

(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 11, November 2015

Improved Integrated Data-Hiding and Compression Scheme for Digital Images using Vector Quantization (VQ) and Side Match

Aswathy Subash, Nimmy George

PG Scholar, Dept. of CSE, VJCET, Vazhakulam, Kerala, India Asst. Professor, Dept. of CSE, VJCET, Vazhakulam, Kerala, India

ABSTRACT: In this paper we propose an improved integrated data-hiding and compression scheme for digital images using vector quantization (VQ) and side match. The two image processing functions, data hiding and image compression, can be accomplished in a single module. On the sender side, except for the blocks in the leftmost and topmost of the image, each of the other residual blocks are used for embedding secret data and compressed simultaneously by VQ or side match. In side match method the similarity between neighbouring blocks is evaluated. The residual blocks are encoding by utilizing the redundancy between the image blocks. Huffman encoding is also applied to the resulting compressed code stream to further improve the compression performance. After segmenting the image compressed codes which is the result of Huffman decoding into sections by the indicator bits, the receiver can achieve the extraction of secret bits and image decompression successfully. Experimental results demonstrate the effectiveness of the proposed scheme.

KEYWORDS: Data hiding; image compression; vector quantization (VQ); Huffman coding; side match

I. INTRODUCTION

The advancement in the information and internet technology enable the people to transmit and share digital contents conveniently. In such a scenario ensuring communication efficiency and save network bandwidth, compression techniques can be implemented on digital content. In many applications, most digital content, especially digital images and videos are converted into the compressed forms for transmission. Nowadays large numbers of privacy related problems are there in an open network environment about how to transmit secret or private data securely. Information hiding techniques contributes many excellent solutions for this problem, which can embed secret data into the cover data imperceptibly.

Many researches have been conducted in the field of data hiding for digital images [2], [3], [4], [5], [7], [9]. However, in these methods, data hiding and image compression process are performed as two independent modules on the sender side. Under this circumstance, there may have many chances for an attacker to destroy or exploit the compressed image without the watermark information embedded, and performing data hiding and image compression independently may lead to lower efficiency in the applications. In 2014, Chuan Qin et.al proposed a novel joint data hiding and compression scheme based on SMVQ and image inpainting [1], which integrate data hiding and compression into single module.

Thus, in the proposed work, it is not only focus on improving the high hiding capacity and compression ratio, but also integrate the data hiding and the image compression into a single module seamlessly, which can avoid the risk of the attack from interceptors and increase the implementation efficiency. The image compression in IIDHC scheme is based mainly on the VQ and horizontal and vertical side match. Huffman encoding is performed with the resulting compressed code streams.

On the sender side, except for the blocks in the leftmost and topmost of the image, each of the other residual blocks in raster-scanning order can be embedded with secret data and compressed simultaneously by VQ or side match. In side match method the similarity between neighbouring blocks is evaluated. The residual blocks are encoding by utilizing the redundancy between the image blocks. Vector quantization is also utilized for some complex residual blocks.



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Huffman encoding is also applied to the resulting compressed code stream to further improve the compression performance. After segmenting the image compressed codes which is the result of Huffman decoding into a series of sections by the indicator bits, the receiver can achieve the extraction of secret bits and image decompression successfully according to the index values in the segmented sections.

II. RELATED WORK

Many researches are being done in the field of data hiding and compression for digital images which hide secret data into the compression code of the image. VQ [9] is one of the most popular and widely used compression techniques for digital images due to its simplicity, i.e. simple encoding and decoding, and cost effectiveness in implementation. During the vector quantization compression process, the Euclidean distance is utilized as a measure to evaluate the similarity between each image block and the codewords in the VQ codebook. The index of the codeword with the smallest distance is used ad a reference to represent the block. Thus, VQ encoding process results in a VQ index table. Instead of pixel values, only the index values are stored, therefore, the compression is achieved effectively. The VQ decompression process can be implemented easily and efficiently because each received index can be recovered by simple table lookup operation.

Many vector quantization based data hiding methods have been proposed so far [8], [9]. As an improved version of VQ, SMVQ has been introduced which make use of original VQ codebook and a sub-codebook to perform compression of digital images. Recently, many researchers have studied on embedding secret information by SMVQ [2]-[5], [7], [9]. In 2006 Chin-Chen Chang et.al proposed a reversible data hiding based on side match vector quantization [2]. In this work each block in the SMVQ compressed cover image was represented by appropriate codeword in the sub-codebook.

In most of these schemes data hiding and image compression are conducted as two separate modules which mean the two processes, image compression and the data hidings are two independent modules on the sender side. This will lead to some security problems and make it easy for an attacker to compromise the compressed image without the watermark information embedded. Also performing these two functions independently leads to inefficiency in applications. To solve this problem Chuan Qin et.al proposed joint data hiding and compression scheme based on SMVQ and image inpainting [1] which integrate the two process, data hiding and image compression, into a single module seamlessly.

In [1] all image block except for the non residual blocks (blocks in the leftmost and topmost of the image), each of the other residual blocks in raster-scanning order can be used for embedding secret data and compressed simultaneously by SMVQ or image inpainting. The compression methods change adaptively according to the current embedding bit. To control the visual distortion and error diffusion caused by the progressive compression of some residual blocks, vector quantization was used. The receiver can achieve the extraction of secret bits and image decompression successfully after segmenting the image compressed codes into a series of sections by the indicator bits, according to the index values in the segmented sections. The problem with this method is that, it was a threshold depended approach. i.e. The number of SMVQ blocks and inpainting blocks increases with *T*. The secret bits are only embedded in the SMVQ and inpainting blocks.

III. **PROPOSED METHOD**

In this paper, only a single module is utilized to realize image compression and secret data embedding simultaneously [1]. In improved integrated data hiding and compression scheme, the image compression is based on VQ and side match. In side match mechanism both horizontal and vertical side match methods are utilized for image compression. Secret bits are embedded into side match encoded image blocks. Huffman coding is also utilized to further improve the compression performance.

A. Overview

The improved integrated data hiding and compression scheme can be mainly divided into two phases. First phase is the image compression and secret data embedding and the second phase is the image decompression and secret data extraction.

In image compression and secret data embedding phase or at sender side, the original uncompressed input image I, is divided into the non-overlapping $n \times n$ blocks. The blocks in the leftmost and topmost of the input image I called as non-



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

residual blocks. These non-residual blocks are encoded by standard VQ method and these VQ encoded blocks are not used to embed secret bits. When the encoding process is complete, each non-residual blocks is simply represented by the index of the codeword with smallest distance.



Fig.1. Flow Chart of the Compression and Secret Data Embedding in the Residual Block.

The blocks except in the leftmost and topmost of the input image are called residual blocks. These blocks are encoded successively in raster scanning order, and their encoded methods switches between VQ or side match according to whether current block has any similarity between its top or left block or not. i.e., the residual blocks are used for embedding secret data and compressed simultaneously by VQ and horizontal and vertical side match. Fig.1 shows the compression and secret data embedding in the residual block.

Each residual block is encoded with respect to its top and left block. The current block denoted as $X_{i, j}$ is encoded using the index of the nearest code word of the block such that the total storage space for an image is significantly reduced. I.e. similarity of the current processing block with its top and left neighbouring blocks is considered. If the index values does not matches with index value of the up blocks or left block, then this block $X_{i,j}$ is encoded with standard VQ and no secret bits are embedded. Otherwise the block $X_{i,j}$ is encoded with horizontal or vertical side match depending up on whether it is similar to up block or left block and secret bits are embedded. Indicator bits are used to distinguish the blocks which are encoded using standard VQ and side match methods. Secret bits are embedded only in side match encoded residual blocks. The detailed procedure is described in section IV. For improving the compression performance the compressed code streams are again decoded with Huffman coding.

The second phase of the proposed system is the image decompression and secret data extraction, which is performed at the receiver side. In this phase the received compressed codes are first decompressed using Huffman decoding. Fig. 2 shows the flowchart of decompression and secret bit extraction for each residual block. Then in order to recover the original image and extracting secret bits, the compression code stream, which is the result of Huffman decoding, are divided into a no. sections depending on the indicator bits.

Explicitly, if the current indicator bit is 0, which means this section corresponds to a VQ compressed block with no embedded secret bit. So this section is decoded using VQ compression method. Otherwise, if the current indicator bit is



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

1, which means this section corresponds to horizontal or vertical side match. The original secret message can be extracted by decrypting the received secret bits by the security key. After all the segmented sections in the compressed codes complete the above described procedure, the embedded secret bits can be extracted correctly, and the decompressed image can be obtained successfully.



Fig. 2.Flow Chart of the Decompression and Secret Data Extraction in the Residual Block

IV. METHODOLOGY

This section gives the detailed technical description of the proposed system. The improved integrated data hiding and compression scheme using VQ and side match belongs to the area of image processing. The complete implementation of the system is done using MATLAB R2010a.

In this system instead of two separate modules only a single module is utilized to realize image compression and secret data embedding simultaneously. So the proposed IIDHC is realized mainly with two main modules as follows.

A. Image Compression And Secret Data Encryption

Vector quantization (VQ) [9] has widely been used for signal processing due to its excellent compression performance. Data hiding techniques in the VQ-compressed domain can relish advantages of both data hiding and compression for a multimedia distribution, achieving a secure channel and bandwidth/space saving for data transmission/storage. In this scheme, along with VQ, redundancy between current encoding block and neighboring block is utilized for better compression performance. This technique is known as horizontal and vertical side match. The detailed procedure is described as follows.



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

In this work, the sender performs image compression with the help of VQ code book containing codewords of length n^2 . Denote the original uncompressed input image I of size M×N. The image I is divided into the non-overlapping blocks of size n×n. In this method, it is assumed that M and N can be divided by n with no remainder. Denote all k divided blocks in raster scanning order as $X_{i, j}$, where $k = M \times N/n^2$, i = 1, 2...M/n, and j = 1, 2...N/n. The blocks in the leftmost and topmost of the image I called non-residual blocks, i.e., $X_{i,1}$ (i = 1, 2, ..., M/n) and $X_{1,j}$ (j = 2, 3, ..., N/n), are encoded by direct VQ method and these VQ encoded blocks are not used to embed secret bits. The VQ consists of three phases, namely, codebook construction, encoding process, and decoding process. In VQ compression process, a set of training images are used as a reference to generate a codebook using the LindeBuzoGray (LBG) algorithm [9], where the training images are divided into non-overlapping blocks and the most representative blocks are selected to generate codebook. The codebook consists of a set of elements called codewords. With this codebook, each block in the uncompressed input image is encoded using the index of the most similar codeword of the block. The squared Euclidean distance [8] used to measure the similarity between a block, $X=(X_1; X_2; ; ; ; ; X_K)$ and a codeword, $C_{tk}(C_{t1}; C_{t2}; ; ; ; ; ; C_{tk})$ is defined as eq (1), where C_t denotes the t_{th} codeword in the codebook.

$$D(X, C_t) = \sqrt{\sum_{i=1}^k (X_i - C_{t,i})^2}$$
(1)

At the end of the encoding process, each image block X is represented by the index of the codeword having smallest distance ie. index of the closest codeword in the codebook. This method can significantly reduce the total storage space when a codebook of a small size is employed. In the decoder process, with the same codebook as that used in the encoder, the original image is restored from the indices using a table-lookup operation. The same codebook containing T codewords is used by both sender and receiver in order to perform encoding and decoding.

The residual blocks are encoded successively in raster scanning order ie. from left to right and top to bottom. The encoding methods of these residual blocks are related to the whether current block has any similarity between its top or left block or not. i.e. The residual blocks can be embedded with secret data and compressed simultaneously by VQ and horizontal and vertical side match. Fig 1 shows the compression and secret data embedding in the residual block.

The current processing block can be denoted as $X_{i,j}$ ($2 \le i \le M/n$, $2 \le j \le N/n$). The left blocks of $X_{i,j}$ are $X_{i,j-1}$ and up blocks of $X_{i,j}$ are $X_{i,j-1}$. The current block is encoded using the index of the most similar code word of the block. The similarity between a block, $X_{i,j}$ and a codeword, $C_{tk}(C_{t1}; C_{t2}; ;;;; C_{tk})$ is defined as eq (1). If the index values does not matches with index value of the up blocks or left block, then this block $X_{i,j}$ sencoded with standard VQ and no secret bits are embedded. Otherwise the block $X_{i,j}$ is encoded with horizontal or vertical side match depending up on whether it is similar to up block or left block and secret bits are embedded. Note that if standard VQ is adopted, an indicator bit i.e., bit 0 should be added before the compression code of the VQ index for $X_{i,j}$. If side match is used, the indicator bit i.e., bit 1 is added ad prefix of the compressed code for $X_{i,j}$. If index value of current block is same as the index of the up block, which means that these blocks are redundant, then bit 1 is used to represent the block $X_{i,j}$ in the compressed code and along with this n/2 bits of secret data is appended. Else if index value of current block is same as the index of the left block, which means that these blocks are redundant, then bit 0 is used to represent the block $X_{i,j}$ in the compressed code and along with this n/2 bits of secret data is appended. In this scheme, before being embedding the secret data are scrambled by a secret key to ensure the security. So for encryption and decryption, RC4 encryption algorithm is used. So with side match only one bit is required to represent the current block which will reduce the length of compressed code for $X_{i,j}$ than using standard VQ.

After the current block $X_{i,j}$ is processed, the above same procedure is repeated for the remaining residual blocks until the last residual block is processed. Then, the compressed codes of all image blocks (residual and non-residual) are concatenated. For improving the compression performance the concatenated compressed code are again decoded with Huffman coding, which will further reduce the size of transmitting data there by saving bandwidth/space for data transmission/storage.

B. Image Decompression and Secret Data Extraction

The decompression and secret bit extraction of each residual block is performed at the receiver side. For that the received compressed codes are first decompressed using Huffman decoding. Huffman decoding is accomplished by a simple table look up operation.



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

At the receiver side, the receiver should first recover the (M + N - n)/n blocks in the leftmost and topmost of the image since they are used as reference to decompress the remaining residual blocks. These (M + N - n)/n blocks are decompressed directly by VQ indices which are retrieved from the image compressed codes. Each VQ index of each block occupies log2 T bits.

In the decoder process, with the same codebook as that used in the encoder, the original image is restored from the indices using a table-lookup operation. Then, the remaining blocks except the first (M + N - n)/n blocks are processed successively in raster-scanning order.

Fig. 2 shows the decompression and secret bit extraction for residual blocks. Then in order to recover the cover image and extracting secret bits, compression code stream, which is the result of Huffman decoding, are divided into a no. of sections according to the indicator bits.

If the current indicator bit in the compressed codes is 0, this indicator bit and the following $\log_2 T$ bits are partitioned as a section and this section corresponds to a VQ compressed block with no embedded secret bit. The decimal value of the last $\log_2 T$ bits in this section is exactly the VQ index and the block is recovered with respect to this decimal value. Otherwise, if the current indicator bit is 1, this indicator bit and the following \log_2 (2n) bits are then segmented as a section, which means this section corresponds to horizontal or vertical side match. If the indicator bit is 1 then read next bit in this section. If next indicator bit is 1 then this block similar to its upper block and is replaced with the VQ index of up block index that can be used directly to recover the block. If next indicator bit is 0 then this block similar to its left block and is replaced with the VQ index of left block index that can be used directly to recover the block. In both this case remaining bits in the section represent secret bits. The original secret message can be extracted by performing decryption on the received secret bits by the security key.

The above procedure is repeated for each segmented section until all sections are processed. After all the segmented sections in the compressed codes complete the above described procedure, the embedded secret bits can be extracted correctly, and the decompressed image can be obtained successfully. The final decompressed image does not contain the embedded secret bits any longer.



Fig.3. Standard Test Images.

V. EXPERIMENTAL RESULTS

The experiment was conducted on a set of gray-level images of size 512×512. Six standard test images, as shown in Fig. 3, are used for evaluating the performance of the proposed system. In addition to these six standard images, the uncompressed color image database (UCID) was also considered for evaluation process.

The complete implementation of the system is done using MATLAB R2010a programming environment. In the evaluation process, the input images are divided into non-overlapping 4×4 blocks. With respect to the size of image blocks, the length of codewords containing in the codebook was 16. The VQ codebook size T used for evaluation was 256, and secret bits for embedding were generated by RC4 encryption algorithm. After dividing the images into non-



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

overlapping blocks, the blocks in the topmost row and the leftmost column were compressed by VQ. The VQ encoded blocks are not used for embedding secret bit. These blocks are used as a reference for encoding remaining blocks. The secret bits are only embedded in the vertical or horizontal side match encoded blocks. Thus, the hiding capacity of the proposed scheme is equal to the sum of the numbers of vertical or horizontal side match encoded blocks. Fig. 4 shows the input image and output image after decompression by VQ and side match. Tables 1- 7 show the comparison resultsfor six standard test images such as Lena, Airplane, Lake, Peppers, Sailboat, Tiffany and UCID data set.

VI. **PERFORMANCE EVALUATION**

The performance of the proposed system is presented in terms of hiding capacity, compression ratio, PSNR and SSIM value. The hiding capacity of the proposed scheme was compared against the hiding capacity of Chuan Qin.et.al method [1]. In Chuan Qin.et.al method [1] the secret data hiding was performed depending up on the threshold and distortion value. In this method the total hiding capacity is equal to the number of SMNQ and inpainted blocks.

Denote the length of the compressed codes for the image as L. The compression ratio CR [1] can be calculated according to eq. (2). Peak signal-to-noise ratio (PSNR) [1] was utilized to measure the visual quality of the decompressed images Id, see eq. (3).

$$CR = \frac{8 \times M \times N}{L} (2)$$

$$PSNR = 10 \times \log_{10} \frac{255^2 \times M \times N}{\sum_{x=1}^{M} \sum_{y=1}^{N} [l(x, y) - l_d(x, y)]^2} (3)$$

Where M and N represent image size; I (x, y) and Id (x, y) represent the pixel values at (x, y) of the uncompressed image I and the decompressed image Id, respectively. For image quality assessment both PSNR value and the structural similarity (SSIM) value was used.



Fig.4.(a) Original Uncompressed Input Image, (b) Decompression Result

Fig. 4 shows the uncompressed input image and corresponding output image after decompression by VQ and side match. Tables 1 to 7 show the performance comparisons of the compression ratio, the visual quality of decompressed image, hiding capacity for six standard test images such as Lena, Airplane, Lake, Peppers, Sailboat, Tiffany and UCID data set. Note that the results of CR, PSNR, SSIM and HC in Table 7 are the mean values for a set of the images in the UCID database.



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Table 1: Comparison of performance measures for Lena

Schemes	CR	PSNR	SSIM	HC
JDHC	14.2523	29.2923	0.8347	33
Proposed	19.3316	29.2935	0.8349	5522

Table 2: Comparison of performance measures for Airplane

Schemes	CR	PSNR	SSIM	НС
JDHC	14.2554	27.6578	0.8644	397
Proposed	22.5069	27.6594	0.8647	7268

Table 3: Comparison of performance measures for Lake

Schemes	CR	PSNR	SSIM	НС
JDHC	14.2616	24.6629	0.8649	543
Proposed	19.6007	26.5247	0.8692	5742

Table 4: Comparison of performance measures for Peppers

Schemes	CR	PSNR	SSIM	НС
JDHC	14.2522	28.8759	0.7861	32
Proposed	20.355	28.783	0.7864	6287

Table 5: Comparison of performance measures for Sailboat

Schemes	CR	PSNR	SSIM	HC
JDHC	14.2616	28.7814	0.8625	594
Proposed	19.2824	29.3231	0.8645	5435

Table 6: Comparison of performance measures for Tiffany

Schemes	CR	PSNR	SSIM	НС
JDHC	14.2492	25.3505	0.7724	15
Proposed	19.3587	26.4376	0.7739	5449

Table 7: Comparison of performance measures for UCID

Schemes	CR	PSNR	SSIM	HC
JDHC	14.2542	2.9351	0.8030	268
Proposed	21.3623	25.812	0.8030	6860



(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

VII. CONCLUSION

This work proposes an improved integrated data-hiding and compression scheme using VQ and side match. The blocks, except for non residual blocks (those in the leftmost and topmost of the image), can be embedded with secret data. Here compression and data hiding is performed simultaneously and the compression method switches between VQ and side match according to redundancy among the neighbouring blocks. If the current encoding residual block is same as its upper block then vertical side match is used. If the current encoding block is same as its left block then horizontal side match is used. Secret data is scrambled by secret key to ensure the security. These encrypted secret data is placed on the vertical and horizontal side match blocks. To further improve the compression performance Huffman encoding is applied on the secret embedded compression codes of the image. The receiver can extract the original secret data and obtain the input cover image successfully after segmenting the Huffman decoded code stream into various sections during the decompression phase.

Since the proposed method utilizes the redundancy between the blocks it is capable of embedding two bits of secret data with each side match encoded blocks. Also the proposed system follows a threshold independent procedure for integrated data hiding and compression. The experimental results show that this scheme has the better performances for hiding capacity, compression ratio, and decompression quality.

REFERENCES

- C. Qin, C. C Chang and Yi-Ping Chiu, "A novel joint data-hiding and compression scheme based on SMVQ and image inpainting," *IEEE Trans. Image Process.*, vol. 23, no. 3, March. 2014.
- C. C. Chang, W. L. Tai, and C. C. Lin, "A reversible data hiding scheme based on side match vector quantization," *IEEE Trans. Circuits* Syst. Video Technol., vol. 16, no. 10, pp. 1301-1308, Oct. 2006.
- 3. J. D. Lee, Y. H. Chiou, and J. M. Guo, "Reversible data hiding based on histogram modification of SMVQ indices," *IEEE Trans. Informat. Forensics Security*, vol. 5, no. 4, pp. 638-648, Dec. 2010.
- 4. L. S. Chen and J. C. Lin, "Steganography scheme based on side match vector quantization," *Opt. Eng.*, vol. 49, no. 3, pp. 0370081-0370087, Mar. 2010.
- 5. J.D Lee and Y.H Chiou, "Lossless data hiding for VQ indices based on neighboring correlation," *IEEE Trans. Informat. Forensics Security*, vol. 21, no. 10, pp. 1389-1402, Oct. 2010.
- 6. P. Tsai, "Histogram-based reversible data hiding for vector quantization compressed images," *IET Image Process.*, vol. 3, no. 2, pp. 100-114, 2009.
- 7. J.D Lee, Y.H Chiou, and J.M Guo, "Information hiding based on block match coding for vector quantization compressed images," *IEEE Systems Journal*, vol.8,no. 3, pp. 737-748, Sept.2014.
- 8. NopparatPantsaena, M. Sangworasil, C. Nantajiwakornchai and and T. Phanprasit "image compression using vector quantization," *in Proc. IEEE Int. Conf. Signal Processing and Information Technology*, Ho Chi Minh, pp. 000158 -000163, Dec. 2012
- 9. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantization design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

BIOGRAPHY

Aswathy Subash received her B.TECH degrees from SNGCE, MG University, Kottayam, India in 2012. She is currently pursuing her M.TECH degree at VJCET, MG University, Kerala, India. Her area of interest includes image processing, data mining etc.

Nimmy George received her B.TECH degrees from VJCET, MG University, Kottayam, India in 2008 and her M.TECH degree from VJCET, MG University, Kottayam, Kerala, India in 2011. She is currently working as Asst. Professor at VJCET, Vazhakulam, Kerala, India. Her area of interest includescomputer networks.