Palette Partition Based Data Hiding for Color Images

Yu-Chiang Li[1], Piyu Tsai[2], Chih-Hung Lin[1], Hsiu-Lien Yeh[3], and Chien-Ting Huang[1]
[1]*Department of Computer Science and Information Engineering*
Southern Taiwan University, Tainan, Taiwan 710, R.O.C.
E-mail: {lyc002,chuck}@mail.stut.edu.tw; jason721111@hotmail.com
[2]*Department of Computer Science and Information Engineering*
National United University, Miaoli, Taiwan 360, R.O.C.
E-mail: pytsai@nuu.edu.tw
[3]*Department of Business Administration*
National Central University, Taoyuan, Taiwan 320, R.O.C.
E-mail:974401019@cc.ncu.edu.tw

Abstract

In this paper, a palette partition based data hiding for color images is proposed. The palette used in color image quantization is partitioned into sub-palettes according to the similarity of colors. Any size of sub-palettes is permitted and a color can appear in several sub-palettes at the same time. Furthermore, a mapping function is designed to process color embedding for color belongs to three-color sub-palette. The experimental results show that the proposed scheme has a higher hiding capacity than that of the similar methods. Furthermore, a large ratio between embedded secret image and the image distortion is obtained by the proposed scheme.

1. Introduction

Image data hiding employs ordinary image to cover a secret message [1]. The secret message is imperceptibly inserted into the ordinary image, which is called cover image. The cover image with the secret message forms stego-image. Generally, it is difficult to distinguish the difference between cover image and stego-image by the human visual system. Consequently, the stego-image can be transmitted on open channels without suspicion. Only the authorized user can extract the embedded secret message from the stego-image.

Several image data hiding schemes have been proposed to solve the secret transmission. They can be divided into three categories according to the format of the cover image. First, the cover image is in spatial format. The secret message is directly embedded into spatial domain. The simplest and well-known scheme is based on the least-significant-bits (LSBs) substitution [2-3]. Second, the cover image is in the frequency format. The spatial domain image is first transformed into frequency domain by discrete cosine transformation (DCT), discrete wavelet transformation (DWT) and so on. Third, the cover image is in quantization format. Only a few of methods is worked on this category. In fact, the quantization-based images such as vector quantization (VQ), color quantization (CQ)-based (palette-based) images, are widely used on the Internet and web pages such that all browsers can recognize them. In 2002, Jo and Kim proposed a watermarking method based on VQ [4]. In their scheme, the codewords in the codebook are clustered into three sub-codebooks. In 2004, Chang et al. partition color palette for color image data hiding [5]. The colors in palette are clustered into four-color, two-color and single-color sub-palette. The color in four-color sub-palette embeds more secret bits than that of in two-color sub-palette. In 2008, Chiang and Tsai improve codebook partition algorithm to which a codeword can overlap in several sub-codebook [6]. The codeword overlapping increases the number of sub-codebooks and preserves the similarity of codewords.

In this paper, Chiang and Tsai's codebook partition algorithm is employed to partition the color palette and an improved secret embedding procedure is performed to increase the hiding capacity.

2. Related Work

In quantized image data hiding, Chiang and Tsai's scheme divides the codebook into several sub-codebooks with codeword overlapping [6]. In their codebook partition algorithm, a set of codewords with strong similarity is clustered into a sub-codebook. If a codeword to be overlapped in many sub-codebooks, two conditions are detected. First, the location of an overlapping codeword in overlapping sub-codebooks should be identical. Second, the size of the overlapping sub-codebooks should also be the same. According that, the codebook is partitioned into four-codeword, three-codeword, two-member and one-codeword sub-codebooks with codeword overlapping.

In the secret embedding, the sub-codebook of the codeword belongs is first identified. The size of the identified sub-codebook determines the secret embedding. Four cases are considered. First, if the size of the identified sub-codebook is two, 1-bit secret message can be embedded. The codeword whose location in sub-codebook matches the secret bit is selected to replace the original codeword. Second, if the size of the sub-codebook is four, 2-bit secret message can be embedded. Similarly, the codeword whose location matches the embedded 2-bit secret message is employed. Third, the size of the sub-codebook is three, two codewords belonging in three-codeword sub-codebook are considered together. The combination of codewords in two three-codeword sub-codebooks, by production rule, matching the embedded 3-bit secret message is selected to embed 3-bit secret message. Last, the size of sub-codebook is one, this codeword is skipped for it is not used to embed secret message.

The secret embedding is further considered when overlapping codeword is encountered. A set of candidate codewords from the overlapping sub-codebooks are examined. One of the candidate codewords with the least image distortion is selected to replace the original codeword and to improve image quality.

3. The Proposed Scheme

3.1. The Palette Partition Algorithm

Palette is a set of colors employed in color image quantization. Each color in palette is the representative of sub-set of colors in an image. The similarity between colors in palette is explored to partition the palette into sub-palettes. Each color finds a set of closer colors in palette to form a sub-palette using the squared Euclidean distance (*SED*). If the summation of *SED* between colors is less than a predefined threshold (*TH*), a sub-palette is generated. The proposed palette partition algorithm divides the color palette into several sub-palettes with different sizes according to their similarity.

To increase the number of sub-palettes, the color overlapping among sub-palettes is considered. In Chiang and Tsai's scheme, the codebook is partitioned into several sub-codebooks with codeword overlapping. The codeword overlapping concept is employed in the proposed palette partition. The color overlapping preserves the colors in sub-palette with stronger similarity.

The proposed scheme computes and sorts the total *SED* among a set of colors. According to the order of the total *SED* and *TH*, it forms several sub-palettes. In the meanwhile, the proposed employs the overlapping technology to generate sub-palettes. Generally, a large size of sub-palette provides a great number of hiding capacities. According that, the color palette is first partitioned into four-color sub-palettes and then three-color sub-palettes, two-color sub-palettes and so on. The remained colors form single-color sub-palette with itself.

3.2. The Embedding Procedure

The original color image is first quantized with palette. The index of the searched closest color in palette is taken as the encoded color. Then, a color image is encoded to an index table. The index table is taken as the cover image to embed a secret message. In the secret embedding procedure, the sub-palette of the color (index) belongs is first identified. If the size of the identified sub-palette is greater than one, this color can be used to carry a secret message.

Three scenarios are considered. In the first scenario, the size of identified sub-palette is four, two secret bits can be embedded. In the second scenario, the identified sub-palette size is two, only one secret bit can be embedded. In the last scenario, the size of identified sub-palette is three, the special relationship of embedded secret message and the color location are listed in Table 1. One or two secret bits can be embedded by a color of sub-palette with size of three.

Table 1. The relationship of embedded secret message and color location

Location	0	1	2	3
Four-color	$(00)_2$	$(01)_2$	$(10)_2$	$(11)_2$
Three-color	$(00)_2$	$(1)_2$	$(01)_2$	NA
Two-color	$(0)_2$	$(1)_2$	NA	NA

Since color may overlap in several sub-palettes, the overlapping color embedding should be reconsidered. If an embedding color is overlapped in several sub-palettes, the colors in overlapped sub-palettes, which location matches the embedded secret message, are considered. One of colors with the least *SED* between the color and the original one is selected to reduce the distortion.

For example, in Table 2, a set of quantized colors is used to embed a secret message. The sub-palette size of each color belonged indicates the secret message to be embedded. For example, color 146 belongs to three-color sub-palette. It can embed 1-bit secret value of $(1)_2$ or 2-bit secret value of $(00)_2$ or $(01)_2$. If 2-bit secret $(01)_2$ is read, the corresponding color 145 is selected to replace the quantized color 146. It is noted that if 1-bit secret message value of $(1)_2$ to be embed, one of colors 147 and 137 with the least distortion is selected. This is because color 146 is an overlapping color. Similarly results are shown in colors 39 and 149 to which one of colors with the least distortion is selected.

Table 2. Example of secret embedding

Quantized Color	146	67	39	149
Sub-palette Size	3	1	4	2
Embedded Color	$147(1)_2$ $137(1)_2$ $146(00)_2$ $145(01)_2$	67	$32(00)_2$ $53(01)_2$ $15(01)_2$ $39(10)_2$ $47(11)_2$ $16(11)_2$	$125(0)_2$ $131(0)_2$ $149(1)_2$

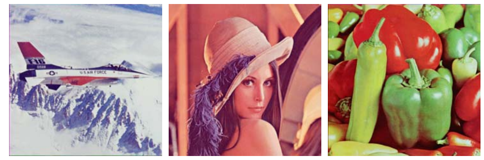
3.3. The Extraction Procedure

To extract the embedded message, the size of identified sub-palette determines whether a secret message is embedded or not. If the sub-palette size is greater than one, a secret message is hidden. Otherwise, no secret message is embedded. Three scenarios are considered to extract the embedded secret message. In the first scenario, the size of the identified sub-palette is equal to four. The location of color in sub-palette is the extracted secret message. In the second scenario, the size of the identified sub-palette is equal to two. Also, the location of color is the extracted secret message. Last, the identified sub-palette size is three. In this scenario, the extracted secret message is extracted according to Table 1.

The overlapping color in secret extraction procedure can be ignored, because the size of sub-palettes and the location of colors in sub-palette are the same. So the extracted secret message is identical. Therefore, the color overlapping is nothing to do with in the secret extraction.

4. Experimental Results

In this experiment, three RGB color images, "Airplane," "Lena," and "Peppers" of 512 × 512 pixels were encoded by the color quantization technique. The palette of each color image with 256 3-dimensions was generated by Photoshop version 6.0 with optimal option. The quantized color images were shown in Figure 1 and taken as cover images. The secret message is composed of a bit stream generated by a pseudo random number generator (*PRNG*).



(a) Airplane (b) Lena (c)Peppers

Figure 1. The cover images of 512×512 pixels

The palette of each color image was partitioned into sub-palettes with different sizes. Table 3 showed the partition results in which the number of sub-palettes and the number of individual colors were involved. For example, the number of four-color sub-palette in image "Lena" is 146 and the number of individual colors is 236. Generally, the number of colors is 584 for 146 four-color sub-palettes if color overlapping is disallowed. In other words, the number of overlapping color in this case is 348 (i.e., 584 –236). This result shows that the color overlapping efficiently increases the number of sub-palette. Similarly, the numbers of overlapping colors in three-color and two-color sub-palettes are 2 and 0, respectively.

Table 3. The numbers of sub-palettes and individual colors

Numbers / Images	Sub-Palette size		
	Four-color	Three-color	Two-color
Airplane	113(214)	3(9)	2(4)
Lena	146(236)	2(4)	1(2)
Peppers	71(141)	5(12)	22(42)

Table 4 showed the number of colors of different sub-palette sizes in each cover image. The hiding capacity can be estimated according to Table 1. The hiding capacity of color in three-color sub-palette is averaged to 1.67 (i.e., 5/3). From Table 4, it is found that the greatest number of sub-palettes is four-color sub-palette. This is because most of colors are clustered into four-color sub-palettes. In other words, four-color sub-palette provides the greatest hiding capacity than other sub-palette sizes. From Table 4, the estimated capacity of image "Airplane" is about 515K bits.

The image quality is measured by peak signal-to-noise ratio (*PSNR*) and the hiding capacity is measured by the number of embedded secret bits. An experiment was performed to compare the performance of Jo and Kim's, Chiang and Tsai's and the proposed methods. Jo and Kim's and Chiang and Tsai's schemes were implemented in color images. The number of embedded secret message is determined according to each scheme provided. Table 5 showed the performance of schemes. From Table 5, it is seen that a best image quality and a lowest hiding capacity was provided by Jo and Kim's scheme. The proposed method provides the largest hiding capacity than other methods. In comparison result with Chiang and Tsai's, a better stego-image quality and a higher hiding capacity are obtained by the proposed method. The ratio shown in last row of Table 5 means the proportion of the number of embedded secret bits and the image distortion caused by secret embedding. For example, the average image quality of cover image is 36.15 dB and the average stego-image quality of the proposed method is 29.84 dB. The image distortion is 6.31 dB (i.e., 36.15-29.84). The large ratio implies a great number of secret bits are embedded and a small amount of image distortion is caused. From Table 5, it is shown the largest ratio is provided by the proposed method. According that, the proposed method has improved the hiding capacity and stego-image quality from Chiang and Tsai's method.

Table 4. The number of colors with different sub-palette sizes and the estimated hiding capacity

Sub-palette	Four-color	Three-color	Two-color	Capacity
Airplane	255,216	2,487	893	515,478
Lena	251,938	3,056	553	509,532
Peppers	197,045	14,947	26,457	445,508

Table 5. The performance of Jo and Kim's, Chiang and Tsai's, and the proposed methods

Method / Covers	Jo & Kim's		Chiang& Tsai's		Proposed	
	PSNR	Capacity	PSNR	Capacity	PSNR	Capacity
Airplane	35.58	207,966	32.13	512,568	32.11	515,037
Lena	33.52	206,248	30.08	505,957	30.11	509,036
Peppers	30.67	150,756	27.28	428,020	27.31	442,989
Average	33.26	188,323	29.83	482,181	29.84	489,020
Ratio	65,164		76,295		77,499	

5. Conclusions

In this paper, an overlapping color palette partition based data hiding with improved secret embedding procedure has been presented. The similarity of colors in palette is explored to generate any size of sub-palettes. During color clustering, color overlapping not only increases the number of sub-palettes but also preserves the similarity of colors in sub-palette. Since most of colors are clustered into sub-palette with large size, the hiding capacity is therefore increased. In addition, the improved secret embedding procedure further enhances the hiding capacity in three-color sub-palette. The experimental results have shown that a large hiding capacity and a better stego-image quality are provided.

The comparison results also indicate that the proposed method provides a largest hiding capacity and supports better stego-image quality than Chiang and Tsai's method. In comparison with Jo and Kim's method, the proposed method provides approximately two and half hiding capacity while the image degradation is about 3.5 dB. However, the ratio of embedded secret message and image distortion indicates that the proposed method obtains a higher value.

Acknowledgements

This study was supported in part by the National Science Council, Taiwan, grants NSC97-2218-E-218-001 and NSC97-2221-E-218-046.

References

[1] D. Artz, "Digital Steganography: Hiding Data within Data", *IEEE Internet Computing*, Vol. 5, No.3, 2001, pp. 75-80.

[2] Y. K. Lee and L. H. Chen, "High Capacity Image Steganographic Model", *Proceedings of IEE International Conference on Vision, Image and Signal Processing*, Vol. 147, No. 3, 2000, pp. 288-294.

[3] R. Z. Wang, C. F. Lin and J. C. Lin, "Image Hiding by Optimal LSB Substitution and Genetic Algorithm", *Pattern Recognition*, 34(3), 2001, pp. 671-683.

[4] Jo, M. and Kim, H. D., "A Digital Image Watermarking Scheme Based on Vector Quantization", *IEICE Transactions on Information and Systems* , Vol. E85-D, No. 6, 2002, pp. 1054-1056.

[5] C. C. Chang, P. Tsai, and M. H. Lin, "An Adaptive Steganographic for Index-Based Images Using Codeword Grouping", *The fifth International Pacific-Rim Conference on multimedia*, Tokyo, Japan, Nov. 30- Dec. 3, 2004 (PCM-2004) , pp. 731-738.

[6] Y. K. Chiang and P. Tsai, "Steganography Using Overlapping Codebook Partition", *Signal Processing*, Vol. 88, No. 5, 2008, pp. 1203-1215.