가변 길이 자료 은닉이 가능한 이미지 스테가노그래픽 방법 연구

Image Steganographic Method using Variable Length for Data Embedding

정 기 현*
Jung, Ki-Hyun

ABSTRACT

Wu and Tsai's pixel-value differencing method and Chang and Tseng's side-match method are based on the theory that the number of bits which can be embedded is determined by the degree of the pixel's smoothness, or its proximity to the edge of the image. If pixels are located in the edge area, they may tolerate larger changes than those in smooth areas. However, both methods are subject to the fall off the boundary problem(FOBP). This study proposes a new scheme that can solve the FOBP. The experimental results demonstrate that the proposed method resolves the problem, and achieves a higher image quality index value than other methods.

주요기술용어(주제어) : Steganography(스테가노그래피), Information Hiding(정보은닉), Run Length Encoding(반복 길이 부호화), Data Hiding(자료은닉)

1. Introduction

Steganography is an ancient technique of hiding information[1]. As the transmission of digital media via the Internet becomes more and more widespread, steganographic techniques for data hiding in digital media have become a popular subject for industrial applications. Two important uses are to provide proof of copyright, and assurance of content integrity. Code breaking in a steganographic system is different from that of a cryptographic system. In cryptography, the system is broken when the attacker can read the secret message, whereas breaking a steganographic system has two stages : the attacker can firstly detect that steganography has been used and secondly, the attacker is able to read the embedded message[2]. Hiding the message into cover objects successfully is therefore critical to a steganographic system.

The key to successful data hiding is the finding of holes that are not suitable for exploitation by compression algorithms. A further challenge is to fill these holes with data in a way that remains

† 2008년 4월 11일 접수~2008년 5월 16일 게재승인
* 영진전문대학 컴퓨터정보계열
주저자 이메일 : kingjung@paran.com

unaffected by a large class of host signal transformations[3]. A trade-off exists between the quantity of embedded data and the degree of immunity to host signal modification. By constraining the degree of host signal degradation a data hiding method can operate with either a high rate of embedded data, or a high resistance to modification, but not both. The quantity of embedded data and the degree of host signal modification vary from application to application.

Data hiding techniques can have access to any pixel or block of pixels at random. Some techniques are more suited to dealing with small amounts of data, while others are suited to large amounts. Some techniques are highly resistant to geometric modifications, while others are more resistant to non-geometric modifications, like as filtering. An image quality index and bit capacity are employed to evaluate the proposed method, in which the secret data are stored behind the host image in a way imperceptible to human vision.

This paper proposes a scheme that can provide a superior secret data embedding mechanism that is imperceptible to human vision. Moreover, the new method solves the fall off the boundary problem(FOBP).

This paper is organized as follows. Section 2 reviews two methods reported recently relating to the proposed method and the FOBP. In Section 3, the details of our newly proposed scheme are described. In Section 4, the experimental results are presented and discussed. Finally, our conclusions are presented in Section 5.

2. Related Work

In general, if the pixels are located in edge areas they can tolerate larger changes than those in smooth areas. The range of changeable pixel value in smooth areas is small, whereas in edge areas it is large so that the stego-image maintains good perceptual quality when embedded. Wu and Tsai's pixel-value differencing method[4] and Chang and Tseng's side match method[5] are based on the premise that the number of bits which can be embedded depends on the proximity to the edge of the image or degree of smoothness. But, both these methods display the FOBP in the edge area, although it may be hidden within large amounts of data.

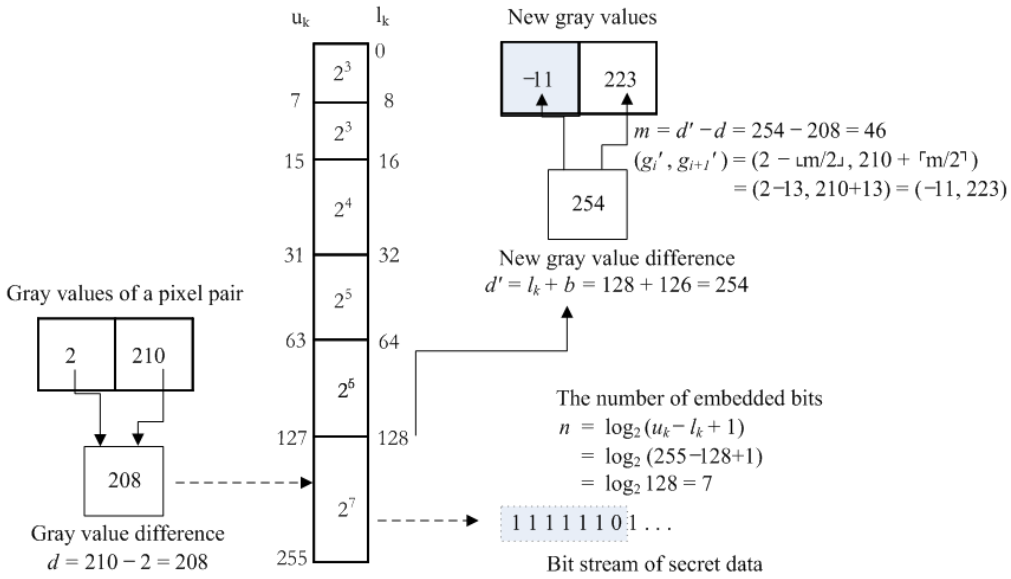
Wu and Tsai proposed a pixel-value differencing method, whereby a cover image is partitioned into non-overlapping blocks of two consecutive pixels. A difference value is calculated from the values of the two pixels in each block. Secret data are embedded into a cover image by replacing the difference values of the two-pixel blocks of the cover image with similar ones, in which bits of the embedded data are included.

Chang and Tseng employed two-sided, three-sided and four-sided side match schemes. The two-sided side match method uses the side information of the upper and left neighboring pixels in order to make estimates. The three-sided side match scheme utilizes not only the upper and left pixels, but also one of the other neighboring pixels, to the right or below. In order to make more precise estimates, the four-sided side match method uses all neighboring sides ; upper, left, right and below a given pixel, instead of only two, as in the two-sided side match scheme.

Above two methods do not use the pixel blocks which the FOBP occurs.

2.1 The FOBP Occurring with Wu and Tsai's Method

Wu and Tsai's pixel-value differencing is often referenced to embed secret data. In the proposed



[Fig. 1] An example of Wu and Tsai's steganographic method

method, the experiment was based on selecting the range widths of 8, 8, 16, 32, 64 and 128, which partition the total range of [0, 255] into [0, 7], [8, 15], [16, 31], [32, 63], [64, 127], [128, 255]. We will explain the FOBP issue as it affects range width when using Wu and Tsai's method.

Fig. 1 shows why we cannot use the pixel pair (2, 210) for embedding secret data. This pixel value belongs to the edge area, so it should be able to hide many bits of secret data. However, as we show in Fig. 3, the resulting pixel value (11, 223) is excluded because it extends over the gray-scale pixel boundary [0, 255]. Therefore, Wu and Tsai's method does not use this pixel value.

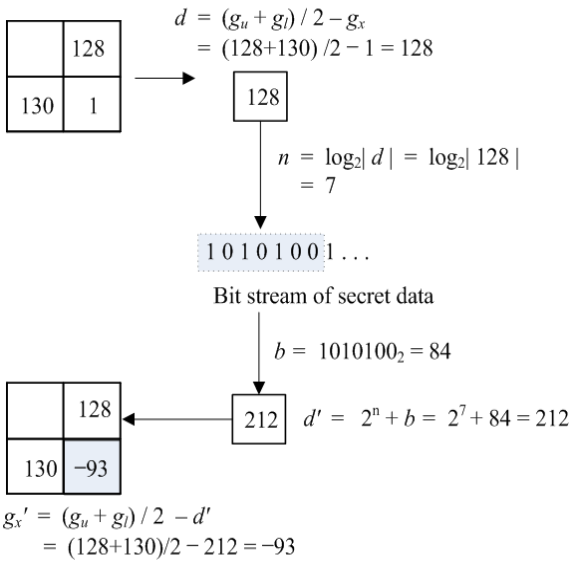
2.2 The FOBP Occurring with Chang and Tseng's Method

In this proposed method, ranges with smaller widths are created when d is close to 0, and ranges with greater widths when d is distant from 0, for the purpose of yielding better imperceptibility results.

In an earlier section we stated that we can hide more secret data in the edge areas. However, we also note that the FOBP occurs in the edge areas. This means that we cannot use the pixel value which creates this problem even though it enables us to hide larger quantities of secret data. Fig. 1 illustrates the data embedding process and the weak point. In the following discussion we focus on the FOBP.

Case 1. $d = (g_u + g_l) / 2 - g_x \rangle 1$.

Fig. 2 shows the FOBP when the d value is greater than 1 in the Chang and Tseng's two-sided side match scheme. The given pixel p_x is assumed to be 1. The two neighboring pixels have values 128 and 130. Thus the difference d = 128 and n = 7. Assume that the seven bits of the embedding data are 1010100 randomly, with the integer value b = 84. Then the new difference is d' = $2^n + b = 2^7 + 84 = 212$. Finally, the new value of pixel p_x' is g_x' = (128 + 130) / 2 - 212 = -93.



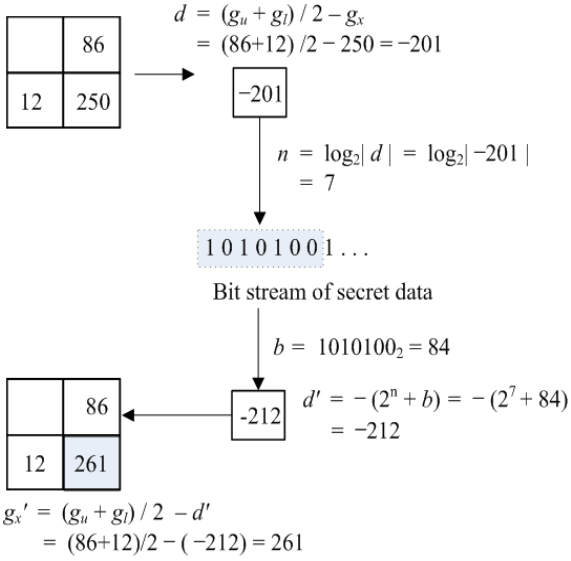
[Fig. 2] An example of the FOBP(d > 1)

Since the new value is less than 0, the pixel does not belong to the range [0, 255] in the gray-scale image. So, pixel p_x is not used for embedding data.

In Wu and Tsai's method, the calculated value, $g_x' = (g_u + g_l) / 2 - d' = (g_u + g_l) / 2 - (2^n + b) = (g_u + g_l) / 2 - (2^n + 2^n - 1)$, where we assume that secret b value is the worst value in terms of its usefulness. Therefore, if the $(g_u + g_l) / 2 - (2^{n+1} - 1)$ value is less than 0, this pixel cannot be used for secret data embedding.

Case 2. $d = (g_u + g_l) / 2 - g_x < 1$.

Fig. 3 shows the FOBP when the d value is less than 1. The given pixel p_x is assumed to be 250. The two neighboring pixels have values 86 and 12. Thus, the difference d = -201 and n = 7. Assume that the seven bits of the embedding data are 1010100 in random, the integer value b - 84. Then the new difference is $d' = -(2^n + b)$ = -(27 + 84) = -212. Finally, the new value of pixel p_x' is g_x' = (86 + 12) / 2 - (-212) = 261. Because the new value is greater than 255, the pixel falls off the boundary [0, 255] in the gray-scale image. So, the pixel p_x is not used for embedding data.



[Fig. 3] An example of the FOBP(d < 1)

In the case of $d < 1$, $g_x' = (g_u + g_l) / 2 - d' = (g_u + g_l) / 2 + (2^n + b) = (g_u + g_l) / 2 + (2^n + 2^n - 1)$, where we assume that secret b value is the worst value in terms of the ability of the pixel to hide secret data. Therefore, if $(g_u + g_l) / 2 + (2^{n+1} - 1)$ value is more than 255, the pixel is not able to hide secret data, and must be excluded from data embedding.

If we assume that the secret data bits stream is random, the existing pixel value does have the FOBP. Depending on whether we use the cover image or the secret image, the pixel value is different and suffers from the FOBP. The problem occurs also in the three-sided and four-sided side match method. The difficult thing is that although the neighboring pixel value difference is large, we can't use the pixel because of the FOBP. This creates the belief that its secret data hiding capacity is low.

3. Proposed Method

In Section 2, we described how the FOBP occurred in both Wu and Tsai's method and Chang and Tseng's method. This study proposes a new method resolving the FOBP and which therefore improves on the method of Chang and Tseng. The data embedding and extracting processes are included in this paper. Because the FOBP occurs in both methods, the proposed method can also be applied to the method of Wu and Tsai.

The method uses the side information of the upper g_u and left g_l neighboring pixels to calculate the new pixel p_x' from the original pixel p_x value. The cover data is embedded in raster-scan order except for the pixels of the first row and the first column, as treated in Chang and Tseng's method.

3.1 Data Embedding Algorithm

The proposed secret data embedding scheme is as follows :

Given an input pixel p_x with gray-scale value g_x, let g_u, g_l and g_{lu} be the gray-scale values of its upper p_u, left p_l and left-upper p_{lu} pixel, respectively.

Then a difference value d is computed as

$$d = (g_u + g_l + g_{lu}) / 3 - g_x \quad (1)$$

The boundary value b is calculated by

$$b = \log_2 | d | \quad (2)$$

Embedding a bit count value, say n, is defined. $n = b - 1$ is applied when the value $T = (g_u + g_l + g_{lu}) / 3 - (2^{b+1} - 1)$ is lower than zero or greater than 255. Here, we select the best value m by experiment, where $K = 2^m - 1$ ($3 \le m \le 7$).

$$n = \begin{cases} 3 & \text{if } 0 \le T < K, \\ b + 1 & \text{if } K \le T < 2^8 - 1, \\ b - 1 & \text{otherwise} \end{cases} \quad (3)$$

A sub-stream with n bits in the embedding data is selected and is converted to integer value i. Then a new difference d' is computed as

$$d' = \begin{cases} 2^n + i & \text{for d} \ge 0, \\ -(2^n + i) & \text{for d} < 0 \end{cases} \quad (4)$$

Finally, the new value of the pixel p_x is defined to be

$$g_x' = (g_u + g_l + g_{lu}) / 3 - d' \quad (5)$$

3.2 Data Extracting Algorithm

The extraction of embedded data is directive. As with the embedding process, the data is extracted in raster-scan order like as Chang and Tseng's method. Given an input pixel p_x' with gray-scale value g_x' for the stego-image, let g_u', g_l' and g_{lu}' be the gray-scale values of its upper p_u', left p_l' and left-upper p_{lu}' pixel, respectively. Then a difference value d^* is computed as

$$d^* = (g_u' + g_l' + g_{lu}') / 3 - g_x' \quad (6)$$

The boundary value b' is calculated by

$$b' = \log_2 | d^* | \quad (7)$$

The number of bits n can be derived from Eq. (8), where $T' = (g_u' + g_l' + g_{lu}') / 3 - (2^{b'+1} - 1)$.

$$n = \begin{cases} 3 & \text{if } 0 \le T' < K, \\ b' + 1 & \text{if } K \le T' < 2^8 - 1, \\ b' - 1 & \text{otherwise} \end{cases} \quad (8)$$

Finally, the embedded value i is extracted using the following Eq. (9).

$$i = \begin{cases} d^* - 2^n & \text{for } d^* \geq 0, \\ -(d^* + 2^n) & \text{for } d^* < 0 \end{cases} \quad (9)$$

The value i is integer, so the corresponding binary value is converted finally.

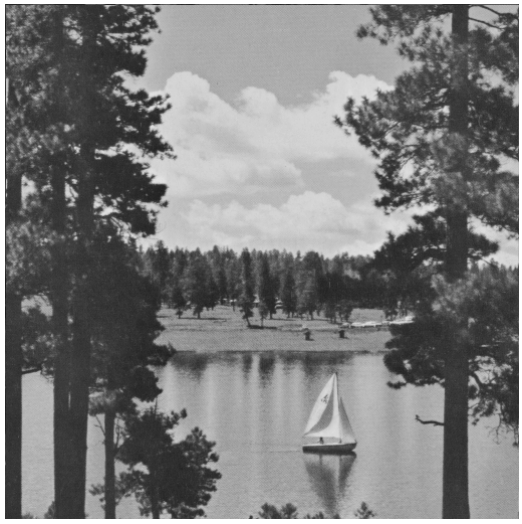
4. Experimental Results

A universal image quality index was adopted to measure any image distortion[10] and the capacity for embedding secret data. Because the FOBP arises in both, as mentioned already, the proposed method is compared with Wu and Tsai's pixel-value differencing method and Chang and Tseng's two-sided side match scheme.

In the experimental results, two 512×512 gray-scale images as shown in Fig. 4 were used as cover images and four 256×256 secret images as shown in Fig. 5 were also selected for comparison.

Tables 1 and 2 show the results of detailed comparisons of two methods in terms of quality index value and bit capacity, using the two cover

(a) House



(b) Boat

[Fig. 4] The two cover images

(a) Lena

(b) Pentagon

(c) Airplane

(d) Fishing boat

[Fig. 5] The four secret images

[Table 1] The results of embedding method using House as the cover image

Secret Image	Two-sided Side Match		Proposed Method (2^5 - 1)	
	Capacity(bit)	Quality Index	Capacity(bit)	Quality Index
Lena	571,716	0.888782	432,940	0.944540
Pentagon	571,716	0.888926	432,940	0.943862
Airplane	571,716	0.887258	432,940	0.941392
Fishing boat	571,716	0.889273	432,940	0.945240

[Table 2] The results of embedding methods using Boat as the cover image

Secret Image	Two-sided Side Match		Proposed Method (2^5 - 1)	
	Capacity(bit)	Quality Index	Capacity(bit)	Quality Index
Lena	566,648	0.956386	453,402	0.965278
Pentagon	566,648	0.955567	453,402	0.963945
Airplane	566,648	0.952822	453,402	0.961564
Fishing boat	566,648	0.957597	453,402	0.965463

images. The two-sided side match method is compared with the proposed method. Results show that the proposed method has a higher quality index value compared with the other method. The bit capacity is small because our proposed method references the neighboring three pixels which can provide an image more similar to the cover image.

In the experiment most of the distortions are found on the edge areas of the images. Therefore, such distortions will be less noticeable because changes in the edge parts of images are generally less obvious to the human eyes.

5. Conclusions

We proposed a novel data hiding method that resolved the fall off the boundary problem(FOBP) that occued in both Wu and Tsai's pixel-differencing method and Chang and Tseng's side match method. These earlier methods were both based on the premise that the number of bits which could be embedded was determined by the degree of pixel smoothness or proximity to the image edge. If the pixel was located in the edge area it could tolerate larger changes than those in smooth areas. However, pixels in the edge areas were subject to the FOBP, and hence could not be used to embed secret data, even though they were more capable of holding secret data bits than those in smooth areas. In this paper, we proposed a new method that could both solve the FOBP and provided a good quality index value that was imperceptible to the human visual system. Moreover, there was no need to reference the cover image when extracting the embedded data from the stego-image. Our experimental results have shown that the proposed scheme provides an improved way to hide secret data.

References

[1] Neil F. Johnson, Sushil Jajodia, Exploring Steganography : Seeing the Unseen, Computer Practices, 26~34, 1998.

[2] J. Zollner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke, G. Wolf, Modeling the security of steganographic systems, 2nd Workshop on Information Hiding, 345~355, 1998.

[3] W. Bender, N. Morimoto, A. Lu, Techniques for data hiding, IBM Systems Journal, 35(3/4), 313~336, 1996.

[4] D. C. Wu, W. H. Tsai, A steganographic method for images by pixel-differencing, Pattern Recognition Letters 24(9-10), 1613~1626, 2003.

[5] C. C. Chang, H. W. Tseng, A steganographic method for digital images using side-match, Pattern Recognition Letters 25(12), 1431~1437, 2004.

[6] R. C. Gonzalez, R. E. Woods, Digital Image Processing, Prentice Hall, Upper Saddle River, NJ, 2002.

[7] R. J. Anderson, F. A. P. Petitcolas, On the limits of steganography, IEEE Journal on Selected Areas in Communications 16, 474~481, 1998.

[8] F. A. P. Petitcolas, R. J. Anderson, M. G. Kuhn, Information hiding - a survey, Proceedings of the IEEE, special issue on protection of multimedia content, 87(7), 1062~1078, 1999.

[9] K. H. Jung, J. G. Yu, S. M. Kim, K. J. Kim, J. Y. Byun, K. Y. Yoo, The hiding of secret data using the run length matching method, LNCS, 1027~1034, 2007.

[10] Z. Wang, A. C. Bovik, A universal image quality index, IEEE Signal Processing Letters 9, 81~84, 2002.