انجمن رمز ایران
Iranian Society of Cryptology

8th International ISC Conference
on Information Security and Cryptology
(ISCISC'11)

Ferdowsi University of Mashhad

Ferdowsi University of Mashhad-September 14-15, 2011

A Reversible Data Embedding Scheme Based on Search Order Coding for VQ Index Tables

Peyman Rahmani, Gholamhossein Dastghaibyfard
Department of Computer Science and Engineering
Shiraz University
Shiraz, Iran
rahmani@cse.shirazu.ac.ir, dstghaib@shirazu.ac.ir

Ehsan Rahmani
Islamic Azad University
Nourabad Mamasani Branch
Nourabad Mamasani, Iran
ehsan.rahmani@mamasaniiau.ac.ir

***Abstract*—This paper proposes a reversible data hiding scheme for vector quantization (VQ)-compressed images based on search order coding (SOC). Data is embedded by choosing one of the possible ways to represent each embeddable index and outputted code is legitimate SOC code. The proposed scheme has two advantages. First it is more secure, comparing with previous SOC based schemes that generate non-legitimate codes as output. Second it can be used beside other schemes to embed more data and also it is a solution for the problem of transmission of side information of schemes which their outputs are legitimate VQ codes.**

Keywords-Image compression; Vector quantization; Search order coding; Data embedding

I. Introduction

One of the great challenges in today's fast developing internet is how to transfer secret data. Data hiding is a protective technique for secret data transmission. Secret data can be embedded into digital images, audios, videos… in such a way that it is undetectable to unauthorized interceptors. This paper considers data hiding in digital images.

Data hiding techniques are either irreversible or reversible. Irreversible data hiding techniques distort cover image in an unrecoverable manner. In some applications, such distortion may not be acceptable. In reversible or so called lossless data hiding techniques original image can be recovered after extraction of the embedded data.

Due to limitation of bandwidth and storage, typically images are transmitted in compressed formats. Vector Quantization (VQ) [1] is a widely used image compression technique with properties of simple structure and fast decoding process. Some methods have been proposed to further improve compression ratio of VQ. Search order coding (SOC) proposed in 1996 by Hsieh and Tsai [2] is one of them.

In recent years some VQ-based image data hiding schemes have been proposed that some of them have reversibility property [3-9]. Reversible VQ-based image data hiding techniques must be able to recover original VQ index table after extracting secret data. Some VQ based reversible data hiding schemes [3, 4, 7-9] generate non-legitimate codes as output and their outputs cannot be decoded by standard VQ (or VQ based such as SOC) decoders. However output of such schemes may raise attention of interceptors, therefore they are not suitable for steganographic applications.

In the existing schemes which generate legitimate code as output [5, 6], an index replacement strategy is used to embed data. Such schemes generate VQ codes with some artifacts in the image blocks and also some spots may appear in their resulting stego image. In the scheme proposed by Chang and Lin [5], three state codebooks are created for each block. First state codebook is created based on side match distortion. Codewords in the second state codebook are selected based on similarity with codewords of the first state codebook. Codewords in the third state codebook are selected based on similarity with codewords of the second state codebook from positions of the sorted codebook with zero hit rates. An index located in the first state codebook which its corresponding positions at the other two state codebooks are not null, is embeddable. An embeddable index isn't changed to hide secret bit 0 and it is replaced with corresponding index in the second state codebook to hide secret bit 1. To distinguish between indices which have been embedded by bit 1 and indices which originally located in the second state codebook, indices belong to the second case are replaced with their corresponding indices in the third state codebook. However multiple hit maps for various smoothness levels of blocks are used and they must be sent to the receiver as side information for recovery purpose. Recently Yang *et al.* [6] have improved this scheme.

Existing SOC based reversible data hiding schemes [3, 4] generate non-legitimate codes as output. This paper proposes a new data hiding scheme for SOC codes which its output is legitimate SOC code and it can be decoded by standard SOC decoder. So the existence of secret data is concealed.

In the new scheme, instead of changing indices, data is embedded by choosing one of the possible ways for representation of the indices. In this scheme indices will not change but the way to represent them may be changed. Hence, there is no quality degradation in stego image, and only compression performance is affected. However it is the cost must be paid.

The rest of this paper is organized as follows. In Section II, VQ and SOC are explained. The proposed scheme is presented in Section III. Section IV includes experimental results and discussions. Finally, Section V concludes the paper.

II. Background

A. Vector Quantization

VQ has three phases: training codebook, encoding and decoding. Codebook consists of some representative image blocks called codewords. VQ encoder partitions image to be encoded into non-overlapping blocks and for each image block searches for the closest codeword from codebook and then encodes that image block by the index of the closest codeword in the codebook. Decoding is done by a simple look-up table operation.

B. Search Order Coding

In SOC, for each index in the index table, it tries to find the same index around the current index. It searches indices appearing in a pre-defined path which are called search points (SP). If a match is found, then search order code of that index is used to encode that block, otherwise (i.e. if a match isn't found) original index value (OIV) must be used. An extra indicator bit is placed before each SOC code or OIV to distinguish between SOC codes or OIVs. Fig 1 shows an example of SOC path.

In a SOC, that each SOC code are represented in *m* bits, encoder searches the path until a match is found or number of distinct indices in the search path reaches 2^m.

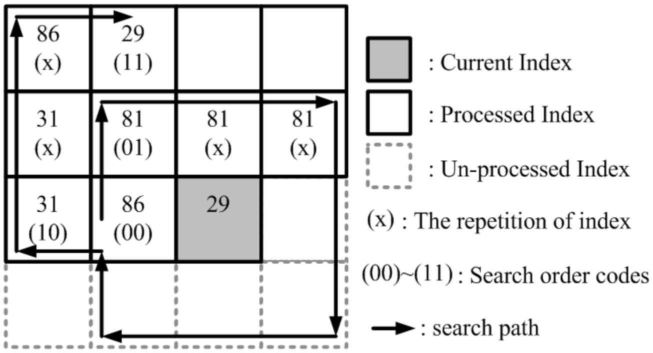

Figure 1. An example of SOC path

III. Proposed Scheme

The proposed scheme is based on choosing one of the index representation methods. For indices which aren't found in their corresponding SOC path, only one way is possible to represent them. For other indices (i.e. indices that are found in their corresponding SOC path), two ways are possible to represent, they can be encoded by SOC or by OIV. Therefore one bit can be embedded in such indices. Data embedding procedure is described in detail as follows:

Data embedding procedure:

```
Input: Index table $T$, secret message $S$, and parameters $m$
        and $n$ (number of bits used for each SOC code and
        OIV respectively).
Output: The SOC compressed code of the index table $T$ with
        embedded data $C$.
Process input index table $T$ in raster scan order. For each
index $I_i$:
    Search for index $I_i$ in the SOC path. Encode $I_i$ based on
    one of the following cases:
        Case 1 (if it is found):
            Retrieve next secret bit $b$ from $S$. Two cases are
            possible:
                Case 1.1 ($b = 0$):
                    Encode $I_i$ by its SOC code, i.e. indicator
                    bit 0 followed by $m$ bit SOC code.
                Case 1.2 ($b = 1$):
                    Encode $I_i$ by its OIV, i.e. indicator bit 1
                    followed by $I_i$.
        Case 2 (if it is not found):
            There is no data embedding. Encode $I_i$ by its
            OIV, i.e. indicator bit 1 followed by $I_i$.
```

Finally the compressed code and the codebook used in VQ compression procedure must be sent to the receiver. Obviously output of this scheme is legitimate SOC code and it can be decoded by standard SOC decoder. On the other hand, the proposed scheme does not change the indices of the input index table for the embedding purpose. When an unauthorized user decodes compressed code to the original index table by using standard SOC decoder and then decodes the obtained index table to an image by using standard VQ decoder, nothing will be suspicious.

Within data extraction procedure original index table is obtained and also it can be compressed to SOC code. Data extraction procedure is described in detail as follows:

Data extraction procedure:

```
Input: SOC compressed code of the index table $T$ with
        embedded data $C$, and parameters $m$ and $n$ (number
        of bits used for each SOC code and OIV
        respectively).
Output: The index table $T$ (or SOC compressed of the index
        table $T$), and secret message $S$.
While whole input code is not processed do followings:
    Retrieve next bit from input code as indicator bit $d$. Two
    cases are possible:
        Case 1 ($d = 0$):
            Concatenate bit 0 to the extracted secret data $S$.
            Retrieve next $m$ bits from input code as SOC
            code of current index $I_i$ and find out $I_i$ in the SOC
            path.
        Case 2 ($d = 1$):
            Retrieve next $n$ bits from input code as OIV of
            current index $I_i$ from input code. Search for $I_i$ in
            the SOC path. Two cases are possible:
                Case 2.1 (if it is found):
                    Concatenate bit 1 to the extracted secret
                    data $S$.
```

Case 2.2 (if it is not found):
I_i doesn't carry secret bit.

Reversible data hiding schemes which generate legitimate VQ codes as output [5, 6] usually have low capacity and in some cases they need some side information to achieve reversibility. Side information can be embedded along secret data before transmission. However side information occupies some embeddable spaces and reduces pure embedding capacity. On the other hand, side information may be needed before data extraction. The proposed scheme can be used to further improve capacity of such schemes and to embed side information of them.

IV. Experimental Results

This section includes some experiments and discussions on experimental results to verify the effectiveness and feasibility of the proposed scheme. For experiment, nine gray-level test images were used (Fig. 2). All images are of the size 512×512. For VQ encoding, each image were divided into 16,384 blocks of size 4×4 pixels. Two codebooks include 256 and 512 16-dimensional codewords were trained by using the well-known Linde–Buzo–Gray (LBG) [10] algorithm and four test images Airplane, Lena, Peppers and Toys involved in training codebook. A pseudo random number generator was used to form the data to be embedded.

Since embedding procedure of the proposed scheme does not make any distortion in the input index table, the quality of the stego image is not considered. Two aspects of performance are considered, i.e. hiding capacity and bit rate of compressed code. Bit rate is used to evaluate the compression performance, which is the ratio of the size of the output code to the number of pixels in the image. Bit rate is represented by bpp (bits per pixel).

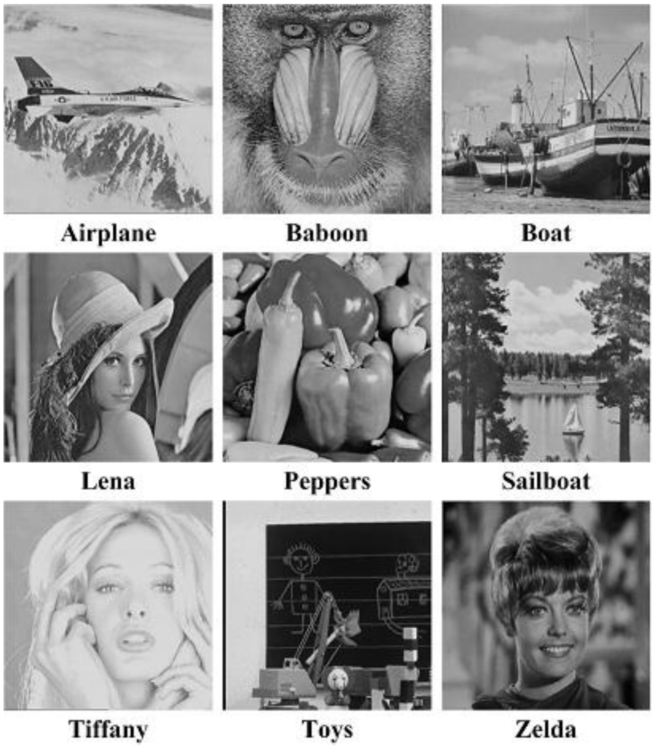

Figure 2. Nine VQ-compressed test images

Tables I and II illustrate bit rate of SOC compressed code without data hiding, payload (hiding capacity) and bit rate of the output code obtained by the proposed scheme for two values of *m* for codebook size 256 and 512 respectively.

The proposed scheme is compared with scheme recently developed by Yang *et al*. [6]. Fig. 3 shows this comparison for codebook size 512. Because for scheme [6] side information occupies some embeddable spaces, pure capacity is considered. It can be seen that in scheme [6] to provide large capacities, quality of stego image degrades sharply, where for the proposed scheme quality of stego image is the same as cover. It also should be noted that scheme [6] generate VQ code as output and its bit rate is 0.563 for codebook size 512, where bit rate of the output code for the proposed scheme usually is lower than it.

Table I. Results of the proposed scheme for codebook size 256

Images	VQ	*m* = 2			*m* = 3		
	PSNR	SOC rate (bpp)	Payload (bits)	Rate (bpp)	SOC rate (bpp)	Payload (bits)	Rate (bpp)
Airplane	31.49	0.329	10,208	0.447	0.346	11,348	0.456
Baboon	23.84	0.476	3,774	0.519	0.458	5,482	0.510
Boat	29.12	0.333	10,007	0.448	0.353	10,993	0.457
Lena	31.66	0.329	10,183	0.446	0.345	11,390	0.452
Peppers	31.90	0.343	9,586	0.453	0.357	10,787	0.461
Sailboat	28.35	0.360	8,844	0.461	0.372	10,011	0.467
Tiffany	29.62	0.289	11,941	0.427	0.312	13,113	0.437
Toys	31.84	0.268	12,857	0.416	0.307	13,371	0.434
Zelda	29.55	0.345	9,490	0.454	0.351	11,076	0.457

Table II. Results of the proposed scheme for codebook size 512

Images	VQ	*m* = 2			*m* = 3		
	PSNR	SOC rate (bpp)	Payload (bits)	Rate (bpp)	SOC rate (bpp)	Payload (bits)	Rate (bpp)
Airplane	32.38	0.405	8,223	0.515	0.404	9,642	0.514
Baboon	24.31	0.564	2,278	0.595	0.546	3,456	0.587
Boat	29.70	0.394	8,644	0.508	0.405	9,606	0.513
Lena	32.63	0.398	8,515	0.509	0.402	9,734	0.515
Peppers	32.85	0.422	7,619	0.524	0.422	8,859	0.524
Sailboat	28.93	0.435	7,097	0.529	0.434	8,356	0.529
Tiffany	30.47	0.356	10,082	0.490	0.367	11,268	0.486
Toys	32.68	0.340	10,675	0.483	0.348	12,123	0.486
Zelda	30.06	0.422	7,609	0.525	0.417	9,067	0.521

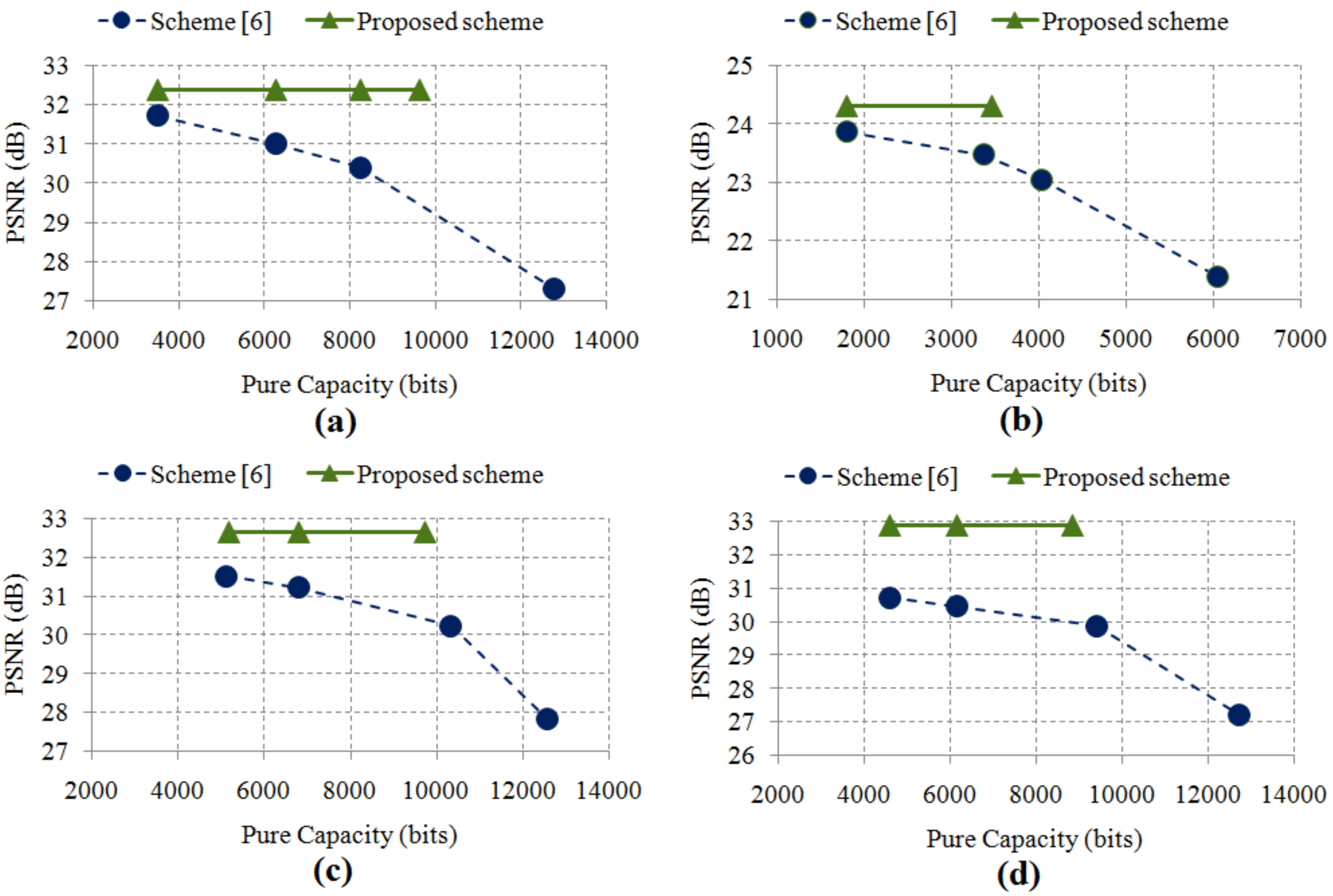

Figure 3. Comparison between scheme [6] and the proposed scheme (a) Airplane (b) Baboon (c) Lena (d) Peppers

V. Conclusion

This paper presented a reversible data hiding scheme for VQ-compressed images based on search order coding. As far as we know, the proposed scheme is the first SOC based reversible data hiding scheme which generates legitimate SOC code as output. In the proposed scheme instead of changing indices, choosing the way to represent them is used to embed data. The proposed scheme can be used beside other reversible VQ-based data hiding schemes to further improve capacity of these schemes and embed side information of them.

References

[1] R. M. Gray, "Vector quantization," *IEEE Acoustics, Speech and Signal Processing Magazine*, vol. 1, no. 2, pp. 4-29, April 1984.

[2] C. H. Hsieh, J.C. Tsai, "Lossless compression of VQ index with search-order coding," *IEEE Transactions on Image Processing*, vol. 5, no. 11, pp. 1579–1582, November 1996.

[3] C. C. Chang, G.M. Chen, M.H. Lin, "Information hiding based on search-order coding for VQ indices," *Pattern Recognition Letters*, vol. 25, no. 11, pp. 1253–1261, 2004.

[4] S. C. Shie, S.D. Lin, "Data hiding based on compressed VQ indices of images," *Computer Standards & Interfaces*, vol. 31, no. 6, pp. 1143–1149, 2009.

[5] C. C. Chang, C.Y. Lin, "Reversible steganography for VQ-compressed images using side matching and relocation," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 493-501, December 2006.

[6] C. H. Yang, W.J. Wang, C.T. Huang, S.J. Wang, "Reversible steganography based on side match and hit pattern for VQ-compressed images," *Information Sciences*, vol. 181, no. 11, pp. 2218-2230, June 2011.

[7] C. C. Chang, C.Y. Lin, "Reversible steganographic method using SMVQ approach based on declustering," *Information Sciences*, vol. 177, no. 8, pp. 1796-1805, April 2007.

[8] C. H. Yang, Y.C. Lin, "Reversible data hiding of a VQ index table based on referred counts," *Journal of Visual Communication and Image Representation*, vol. 20, no. 6, pp. 399–407, August 2009.

[9] P. Rahmani, G. Dastghaibyfard, "Reversible data hiding for VQ-compressed images based on an index replacement technique," in: *Proceedings of 3rd International Conference on Signal Acquisition and Processing*, Singapore, 2011.

[10] Y. Linde, A. Buzo, R.M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84-95, Janury 1980.