge CB2 3QG U.K.

Publisher Item Identifier S 0018-9219(99)04946-4.

Table 1
Number of Publications on Digital Watermarking During the Past Few Years According to INSPEC, January 1999 (Courtesy of J.-L. Dugelay [5])

Year	1992	1993	1994	1995	1996	1997	1998
Publications	2	2	4	13	20	64	103

being located; temporary mobile subscriber identifiers, used in digital phones to provide users with some measure of location privacy; and anonymous remailers, which conceal the identity of the sender of an e-mail message [3].

An important subdiscipline of information hiding is steganography. While cryptography is about protecting the content of messages, steganography is about concealing their very existence. It comes from Greek roots (στεγανός, γράφειν), literally means "covered writing" [151], and it is usually interpreted to mean hiding information in other information. Examples include sending a message to a spy by marking certain letters in a newspaper using invisible ink, and adding subperceptible echo at certain places in an audio recording.

Until recently, information-hiding techniques received much less attention from the research community and from industry than cryptography, but this is changing rapidly (Table 1), and the first academic conference on the subject was organized in 1996 [4]. The main driving force is concern over copyright, as audio, video, and other works become available in digital form, the ease with which perfect copies can be made may lead to large-scale unauthorized copying, and this is of great concern to the music, film, book, and software publishing industries. There has been significant recent research into digital "watermarks" (hidden copyright messages) and "fingerprints" (hidden serial numbers); the idea is that the latter can help to identify copyright violators, and the former to prosecute them.

In another development, the DVD consortium has called for proposals for a copyright marking scheme to enforce serial copy management. The idea is that DVD players available to consumers would allow unlimited copying of home videos and time-shifted viewing of TV programs but could not easily be abused for commercial piracy. The proposal is that home videos would be unmarked, TV broadcasts marked "copy once only," and commercial videos marked "never copy"; compliant consumer equipment would act on these marks in the obvious way [6], [7].

0018–9219/99$10.00 © 1999 IEEE

1062 PROCEEDINGS OF THE IEEE, VOL. 87, NO. 7, JULY 1999

(a) (b)

(c) (d)

Fig. 7. Alteration detection for scanned text. (a) the marked image with size 1048×1380, after embedding 10,500 bits, (b) removing the radix point in the marked image, (c) the critical data (panda logo with size 84×125), (d) the extracted critical data after alteration.

4 Conclusion

This paper proposes a new data hiding algorithm for two-color images. The main idea is to use subblock pattern classification and supblock mechanics to select the pixels with least visual quality reduction for embedding. Analysis and experimental results both show that the proposed method can provide the superior performance and greatly outperforms the previous approaches in visibility transparency. It can applied to tampering detection, invisible annotation, and convert communication.

References

1. R.Z. van Schyndel, A.Z. Tirkel, and C.F.Osborne: A Digital Watermark. Proc. IEEE Int. Conf. Image Processing, vol.2, p86–90, 1994
2. W.Bender,N.Morimoto and D.Gruhl: Techniques for DataHiding,IBM System Journal. vol.25, p313–335, 1996
3. I.J.Cox, J.Kilian, T.Leighton, T.Shamoon: Secure spread Spectrum Watermarking for Images,Audio and Video. Proc. IEEE Int. Conf. On Image Processing, Lausanne, Switzerland, Sep. 1996
4. D.Kundur,D.Hatzinakos: Digital Watermarking Based on Multiresolution Wavelet Data Fusion. Proc. IEEE Special Issue on Intelligent Signal Processing,1997
5. C.Podilchuk, W.Zeng: Image Adaptive Watermarking Using Visual Models. IEEE Journal on Selected Areas in Comm., Vol.16, No.4, 1998
6. A.Piva, M.Barni, F.Bartolini, and V.Capellini: DCT-based watermark recovering without resorting to the uncorrupted original image. in Proc. IEEE ICIP'97, Santa Barbara, CA, Oct. 1997, vol.1, pp.520-523.
7. M.Y. Wu, J.H. Lee: A Novel Data Embedding Method for Two-Color Facsimile Images. Proc. Int. Symposium on Multimedia Information Processing, Taiwan, Dec. 1998
8. H.K. Pan, Y.Y. Chen, and Y.C. Tseng: A Secure Data Hiding Scheme for Two-Color Images. Proc. IEEE Symposium on Computers and Communications (ISCC 2000), France, July 3–6 2000, p750–755