J. Vis. Commun. Image R. 20 (2009) 399–407

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Reversible data hiding of a VQ index table based on referred counts

Cheng-Hsing Yang *, Yi-Cheng Lin

Department of Computer Science, National Pingtung University of Education, Pingtung 900, Taiwan

ARTICLE INFO

Article history:
Received 21 October 2008
Accepted 22 April 2009
Available online 3 May 2009

Keywords:
Data hiding
Reversible embedding
Vector quantization
Codebook clustering
Lossless compression
Steganography
Side-match
Predictive coding

ABSTRACT

This paper presents a new reversible VQ-based hiding scheme that can recover the original VQ compressed codes after data extraction. Our scheme sorts a VQ codebook using the referred counts. The VQ codebook is then divided into 2^B clusters and half of these clusters are used to embed secret data, in which B denotes the size of the secret data embedded into each VQ index. Compared to Chang et al.'s scheme, which divides a sorted VQ codebook into $2^{B-1} \times 3$ clusters and uses the front one-third clusters to embed secret data, our method can embed more data. Moreover, indicator, index exchanging, and side-match prediction schemes are proposed to further improve our scheme. Under the same sorted VQ codebook, the experimental results demonstrate that our data hiding algorithm has higher capacities and better compression rates.

1. Introduction

As multimedia and the Internet have become popular, the problem of protecting transmitted media has become more important. Enhancing the safety of information transmissions using technologies based on data hiding [1,2] has attracted great attention. Data hiding usually embeds secret data into media, such as images and videos, for the purpose of secret transmission or copyright protection. This paper uses images as the embedded media. Images before and after data hiding are called cover images and stego-images, respectively. Because data hiding could damage the original images, a good data hiding technology should have high capacity to embed a lot of data with imperceptive impact upon the quality of stego-images and undetectability to pass program detection [3,4].

Typically, digital image data are transmitted in compressed format. Compression techniques such as JPEG, JPEG 2000, vector quantization (VQ), and block truncation coding have been proposed. Some hiding schemes based on VQ have been proposed [5–7]. Fig. 1 depicts an example of the VQ encoder in which the codebook size and vector dimensions are set to 256 and 16, respectively. When a 4×4 block (or vector) image is imported, the VQ encoder seeks the most similar codeword from the codebook to substitute for the input block. In this case, a codeword with index 2 is selected. The value 2 is exported as the compressed code for the input block.

In 1992, side-match VQ (SMVQ), which can improve vector quantization (VQ) compressing performance, was proposed by Kim [8]. In the SMVQ method, the blocks of an image in the first row and the first column are encoded using VQ. The residual blocks are predicted using the neighboring encoded blocks. Fig. 2 shows an example of SMVQ. Instead of using the original pixels to encode the X block, the SMVQ method uses the upper U block and the left L block to encode the X block. Let $X_0 = (U_{12} + L_3)/2$, $X_1 = U_{13}$, $X_2 = U_{14}$, $X_3 = U_{15}$, $X_4 = L_7$, $X_8 = L_{11}$, and $X_{12} = L_{15}$. The seven assigned pixels are used to search the m closest codewords from the super codebook. The m closest codewords are then used to construct a state codebook. Finally, the codeword for the state codebook with the minimum Euclidean distance from X is used to encode X.

Some VQ-based hiding methods have the reversibility characteristic [9–15]. Reversible data hiding based on VQ generally refers to possessing the ability to extract hidden data and recover the images into the original VQ coding or the SMVQ coding.

We placed the developed reversible data hiding technologies based on VQ into three categories using the output characters as follows.

1.1. Images as outputs

After data hiding, some approaches are limited to producing images as outputs [9,11]. Literature [9] presented a reversible data-hiding scheme based on side-match vector quantization (SMVQ) for digitally compressed images. Another literature [11] presented a reversible information hiding scheme based on VQ.

* Corresponding author. Fax: +886 8 7215034.
E-mail addresses: chyang@mail.npue.edu.tw, bm096114@mail.npue.edu.tw, cathy8.wang@msa.hinet.net (C.-H. Yang).

1047-3203/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2009.04.001

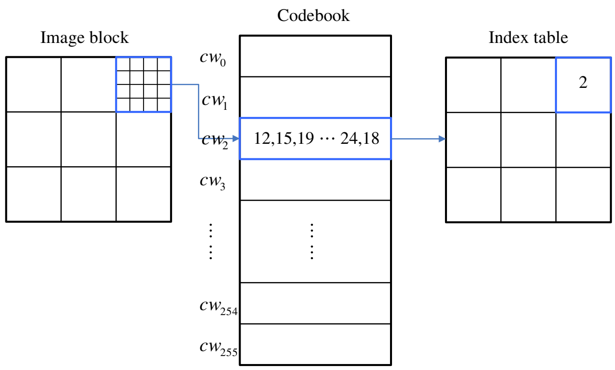

Fig. 1. An example of the VQ encoder.

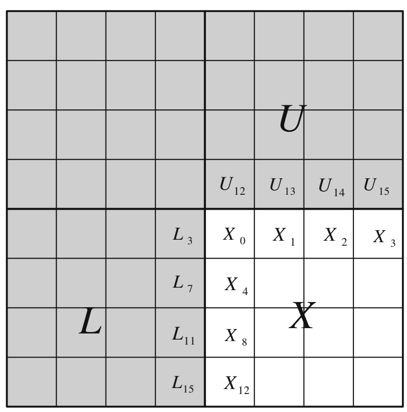

Fig. 2. An example of SMVQ.

1.2. Legitimate VQ coding or SMVQ coding as outputs

After data hiding, a formal VQ coding or SMVQ coding is created as outputs [10,12]. Generally speaking, approaches in this category require more skills. Of course, a formal VQ coding or SMVQ coding can be transformed into images. Literature [10] proposed a reversible embedding scheme for VQ-compressed images that is based on side matching and relocation. The method achieves reversibility without using the location map. To achieve the goal that the compressed cover image can be used repeatedly by different users, in [12] a reversible data-hiding scheme based on a modified side-match vector quantization (SMVQ) technique was proposed.

1.3. VQ coding or SMVQ coding with additional control messages as outputs

Approaches in this category add control messages into the formal VQ coding or SMVQ coding as outputs [13–15]. The volume of the embedded result is usually smaller than that of the original compressed image plus the hidden data. Therefore these methods can be applied to digital libraries. Literature [13] proposed a reversible scheme for VQ-compressed images that is based on a declustering strategy and takes advantage of the local spatial characteristics of the image. The main advantages of this method are ease of implementation, low computational demands and no requirement for auxiliary data. Further, literature [14] proposed an information hiding scheme based on side-match vector quantization (SMVQ), which conceals the secret information in the SMVQ compressed image indices. The methods in these two references can recover the embedded results into the original SMVQ coding. However, in literature [15], the embedded results can be recovered into the original VQ coding.

A hiding approach belonging to the third category is proposed in this paper. Our approaches can recover the original VQ coding, and also flexibility adjusts the embedding capacity. Compared to Chang et al.'s method [15], our approach has larger capacity. The remainder of this paper is as follows. Chang et al.'s method is introduced in Section 2. Section 3 describes the details of our proposed scheme. Section 4 shows some experimental results. Our conclusions are given in Section 5.

2. Past work by Chang et al

In 2007 Chang et al. provided a VQ-based embedding method that can losslessly recover the VQ index table [15]. The goal of their method is to enlarge the embedding capacity. Keeping the stego-image low in distortion is not considered in their method. Their reversible embedding algorithm could perform well in the digital library. In their method, a codebook is partitioned into some clusters. Some indexes in the codebook are reserved for acting as indicators. Data are embedded into the VQ index table by transferring index values from one cluster into another cluster. Sometimes the index values are led by indicators. For a traditional codebook Ψ with N codewords, the codebook is redesigned as codebook Ψ' with $N' = \left\lfloor \frac{N-2^{B-1}}{2^{B-1}\times 3} \right\rfloor \times 2^{B-1} \times 3$ codewords, where B denotes the size of secret data embedded into each VQ index. The surplus 2^{B-1} index values are used as the indicators. Some standard images were used to train the codebook Ψ' in their paper. The codewords in Ψ' were sorted using the referred counts in descending order. The codebook Ψ' is then partitioned into $2^{B-1} \times 3$ clusters of the same size $m = \left\lfloor \frac{N-2^{B-1}}{2^{B-1}\times 3} \right\rfloor$. The front one-third clusters with the highest referred counts were used to embed secret data.

Fig. 3 shows an example of Chang et al.'s proposed method, where index table T is transformed into index table T'. The original codebook Ψ consists of 16 codewords ($N = 16$) and the new sorted codebook Ψ' has 15 codewords ($N' = 15$). The new codebook Ψ' is divided into three clusters C_0, C_1, and C_2, all of which consist of 5 codewords ($m = 5$). In this example, the secret data $(01101)_2$ will be embedded into the index table. Here, the first index 2 belongs to cluster C_0, so the one-bit secret data 0 is embedded and index 2 is transformed into index 7 in cluster C_1. The second index 1 situated in cluster C_0 is transformed into index 11 in *cluster* C_2 to hide the one-bit secret 1. From index table T', the underlined index means that no secret data is embedded in the index and the index with the symbol || means that an indicator I_0 valued 0 is carried ahead. In this case, each index value or indicator 0 can be represented by $\lceil \log_2 N \rceil$ bits. However, to represent the index following indicator 0 needs only $\lceil \log_2 m \rceil$ bits.

Chang et al. extended their method to hide more secret data. Fig. 4 shows an example of Chang et al.'s proposed method which can embed two-bit secret data into each index value in cluster C_0 or cluster C_1 ($B = 2$). The original codebook Ψ consists of 32 codewords ($N = 32$) and the new sorted codebook Ψ' has 30 codewords ($N' = 30$). The new codebook Ψ' is partitioned into six clusters, each of them contains 5 codewords ($m = 5$). In this case, two indicators

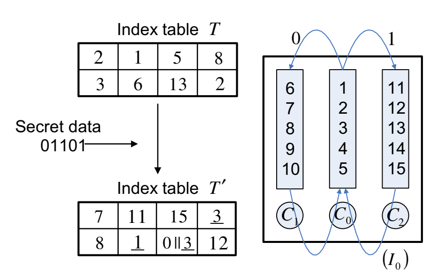

Fig. 3. Example of Chang et al.'s method for embedding one-bit secret data.