Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Reversible data embedding for indices based on histogram analysis

Chin-Chen Chang [a,d,e], Thai Son Nguyen [a,b], Chia-Chen Lin [c,*]

[a] Department of Information Engineering and Computer Science, Feng Chia University, Taichung 40724, Taiwan, ROC
[b] Department of Information Technology, Travinh University, Travinh Province, Viet Nam
[c] Department of Computer Science and Information Management, Providence University, Taichung 43301, Taiwan, ROC
[d] Department of Biomedical Imaging and Radiological Science China Medical University, Taichung 40402, Taiwan, ROC
[e] Department of Computer Science and Information Engineering Asia University, Taichung 41354, Taiwan, ROC

ARTICLE INFO

Article history:
Received 19 October 2013
Accepted 8 June 2014
Available online 19 June 2014

Keywords:
Data hiding
Image compression
Reversible
Steganography
VQ compression
Mapping function
Histogram analysis
High capacity

ABSTRACT

Reversible data hiding is a technique that not only protects the hidden secrets but also recovers the cover media without any distortion after the secret data have been extracted. In this paper, a new reversible data hiding technique for VQ indices which are compressed streams based on the mapping function and histogram analysis of transformed VQ indices is introduced to enhance the performance of some earlier reversible data hiding schemes that are based on VQ indices. As a result, the proposed scheme achieves high embedding capacity and data compression simultaneously. Moreover, the original VQ-compressed image can be perfectly reconstructed after secret data extraction. To estimate the performance of the proposed scheme, variety of test images are used in the experimental testing. As can be seen in the experimental result, our scheme is superior to some previous schemes in term of compression rate and embedding rate while maintaining the reversibility.

1. Introduction

The Internet is a popular channel for allowing more and more users to exchange information without geographical limitations and time restrictions. Therefore, a large amount of digital data, such as digital images, text, audio, and video, is transmitted through the Internet every second. Unfortunately, this means that sensitive digital data can be intercepted by malicious users of the Internet. Therefore, the development of techniques for ensuring the secrecy and security of data during transmission has become a top priority. To solve the above-mentioned problem, researchers have introduced various algorithms, such as encryption [1,2], steganography [4,8,14], and data hiding algorithms [5,9–11]. In encryption, data are transformed into meaningless, undecipherable form. Then, only an authorized user can recover the original form of these meaningless data by using a secret key that is supplied in advance through the security channel. However, the secret data in the meaningless form will be curious to malicious users, and they may try to decrypt the messages to get information. To avoid unnecessarily attracting the attention of malicious users, steganography was developed to conceal the secret data in the cover medium. Different data hiding algorithms have been proposed to guarantee the security of data during transmission by hiding the secret information in the digital data of the cover image, such as the "Hello Kitty" image, to avoid attracting the attention of malicious attackers. In previous data hiding schemes in the literature, the slight changes in the cover data were not expected to attract the attention of possible attackers. However, in such schemes, the digital data may be altered and cannot be recovered after the secret information has been extracted. In some fields, such the military and medical fields, it is essential to be able to reconstruct the original digital data in the cover image without any distortion after the secret information has been extracted. To deal with this issue, various reversible data hiding techniques have been proposed. Reversible data hiding techniques begin with the hiding phase, in which secret information is embedded into the digital data of the cover image. Then, in the extracting phase, the secret information can be extracted correctly, and the digital data of the cover image can be recovered in its original form.

Data hiding can be further classified into three different domains based on the category of the cover image, i.e., the spatial domain, the transform domain, and the compression domain. Spatial domain [3,7,8] techniques can hide secret data by modifying their spatial information. In the transform domain [5,13,18,20,21], the coefficients of the cover image are shifted to embed secret information.

[*] Corresponding author. Address: Department of Computer Science and Information Management, Providence University, No. 200 Chung-chi Rd., Taichung 43301, Taiwan, ROC. Fax: +886 4 24060438.

E-mail addresses: ccc@cs.ccu.edu.tw (C.-C. Chang), thaison@tvu.edu.vn (T.S. Nguyen), ally.cclin@gmail.com (C.-C. Lin).

http://dx.doi.org/10.1016/j.jvcir.2014.06.003
1047-3203/© 2014 Published by Elsevier Inc.

Unfortunately, these two techniques require the use of larger bandwidth space and larger storage space because the sizes of various media files are increased significantly by such processing. Therefore, the spatial and transform domains are not suitable for transmitting secret multimedia files with limited bandwidth. In order to reduce the excessive use of bandwidth in the network and to reduce the storage space required as the result of encoding media files, the compression domain was introduced as the best choice in this case. The main reason is that, after an image is processed in the compression domain, the size of the compression code of the image is considerably smaller than the original cover image. Even the compression code carries secret information; its size still is less than that of the original cover image. Therefore, data hiding algorithm in compression domain can guarantee both of the advantages, data hiding and compression. In the past 10 years, there has been extensive research related to the use of the compression domain to conceal secret information, such as VQ [10,12,19], SMVQ [15,16], and JPEG [17,18]. In 2001, to authenticate digital images in JPEG domain, Fridrich et al. [17] proposed an invertible data hiding scheme by using a second-order function to modify the quantization table, and one bit secret was embedded into the DCT coefficient. Then, in 2004, Xuan et al.'s scheme [20] was presented to embed secret data into the high-frequency wavelet coefficients by depending on the histogram modification algorithm. In 2005 and 2006, Chang et al. [12,15,16] proposed several reversible data hiding schemes by using SMVQ and modified SMVQ. Since the SMVQ-based algorithm uses a sub-codebook created from the super codebook to encode the image blocks, the compression rate of the image is better than the traditional VQ. Although the high compression rate is obtained, the quality of stego-images is reduced as the sub-codebook size is decreased. In [5], Lin and Chang proposed a different reversible data hiding scheme that relied on a pair of dissimilar codewords in the sorting codebook. In this scheme, one bit of secret will be hidden with each pair of dissimilar codewords. However, this scheme cannot ensure that the cover image is reconstructed correctly. In 2007, Chang et al. [22] presented a new reversible data hiding scheme by dividing the traditional VQ codebook into three groups to embed secret data while retaining the feature of reversibility. Then, in 2008, Wang and Lu [19] presented a lossless data hiding scheme by depending on improved VQ index neighboring coding. By using this scheme, the high image quality of the stego-image was obtained, when the average of the *PSNRs* was about 30 dB. This was better than Lin and Chang's scheme. However, Wang et al.'s average embedding rate was just 0.85 bpi, which is still less than that of Lin and Chang's scheme (0.98 bpi). In 2009, Chen and Huang [23] proposed a hybrid scheme called HLIC by using the combination of the dynamic tree coding scheme (DTCS) and modified search order coding (MSOC) to encode the index table completely and efficiently without generating any extra coding distortion. The HLIC algorithm is used to embed secret information by depending on a relative of the current processed block and its neighboring blocks, i.e., the left block and the upper block. Also in 2009, Chang et al. [11] proposed a new reversible data hiding scheme for VQ-compressed images based on joint neighboring coding (JNC). Lu et al. [25] presented a lossless data hiding scheme based on image VQ index residual value coding. In this scheme, the current index is encoded by depending on the relative of the current index and its adjacent indices. Yang and Lin [24] proposed a new reversible data hiding scheme in the VQ domain. In Yang and Lin's scheme, the VQ codebook is sorted and divided into 2^B clusters, where B is defined as the number of secret bits embedded into each VQ index and half of these clusters are used to hide secret data. To achieve high embedding capacity, Wang and Lu [26] introduced a path optional lossless data hiding scheme by using VQ joint neighboring coding without generating any distortion. Wang and Lu's average embedding rate was about 1.95 bpi, which is better than some previous schemes [23–25]. However, the average compression rate of Wang and Lu's scheme was larger than 0.6 bpp, which is worse than the previous schemes. In 2010, to get a better compression rate, Lee et al. [29] proposed a new irreversible data hiding scheme by depending on SMVQ prediction through classification codebooks. However, the average embedding capacity of Lee et al.'s scheme (1.0 bpi) was lower than that of Wang and Lu's scheme when the codebook size of 256 was applied and the threshold *THvar* of Lee et al.'s scheme was set as 50, respectively. In 2013, Chang et al. [30] proposed a new compression and information hiding hydrid scheme based on SMVQ and search order coding (SOC) in order to improve the compression rate. However, the embedding rate obtained by their scheme is small because most of the indices simply embed one bit of secret data. After careful reviewing the existing schemes, we discovered that the main problems in the existing schemes are low embedding capacity or high compression rate. To overcome these problems, in this paper, we proposed a new, reversible data hiding scheme. In the proposed scheme, SMVQ technique is applied to transform the VQ index table to transformed index table. Then, we apply the histogram analysis algorithm proposed in [28] to analyze the histogram of the transformed index table and construct the suitable mapping function. This mapping function is referred in the embedding phase to obtain both high embedding rate and low compression rate. During extracting phase, the mapping function is used to ensure the reversibility. As documented in the experimental results, our scheme is superior to that of existing schemes [24–26,29] in term of embedding rate and compression rate while maintaining the same visual quality of the image as these existing schemes and the traditional VQ algorithm.

The remainder of the paper is organized as follows. Section 2 reviews VQ, SMVQ, and histogram analysis. Then, the details of the proposed scheme are illustrated in Section 3. Experimental results are provided in Section 4, and our conclusions are presented in Section 5.

2. Related work

In our proposed scheme, VQ was used to generate an index table for a cover image. Then, SMVQ was used to transform the index table into a transforms index table. Finally, histogram analysis was used to select one mapping function that offers the highest hiding capacity for the cover image. To give readers enough background knowledge to understand our proposed scheme, in this section, we introduce vector quantization (VQ) and side match vector quantization (SMVQ) in Sections 2.1 and 2.2, respectively. In Section 2.3, Saleh et al.'s histogram analysis [28] is introduced.

2.1. Vector quantization

In the literature, vector quantization (VQ) [6] can be presented as a map function from the k-dimensional space R^k to its finite subset $Y = \{y_i; i = 0, 1, \ldots, N-1\}$, where y_i is denoted as a codeword, and the set Y denotes a codebook. The nearest neighboring region, also called the Voronoi region, is defined as follows:

$$X_i = \begin{cases} a \in R^k : \|a - y_i\| < \|a - y_j\| & \text{for all } j \neq i, \\ a \in R^k : \|a - y_i\| = \|a - y_j\| & \text{for all } j \neq i, \end{cases} \tag{1}$$

Then, the set of nearest neighboring regions divides all space R^k by using Eq. (2).

$$\bigcup_{i=0}^{N-1} X_i = R^k, \quad X_i \cap X_j = \phi, \ \forall i \neq j. \tag{2}$$

In the VQ system, first, the codebook is generated by using a well-known, iterative, clustering algorithm [6] that is often referred to as the generalized Lloyd algorithm (GLA). Fig. 1 shows the flowchart of the VQ encoder.