Improved Data Hiding Scheme using VQ based Index Map Compression and Codebook Extension

P. Uma, S. Vimala

Abstract: Now-a-days, data protection has become inevitable for confidential transactions that happen over net. Transmission of such confidential data has become a challenging issue in today’s scenario. Many Data Hiding techniques are used for transmitting data in a secured way. The secret data can be hidden as part of any type of file such as Text file, Sound file, image file, Video file, etc. It has been proposed to hide secret data as part of cover image. The proposed work adopts Vector Quantization (VQ) which is one of the powerful and simple image compression techniques to compress the size of the cover image and to reduce the cost associated with storage/transmission. VQ transforms the cover image into its corresponding Codebook and IndexMap. The confidential data is then embedded as part of the Codebook and IndexMap. The proposed method helps in improving the performance by increasing the embedding capacity and coding efficiency. The performance of Steganography is improved in three levels. The embedding capacity of the proposed method is increased by 23,726 bits when compared to that of existing similar methods, which is a significant improvement.

Keywords: Data hiding, Embedding Capacity, Steganography, Vector Quantization.

I. INTRODUCTION

Steganography is a technique that enables us to transmit secret data in secured way by embedding the data as part of any media file such as Audio, Video or Image file. In the proposed method, it has been decided to embed the data as part of the image file. The image used for embedding the secret data is called a Cover image. Relatively large amount of data can be hidden using such techniques [1]. Generally image occupies more storage space and needs more time for transmission. To reduce the cost associated with storage and transmission of Stego-image, many image compression techniques such as VQ [2], Side Match Vector Quantization (SMVQ) [3], JPEG [4], and JPEG2000 [5] are used. Due to its simplicity and cost effective implementation, VQ finds its application in Steganography [6]. VQ based techniques such as Side Match Vector Quantization (SMVQ) are also used for data hiding [7,8,9,10,11,12,13,14,15]. In Vector Quantization, the input image is transformed into Codebook and IndexMap. The secret data can be hidden as part of Codebook and IndexMap without causing much distortion to the image. VQ comprises of three stages namely 1. Codebook Generation, 2. Image encoding and 3. Image decoding. The quality of the codebook that is generated improves the performance of VQ. In codebook generation phase, the input image is divided into blocks of size 4x4 pixels, which is an optimal size to maintain trade-off between quality and bit-rate of compressed images. Each block is transformed into a vector of size 16 elements \( \{c_1, c_2, \ldots, c_{16}\} \). Collection of such vectors is called Training Set (TS of size N). The desired set (of size M) of training vectors, called the representative vectors is selected from the Training Set and the collection of such vectors is called the Codebook. From the total set of vectors, representative vectors called codevectors are selected to form the codebook of size M, where \( M < N \). In image encoding phase, each input block is replaced with index \( i \) of the closest codevector \( c_i \). These indices collectively form the IndexMap. The distance between the codevector and the input vector is computed using the Euclidean distance measure using (1).

\[
d(x, y_j) = \sum_{j=1}^{k} (x_j - y_{ij})^2
\]  

(1)

In equation (1), \( x_j \) is the \( j^{th} \) component of the input block \( x \) (training vector) and \( y_{ij} \) is the \( j^{th} \) component of the \( i^{th} \) codevector \( y_i \). The codevector with which the minimum distance obtained is identified as the similar codevector.

\[
d(x, y_i) = \min\{d(x, y_j), y_j \in C\}
\]  

(2)

where \( y_i \) is the most similar codevector to the input vector \( x \). During the decoding phase, each index in the IndexMap is replaced with the corresponding codevector [16]. As a result of VQ, the cover image is converted into a set of \( CI = \{CB, IM\} \) where \( CI \) is the Compressed Image, a set of Codebook (CB) and an IndexMap (IM).

In the proposed method, the \( CB \) and \( IM \) are used for embedding the secret data as part of them. In general data hiding technique, the secret data is hidden as part of cover image, where the cover image takes more storage space that leads to extra overhead. In addition to embedding the secret data as part of cover image, VQ compresses the cover image that leads to reduction in storage cost. In section 2, similar works are discussed. In Section 3, proposed work is explained. In Section 4, Experimental results are discussed and in Section 5, the conclusion follows.

Revised Manuscript Received on February 29, 2020.

* Correspondence Author

P. Uma*, Research Scholar, Department of Computer Science, Mother Teresa Women’s University, Kodaikanal, India. Email: umaopoomi@yahoo.com

S. Vimala, Associate Professor, Department of Computer Science, Mother Teresa Women’s University, Kodaikanal, India. Email: vimalaharini@gmail.com

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-9 Issue-5, March 2020

DOI: 10.35940/ijitee.E2547.039520
II. RELATED WORKS

There are many techniques available to generate codebook [17, 18, 19, 20, 21]. One such technique is Simple Codebook Generation (SCG) [22]. The codebook generated using SCG is improved in terms of quality, using K-Means algorithm. The IndexMap created as a result of Image Encoding is further improved by incorporating the Search Order Coding (SOC) method to improve the coding efficiency [23].

In 2004, the steganography method proposed by Chang et al. [24] is the first one in which the VQ based SOC is used. This method faces the challenge of increased file size and limited capacity. In 2007, Chang et al. proposed a new scheme based on the frequency of the index count [25]. The frequency of each index, i.e. the number of occurrences of each index, is considered. Based on ranking frequency, the secret data is hidden as part of higher frequency indices which leads to improved embedding capacity.

Yang et al.’s scheme [26], the VQ codebook is sorted using referred counts and the codebook is divided into 2^k clusters. Here B denotes the size of the secret data that is embedded as part of each VQ index. The embedding capacity achieved with this scheme is better than a scheme of Chang et al.’s. This scheme proposed by Dushyant Goyal et al. [27] is a reversible data hiding technique based on VQ encoding. The embedding capacity is increased by increasing embeddable index values. The performance of this scheme in terms of embedding capacity is better than that of Yang et al.’s. In this scheme, the original VQ compressed index values can be recovered without any loss after extracting the secret data that is embedded.

This scheme proposed by Wang et al. [28] is based on Adjoining State Codebook Mapping (ASCM). This ASCM maps the contents of an image block to the index of the corresponding code vector in the Codebook. In this scheme, two state codebooks are created by using the adjacent image block of a current image block that is being encoded. So this leads to accurate representation of current image block. This leads to improved coding efficiency and increasing embedding capacity of secret data. In Lin et al.’s [9] is used two techniques of SOC and State Codebook Mapping (SCM) to reduce the VQ index size. This leads to embedded the large amount of secret data. This scheme outperforms Wang et al. in terms of embedding capacity and compression rate.

III. PROPOSED METHOD

The generated initial Codebook is optimized using K-Means Clustering technique to improve the quality of the codebook and the optimized codebook is used in the proposed methods. In this research paper, two different approaches are proposed to increase the embedding capacity and the complexity involved in decoding the secret data. In the first approach, the codebook is modified by replacing few of the elements of code vectors. In the second method, the embedding capacity is further improved by embedding secret data as part of IndexMap in addition to Codebook.

A. Steganography using Vector Quantization with Modified Codebook (SVQMCB)

The codebook consists of M number of codevectors, each of which has 16 elements \{c1, c2, c3, ..., c16\}. The optimal size of Codebook is generally 256 as it helps maintaining a trade-off between quality and coding efficiency. A codebook therefore contains 256 x 16 elements. In each codevector, 4 elements from predetermined positions (p) computed using (3) are replaced with four 8-bit secret data and the embedding capacity is thus 256 x 4 x 8 bits.

\[ p = i \mod 4 \quad \text{where} \quad p = 2 \quad \text{and} \quad 1 \leq i \leq 16 \quad (3) \]

In the receiving end, after retrieving the embedded data, gaps will be introduced. The inter pixel feature of images is exploited and hence these gaps are filled by taking the average of the neighboring values, which will be an approximation of the original value without much distortion in the codebook. The gaps introduced in the aforementioned positions are filled up with the values generated using (4).

\[ C_i = \begin{cases} (C_{i-1} + C_{i+1})/2 & \text{if} \quad (i \mod 4) = 2 \\ C_i & \text{Otherwise} \end{cases} \quad \text{where} \quad 1 \leq i \leq 16 \quad (4) \]

B. Steganography using Vector Quantization with Extended Codebook (SVQECB)

In this approach, the actual elements of codevector are remain untouched. Eight more elements are added to the end of each codevector. Pixel values from four desired positions (computed using (3)) are selected. The absolute differences between the selected pixel values and the four 8-bit secret data elements are computed using (5). These differences are padded to the codevector as c17, c18, c19 and c20.

\[ d_j = |C_i - s| \quad \text{where} \quad 17 \leq j \leq 20 \quad \text{and} \quad i = 2, 6, 10, 14 \quad (5) \]

Where \( C_i \) is the \( i \)th element of a codevector value and \( s \) is the secret data to be embedded as part of codebook. \( d_i \) is the absolute difference values. The elements \( C_{21} \) to \( C_{24} \) of a codebook are entered either as 1 or 0 depending on the sign of the difference. Now the manipulated codevector will have the elements as \( CV = \{ C_1, C_2, ..., C_{16}, C_{17}, C_{18}, C_{19}, C_{20}, C_{21}, C_{22}, C_{23}, C_{24} \} \) where \( C_1, C_2, ..., C_{16} \) are the actual pixel values of the original codevector. The elements \( C_{21}, C_{18}, C_{19} \) and \( C_{20} \) denote the absolute difference values and the remaining elements \( C_{21}, C_{22}, C_{23}, C_{24} \) are the indicator flags.

At the receiving end, while restoring the steganographed data, if flag is ‘1’, the ci value is subtracted from the original pixel value and if the flag is ‘0’, the ci value is added to the respective pixel value to get the secret data using equation (6). The identification of pixel values with which the data hidden is more complex when compared to that of the previous technique and also has an additional overhead of increased 32 bits for each codevector.

\[ SD = \begin{cases} P_i - c_i & \text{if} \quad \text{flag} = 1 \\ P_i + c_i & \text{Otherwise} \end{cases} \quad \text{where} \quad i = 17, 18, 19, 20 \quad (6) \]

where \( P_i \) is the pixel value and \( c_i \) is the absolute difference.
B.1 Algorithm

Encoding:

Step 1: Generate Codebook (CB) and refine using K-Means Clustering Algorithm.
Step 2: Generate IndexMap (IM).
Step 3: For each Codevector (CV) in CB, perform the following:
   Step 3.1: Identify the four-pixel positions in CV.
   Step 3.2: Compute the absolute difference between the selected pixels and four bytes of secret data.
   Step 3.3: Generate a flag sequence F depending on the nature of difference values.
   Step 3.4: Pad the codevector with the set of absolute differences and Indicator Flags.
Step 4: Store or transmit the modified CB and IM.

Decoding:

Step 1: For each codevector (CV) in CB, perform the following to retrieve the secret data back:
   Step 2: Extract secret data using the entire codevector using (6).
   Step 3: Retain the first 16 values for Image reconstruction.

C. Data Embedding in IndexMap using SOC

In this method, a novel idea of embedding the secret data in Index Map using Search Order Coding (SOC) Technique is introduced. When implementing SOC algorithm, the first row and first column of indices do not undergo any process as they do not have left and top neighbors. The remaining indices are read in Raster Scan Order. The index that is currently read is called Search Center (Ic) and is compared against the neighboring horizontal and vertical indices Ic, Io, It, and Ir, where Ic, Ix, and Iy indicate Left, Top Left, Top and Top Right indices. These Search indices are assigned the Search Order Codes 00, 01, 10 and 11 respectively. If Ic matches any of the neighboring indices of the Search-path, a bit sequence (of size 9) is generated by combining the following bits: 1 bit (an indicator bit that indicates if there is a matching neighbor), corresponding search order code of size 2 bits and the remaining 6 bits that represent the secret data to be hidden. Now, the original index is replaced with the new bit sequence that is generated. If there is no match between the Search Centre and its neighboring pixels, the Index Ic (generally 8 bits long) is just prefixed with the indicator bit ‘0’.

In Fig. 1, the highlighted cell is the Search Center and is equal to the top-left index. This index is encoded as 101xxxxx, where the first bit is the indicator bit, the next two bits represent search order code (01) and the remaining 6 bits represent the data to be hidden. This methodology leads to increased complexity in identifying the data.

C.1. Algorithm

Encoding

Step 1: Generate the IndexMap (IM) by encoding the cover image I using a VQ Encoder.
Step 2: Read the VQ index Ic from the IndexMap ‘IM’ that forms the Search Center.
Step 3: If Ic matches with any of SOCs, generate bit sequence using an indicator flag ‘1’ followed by the respective SOC and then concatenate with 6 bits of secret data. Replace Ic with the generated bit sequence.
Step 4: If no match found, prefix Ic with the indicator flag ‘0’.
Step 5: Repeat the above steps till all the indices one by one.

Decoding

Step 1: For each index in the IndexMap, perform the following:
Step 2: If the first bit is 0, the index Ic does not contain any secret data. Go to Step 5.
Step 3: Else, the 2nd and 3rd bits are used to retrieve the Index from the respective position.
Step 4: Get the secret data from the last six bits.
Step 5: Repeat the above steps till all the indices of IndexMap are processed.

IV. EXPERIMENTAL RESULT AND DISCUSSIONS

All experiments are performed with the benchmark gray scale images of size 512 x 512 pixels. The input images are Airplane, Lena, Boats, Man, Peppers, Sailboat, Zelda, Girl and Couple. Fig. 2 shows the input images taken for the study. The algorithms are implemented using MATLAB R2014b. The Hardware used is the Intel® Pentium® 1.90 GHZ processor with 8GB RAM.

<table>
<thead>
<tr>
<th>Index Map</th>
<th>Ic</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>56 (01)</td>
</tr>
<tr>
<td>108</td>
<td>38 (00)</td>
</tr>
<tr>
<td>83</td>
<td>91</td>
</tr>
<tr>
<td>114</td>
<td>83</td>
</tr>
</tbody>
</table>

Fig. 1. Search Order Coding

(a) Airplane  (b) Lena  (c) Boats

(d) Man  (e) Peppers  (f) Sailboat
The bit-rate is used to analyze the coding efficiency of any Compression algorithm. The bitrate is computed using (9), where C represents the size of the compressed image in terms of bits. M x N is the size of the original image in terms of rows and columns. The bitrate obtained is represented in terms of bits per pixel (bpp). The lower the bitrate, higher the coding efficiency is.

\[
\text{bitrate} = \frac{C}{M \times N} \text{bpp} \tag{9}
\]

The PSNR values, bitrate and embedding capacity achieved with the proposed methods are shown in Table I. In SVQMCB, average PSNR and bitrate obtained are 34.38 and 0.69 respectively. In SVQECB, the average PSNR and bitrate are 34.46 and 0.72 respectively. In the existing method, the secret data has been embedded only as part of IndexMap. But in our proposed methods, the data has been embedded as part of both Codebook and IndexMap that leads to increased embedding capacity. On an average, 8192 bits of secret data are embedded in Codebook and 39063 bits of data in IndexMap, which is a significant improvement when compared to that of similar existing techniques. Hence the total average embedding capacity obtained with our proposed method is 47255 bits. For visual comparison, the reconstructed images of our proposed methods are given in Fig. 3 and Fig. 4.

Table I: Comparison of PSNR, Bitrate and Embedding Capacity with respect to the proposed Methods: SVQMCB and SVQECB

<table>
<thead>
<tr>
<th>Images</th>
<th>SCG [22]</th>
<th>Proposed Methods</th>
<th>SVQMCB</th>
<th>SVQECB</th>
<th>Embedding Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>BPP</td>
<td>PSNR</td>
<td>BPP</td>
<td>PSNR</td>
</tr>
<tr>
<td>Airplane</td>
<td>35.75</td>
<td>0.63</td>
<td>35.59</td>
<td>0.69</td>
<td>35.75</td>
</tr>
<tr>
<td>Lena</td>
<td>36.12</td>
<td>0.63</td>
<td>36.03</td>
<td>0.69</td>
<td>36.12</td>
</tr>
<tr>
<td>Boat</td>
<td>34.16</td>
<td>0.63</td>
<td>34.01</td>
<td>0.69</td>
<td>34.16</td>
</tr>
<tr>
<td>Man</td>
<td>33.38</td>
<td>0.63</td>
<td>33.35</td>
<td>0.69</td>
<td>33.38</td>
</tr>
<tr>
<td>Peppers</td>
<td>35.19</td>
<td>0.63</td>
<td>35.10</td>
<td>0.69</td>
<td>35.19</td>
</tr>
<tr>
<td>Sailboat</td>
<td>33.51</td>
<td>0.63</td>
<td>33.39</td>
<td>0.69</td>
<td>33.51</td>
</tr>
<tr>
<td>Zelda</td>
<td>37.74</td>
<td>0.63</td>
<td>37.73</td>
<td>0.69</td>
<td>37.74</td>
</tr>
<tr>
<td>Girl</td>
<td>36.87</td>
<td>0.63</td>
<td>36.85</td>
<td>0.69</td>
<td>36.87</td>
</tr>
<tr>
<td>Couple</td>
<td>33.73</td>
<td>0.63</td>
<td>33.55</td>
<td>0.69</td>
<td>33.73</td>
</tr>
<tr>
<td>Baboon</td>
<td>31.24</td>
<td>0.63</td>
<td>31.21</td>
<td>0.69</td>
<td>31.24</td>
</tr>
<tr>
<td>Average</td>
<td>34.77</td>
<td>0.63</td>
<td>34.68</td>
<td>0.69</td>
<td>34.77</td>
</tr>
</tbody>
</table>

The bitrate and the embedding capacity obtained with the proposed methods are compared with that of the results obtained with the existing techniques proposed by Chang et al., Yang et al. and Dushyant Goyal et. al. in Table II.

Table II: Comparison of Proposed scheme and the existing techniques with respect to Bitrate and Embedding Capacity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BPP</td>
<td>Embedding Capacity</td>
<td>BPP</td>
<td>Embedding Capacity</td>
</tr>
<tr>
<td>Lena</td>
<td>0.63</td>
<td>16003</td>
<td>0.63</td>
<td>26980</td>
</tr>
<tr>
<td>Airplane</td>
<td>0.63</td>
<td>16003</td>
<td>0.60</td>
<td>28840</td>
</tr>
<tr>
<td>Peppers</td>
<td>0.60</td>
<td>16003</td>
<td>0.61</td>
<td>28818</td>
</tr>
<tr>
<td>Zelda</td>
<td>0.65</td>
<td>16003</td>
<td>0.65</td>
<td>27304</td>
</tr>
<tr>
<td>Average</td>
<td>0.63</td>
<td>16003</td>
<td>0.62</td>
<td>27986</td>
</tr>
</tbody>
</table>
On an average, Chang et al. [24] achieved an embedding capacity of 16,003 bits (uniform for all images) and the proposed method (adaptive) leads to an embedding capacity of 55091 bits, which is 244.25% more. The Yang et al. [26] achieved an embedding capacity of 27986 bits and the proposed method leads to an embedding capacity of 55091 bits, which is 96.53% more. The DushyantGoyal et al. [27] achieved an embedding capacity of 31365 bits and the proposed method leads to an embedding capacity of 55091 bits, which is 75.20% more than that of DushyantGoyal et al. There is a significant improvement in Embedding capacity of the proposed method. The performance efficiency of the proposed method in terms of Embedding Capacity is represented graphically in Fig. 5.

Table III. Comparison of experimental results of our proposed scheme with Wang et al. and Lin et al. in terms of Embedding Capacity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>44101</td>
<td>48096</td>
<td>54104</td>
</tr>
<tr>
<td>Fit</td>
<td>37781</td>
<td>47456</td>
<td>55664</td>
</tr>
<tr>
<td>Baboon</td>
<td>19305</td>
<td>12422</td>
<td>34731</td>
</tr>
<tr>
<td>Peppers</td>
<td>34721</td>
<td>44985</td>
<td>64904</td>
</tr>
<tr>
<td>Boat</td>
<td>23181</td>
<td>29276</td>
<td>43824</td>
</tr>
<tr>
<td>Girl</td>
<td>22106</td>
<td>35067</td>
<td>51314</td>
</tr>
<tr>
<td>Toys</td>
<td>49201</td>
<td>70477</td>
<td>56697</td>
</tr>
<tr>
<td>Tiffany</td>
<td>34991</td>
<td>53048</td>
<td>52476</td>
</tr>
<tr>
<td>Average</td>
<td>31923</td>
<td>42603</td>
<td>51714</td>
</tr>
</tbody>
</table>

Table III shows that the maximum embedding capacity achieved with the proposed method is 64904 bits with ‘Peppers’ image and the minimum capacity is 34731 bits with ‘Baboon’ image. Wang et al. [28] and Lin et al.’s [9] have come out with 9305 bits and 12422 bits respectively for the Baboon image. For Toys image, the proposed method has achieved with the embedding capacity of 56697 bits, which is 7496 bits more than that of Wang et al.’s (49201 bits), but slightly less than that of Lin et al.’s (70477 bits).

In an average, the embedding capacity achieved with the proposed method is 51714 bits, which is better than that of Wang et al.’s (31923 bits) and Lin et al.’s (42603 bits). The results are graphically compared in Fig. 6.
Improved Data Hiding Scheme using VQ based Index Map Compression and Codebook Extension

V. CONCLUSION

This paper proposes both reversible and irreversible data hiding schemes based on Vector Quantization. In the first method, being irreversible, secret data is embedded as part of the codevector. Pixels in any four predetermined positions are used to embed secret data. In extracting phase, the replaced pixel values are interpolated by taking the average of neighboring values. This method is irreversible. In the second method, that is reversible, the size of the codebook is increased by padding the difference values and flag values as part of the codevectors that leads to more complexity in getting back the secret data. The difference and flag values are used to recover the data from the compressed data. The usage of SOC as part of IndexMap improves the coding efficiency of VQ and also leads to more embedding complexity. It has been observed from the results that the proposed method outperforms the existing steganography methods in terms of embedding capacity. In an average, the embedding capacity achieved with the proposed method is 51714 bits, which is better than that of Chang et al. (16003 bits), Yang et al. (27986 bits), DushyantGoyal et al. (31365 bits), Wang et al. (31923 bits) and Lin et al. (42603 bits). In our proposed method, the bitrate is little higher than the existing method. In future, the proposed work may be improved to reduce the bitrate without affecting the quality of the Cover Image and to improve the embedding capacity.

REFERENCES

AUTHORS PROFILE

P. Uma, pursued her UG Degree in M.K.University, P.G. Degree in Anna University, M.Phil. Degree in Mother Teresa Women’s University. She is doing her research in Mother Teresa Women’s University. She has authored 10 research papers in reputed International Journals. Her research focus is on Image Compression and Steganography.

S. Vimala, pursued her UG Degree in M.K.University, P.G. Degree in Bharathiar University and M.Phil. & Ph.D. Degrees in Mother Teresa Women’s University. She has authored nearly 40 research papers in reputed international journals and a book on ‘C Programming’. She is a life member of IACSIT. She has been awarded the ‘Best Thesis Award’ by ICFAI Business School, Hyderabad and Michigan State University, USA. Her research focus is on Image Compression, Image Analysis and Cloud Computing. She has twenty years of teaching experience and 10 years of research experience.