Information Sciences 293 (2015) 314–326

Contents lists available at ScienceDirect

Information Sciences

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

Reversible data hiding for VQ-compressed images based on search-order coding and state-codebook mapping

Chia-Chen Lin [a,*], Xiao-Long Liu [b], Shyan-Ming Yuan [b]

[a] *Dept. of Computer Science and Information Management, Providence University, Taiwan, ROC*
[b] *Dept. of Computer Science, National Chiao Tung University, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 4 March 2014
Received in revised form 12 August 2014
Accepted 29 August 2014
Available online 27 September 2014

Keywords:
Data hiding
Vector quantization (VQ)
Search-order coding
State-codebook mapping

ABSTRACT

In the field of digital multimedia, numerous researchers have exploited data hiding for providing secure communications. Because most multimedia is compressed in advance to save storage space and transmission time, this paper presents a reversible data-hiding scheme for vector quantization (VQ)-compressed images. In the proposed scheme, search-order coding (SOC) is first employed to reduce the size of the VQ index table. For image blocks that cannot be processed by SOC, a technique called state-codebook mapping (SCM) is used to map the blocks to the corresponding indices in their state-codebooks, which are generated by a set of neighborhood indices. After SOC and SCM, the size of the output code stream is reduced, and a large amount of space is available for embedding sensitive data. Performance comparisons with similar recent studies are presented for demonstrating the superiority of the proposed scheme. The experimental results indicated that the proposed scheme can achieve a high embedding capacity and a high compression bit rate.

1. Introduction

With the rapid progress of an information-oriented society, increasingly large quantities of digitalized data are being transmitted through open channels over the Internet. Concerns pertaining to the security and confidentiality of the transmitted digitalized data have become critical over the past decade. Traditionally, cryptographic algorithms such as DES [12] and RSA [31] have been used for securing data. Recently, the data-hiding technique [1,2,10,13,32,34,37] has emerged as an alternative because it embeds private data imperceptibly in digital content. Schyndel et al. proposed the first data hiding scheme in 1994 [32]. In 1997, Barton proposed the first reversible data hiding scheme, which facilitates the complete restoration of the original host image once the hidden data are extracted [2]. A reversible data hiding technique can be applied in the spatial [23,38], frequency [4,25], and compression [5,22] domains of the host images. Because compression techniques can be applied to digital content to reduce the required storage space and transmission bandwidth, diverse digital image-compression techniques have been developed by academic researchers and the industry, such as block truncation coding (BTC) [11], JPEG [40], and vector quantization (VQ) [26]. Currently, the compressed image format is one of the most frequently transmitted formats over the Internet. Therefore, scholars have been developing various reversible data hiding techniques for diverse digital image-compression techniques [6,8,18,29].

* Corresponding author at: Department of Computer Science and Information Management, Providence University, No. 200 Chung-chi Rd., Taichung 43301, Taiwan, ROC. Tel.: +886 4 26328001x18108; fax: +886 4 24060438.
E-mail addresses: mhlin3@pu.edu.tw (C.-C. Lin), shallen548@gmail.com (X.-L. Liu), smyuan@cs.nctu.edu.tw (S.-M. Yuan).

http://dx.doi.org/10.1016/j.ins.2014.08.057

VQ, a popular block-based image-compression technique, is considered simple and efficient compared with other compression schemes [11,40]. In traditional VQ, each image block is regarded as a multidimensional vector and is encoded as the index of the closest codeword in the codebook; thus, an index table is generated. Image compression is achieved by transmitting the generated indices, instead of pixels, in the spatial domain. Recently, numerous researchers have studied methods of embedding secret messages in VQ-compressed images. In 1999, Lin and Wang [24] proposed an adjustment method, which is regarded as the first data hiding method for VQ-compressed images; in this scheme, two neighboring codewords are considered as a pair for embedding bits ("0" or "1"). In 2005, Yang et al. [39] proposed a reversible watermarking scheme based on modified fast correlation VQ (MFCVQ) for VQ-compressed images. However, the scheme has a considerably low hiding capacity because its embedding rate is less than 1 bit/image block on average. Therefore, in 2009, Lin et al. [21] presented an adaptive embedding technique for VQ-compressed images. They adjusted the predetermined distance threshold according to the required hiding capacity and arranged numerous similar codewords in one group to embed a secret submessage. Thus, on average, two bits could be embedded in each block with acceptable distortion. In 2013, Lee et al. [18] presented a high-capacity lossless data hiding scheme for VQ indices based on neighboring correlation. They first used a sorted codebook to compress the cover image into an index table. Neighboring indices at the top and left of the current encoding index were then used to generate four sub-codebooks to further encode the current encoding index and embed sensitive data simultaneously. Thus, more than 40,000 bits could be embedded in each image on average. Subsequently, for reducing the size of the output code stream while providing a large embedding capacity, Wang et al. [36] proposed a scheme based on adjoining state-codebook mapping (ASCM). They used two adjacent image blocks of the image block to be encoded for creating two state-codebooks that can not only represent the subject image block more accurately but also embed secret data in the compression codes adaptively. Unlike conventional VQ-compression, the scheme of Wang et al. reduces the output codestream size by 13% on average when 16,129 bits are embedded in each test image. Soon after Wang et al. [36] proposed their scheme, Qin et al. [29] presented an efficient reversible data hiding scheme for VQ-compressed images. To perform reversible data hiding, they constructed a series of index mappings among VQ indices with zero occurrence and those with the highest frequency of occurrence from an index table generated for a given image and secret data. The Qin et al. scheme offers efficient data embedding and extraction procedures at low computation cost, and on average, 18,348 secret bits can be embedded in each image.

To improve the compression rate, side match vector quantization (SMVQ) [15] was developed as a variation of VQ in 1992. In SMVQ, blocks in the leftmost column and the topmost row are processed using traditional VQ, and the residual blocks are processed using SMVQ. For each block encoded using SMVQ, the previously processed blocks located above and to its left are used for predicting the block being processed. In general, SMVQ offers visual quality superior to that of traditional VQ. Therefore, several data hiding schemes [7,8,30,33,35] using SMVQ for embedding secret data have been proposed. In 2006, Chang et al. [7] proposed an SMVQ-based reversible data hiding algorithm in which secret data are hidden in an SMVQ-compressed host image and can be extracted completely in two simple steps. In 2010, Lee et al. [19] proposed an SMVQ-based data hiding scheme in which histogram modification of SMVQ indices was employed to increase the probability of achieving a high embedding rate. Additionally, three configurations – under hiding, normal hiding, and over hiding – were proposed for balancing the embedding capacity and the compression rate. In 2012, Shie and Jiang [33] hid secret data in the SMVQ-compressed codes of an image by using a partially sorted codebook; the codebook was used to not only increase the amount of secret data that could be hidden but also increase the visual quality of the reconstructed image, while maintaining an acceptable bit rate. For integrating data hiding and image compression into a single, seamless module, in 2013, Qin et al. [30] proposed a joint data hiding and compression scheme for digital images that uses SMVQ- and image inpainting. Using this scheme, with the exception of the leftmost and topmost blocks in the host image, all other residual blocks can be embedded with secret data and compressed simultaneously using either SMVQ or image inpainting, according to the secret bit. In 2014, Wang et al. [35] presented a reversible data hiding scheme involving a combination of locally adaptive coding (LAC) and SMVQ. In this scheme, a self-organizing sequence and an adaptive threshold list are first defined to concentrate the distribution of index positions so that the variance among the indices can be reduced. Thereafter, a variable coding rule is employed for embedding secret bits and decreasing the length of the code stream carrying the secret bits. Unlike conventional VQ-compression, the scheme of Wang et al. can reduce the output code-stream size by 10% on average, and the associated average hiding capacity per test image is approximately 30,000 bits.

In addition to SMVQ, the search-order coding (SOC) algorithm is another popular variant of VQ [14] that can be used to further compress the index table generated by conventional VQ. With SOC, a superior compression rate can be achieved by searching nearby identical image blocks that follow a spiral path. Some data hiding schemes [3,6,9,16,28] have been designed to embed secret data in SOC-compressed codes. In 2004, Chang et al. [3] proposed an information hiding scheme in which the SOC algorithm and conventional VQ algorithm were combined. In this scheme, receivers can determine whether each bit of secret data is "0" or "1," depending on whether the received compression code is SOC index value or an original VQ index value. However, this scheme switches the SOC index values with the original VQ index values for concealing secret bits, which leads to a considerable increase in the bit rate compared with the SOC algorithm. To reduce the bit rate, in 2009, Chen and Huang [9] used the combination of a dynamic tree-coding method and a modified SOC method for reencoding the VQ index table. Subsequently, they applied hybrid lossless index coding (HLIC) to the reencoded VQ index table for data hiding. The principle behind their data hiding strategy is that the adjacent left and upper blocks of the block being processed typically provide more useful information than the adjacent left-upper and right-upper blocks do. Unlike conventional VQ- compression, the scheme of Chen and Huang reduces the output code stream size by 30% on average, but their hiding