An Efficient VQ-based Data Hiding Scheme Using Voronoi Clustering

Ming-Ni Wu
Dept. of Information Management,
National Taichung Institute of Technology,
Taichung, Taiwan, ROC.
mnwu@ntit.edu.tw

Puu-An Juang
Dept. of Electrical Engineering,
National Chung-Hsing University,
Taichung, Taiwan, ROC.
puuanjuang@dragon.nchu.edu.tw

Yu-Chiang Li
Dept. of Computer Science and Information Engineering,
Southern Taiwan University,
Tainan, Taiwan, ROC.
lyc002@mail.stut.edu.tw

***Abstract*—In this paper, we propose a vector quantization (VQ) -based information hiding scheme that cluster the VQ codeowrds according the codewords' relation on Voronoi Diagram (VD). The clustering method is fast and easy to implement when projecting the codewords into VD space. The hiding capacities can be adaptive adjusted just modifying the threshold value which didn't need reorganizing the codebook sequence. Besides, the image quality is acceptable when raising the embedding capacity. Experiments demonstrate that the method can efficiently hiding the secret data into VQ indexes by the fast hiding procedure, high hiding capacity and good image quality.**

Keywords-data hiding; Voronoi diagram; vector quantization

I. INTRODUCTION

Image data hiding is a technique that embeds the secret message into a publishable image called cover image and the embedding result call stego-image is as similar as the original cover image to avoid the attention of grabbers. Hiding information in compressed image is more important because most of people compress images before they are stored or transmitted to saving the storage space or the transmission bandwidth. Vector quantization (VQ) [5] is an efficient image compressing method due to the high compression ratio and the easy and fast decompression process. A VQ codebook is pre-generated and shared between compression and decompression parts. The original image is divided into non-overlapping blocks firstly. Each block looks up the codebook to find the nearest codeword. The nearest codeword indexes for all blocks construct an index table as the compression result. In the decompression procedure, each index in the index table directive picks the corresponding codeword from codebook to reconstruct the image.

Some VQ-based image hiding schemes are proposed. Mean gray-level embedding (MGLE) is the simplest scheme that sorting the codewords by the mean value and separate the adjacent codewords pair into two groups. One group is assigned the secret value 0 and the other group is assigned the secret value 1. A codeword in a group can be replaced by the corresponding codeword in another group according the embedding secret value. The drawback of MGLE is the grouping method is not good enough. An improved technique which sorting the codebook by principle component analysis (PCA) is produced by Chang and Lin [3], we call it DHPCA method. Another improved scheme is introduced by Due and Hsu [4], we called it ACE method, which grouping codewords by clustering technique. The hiding capacity and image quality are both improved by ACE. However, ACE method need rearranging the codebook repeatedly for different secret content, different cover image size, and different hiding capacity. Furthermore, the rearranging procedure of ACE method is very time-consuming. In this paper, we will propose a codebook clustering method by a Voronoi Diagram (VD) [1]. The VD partitions the space into some planes which are convex polygons according the site points. Each point in a given plane is closer to its site pint than to any other. The detail algorithm please refers to [6]. By clustering codewords on VD planes, we can easy decide the groups and the group size can be adjusted by the distance between a codeword and its site point. Hence, we can adaptive embed different secret data into cover image to avoid the re-clustering process.

The rest of this paper is organized as follows. Section 2 introduces the details of our proposed scheme. Experimental results are illustrated and discussed in Section 3. Finally, conclusions are given in Section 4.

II. PROPOSED METHOD

The main idea of our proposed method is to clustering the VQ codewords according it's location on VD planes. The clustering VQ codebook and a predefined threshold *TH* therefore become the embedding accordance in our information hiding procedure. There are three phases in our scheme, codeword clustering phase, image encoding and information hiding phase, and image decoding and secret extracting phase. The details of them will be described as following paragraphs.

In the first phase, an original VQ codebook $C = \{c_0, c_1, \cdots, c_{N-1}\}$ with N codewords is generated using the well known LBG algorithm. Each codeword $c_i = (c_{i,0}, c_{i,1}, \cdots, c_{i,l-1})$ is a $l = n \times n$ dimensional vector; where n is the high and width of an image block. We project each codeword into a two-dimensional vector according the PCA algorithm [2]. Hence, a two-dimensional array $P = \{p_0, p_1, \cdots, p_{N-1}\}$ is produced from C, where, $p_i = (p_i^1, p_i^2)$. Next, P is sorted in descending order, that is

$P' = \{p', p'_1, \cdots, p'_{N-1}\}$ and the corresponding sorted codebook is $C' = \{c'_0, c'_1, \cdots, c'_{N-1}\}$. It means each codeword c'_i is corresponding to a point p'_i. We pick up M vectors from P' as the VD site points $P'' = \{p''_0, p''_1, \cdots, p''_{M-1}\}$, where $M \le N$. The residual points belong to a set P^*, where $P' = P^* \cup P''$, are clustered into M VD planes. Finally, we collect this M clustered result as M groups $G = \{G_0, G_1, \cdots, G_{M-1}\}$, where $P' = G_0 \cap G_1 \cap \cdots \cap G_{M-1}$. It denotes there are M groups in the VD. The first group is G_0, the second group is G_1, and so on. Each group contains one site point and some other non-site points. Figure 1. shows the processing steps for this phase. An example is illustrated in Figure 3(a)-(b). Let a codebook C with N=32 codewords is projected to the set P and the sorted result of P is P'. We select M=8 site points from P' to construct the VD as these dark circles in Figure 3(a). The remainder 24 points shown as dark triangles in Figure 3(b) are clustered into each VD plane according there locations on VD. It forms eight groups, for example, group G_0 contains one site point p_0 and five non-site points in Figure 3(b). Finally, the codebook C' remarked with group number and site point number for each codeowrd is produced.

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

The block diagram of embedding phase is shown as Figure 2. A predefined threshold TH decides the substitution range. For each group, we can decide which codeword is in the range and which codeword is out of the range by the Euclidean distance measure between a codeword c'_i and the corresponding codeword c'_j of its site point as Equation (1).

$$d(c'_i, c'_j) = \sum_{k=0}^{l-1} (c'_{i,k} - c'_{j,k})^2 \qquad (1)$$

If $d(c'_i, c'_j)$ is less than TH, c'_i is in the range, otherwise, c'_i is out of the range. Figure 3(c) shows the range boundaries by dash circles. For example, a dash circle of group G_0 covers a site point p''_0 and four non-site points. There are five corresponding codewords are inner the range and a codeword outside the range in group G_0. If a codeword is inner the range, it will follow the inner cluster embedding process, otherwise, it follows the inter cluster embedding process. For example, a codeword c'_j which derived from an image block B_i is belong to group G_0 and is judged belonging to in the range area. There are five codewords in the range and $b = \lfloor \log_2 5 \rfloor$ bits can be embedded in the block as Figure 4. The predefined label according the codeword number sequences of these five codewords are 00_2, 01_2, 10_2, 11_2, and 00_2 respectively (not exceeds 2^b). Let the secret bits r=10_2 and the label of codeword c'_k is just equal to r. c'_k replaces c'_j as the encoding result for B_i. Because the distance between c'_j and c'_k is in the range, the distortion caused from this exchanging is small. The codewords out of the range still can embed secrets. Each VD planes can be previously assigned a secret label 0 or 1 as the white number based dark circles in Figure 3(d). The label in plane of p''_0 is 0, the label in plane of p''_3 is 1, and so on. Let consider the codeword c'_5 belonging to G_0 is an out of range point. It is not suitable joining the substitution with other in range points due to the distortion is too larger. If the distance measure between c'_5 and c'_{20} is less than TH, they are exchangeable. If we want embedding a secret bit 0 into c'_5, it doesn't need any change for c'_5 because the plane label is 0. If we want embedding a secret bit 1 into c'_5, it can exchange with c'_{15} or c'_{20}, because the distance measures are less TH and the plane label of c'_{15} and c'_{20} is equal to 1. This is called inter cluster embedding and the embedding example is shown in Figure 5. Another instance is the codeword c'_{18}. We just look up the surrounding planes of c'_{18} and find the distances between it and the surrounding outside points c'_5, c'_6, c'_9, c'_{20} all are too large. So c'_{18} cannot embed any secret under this threshold.

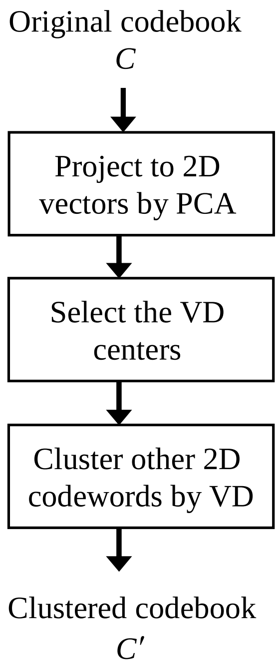


Figure 1. Block diagram of codeword clustering phase

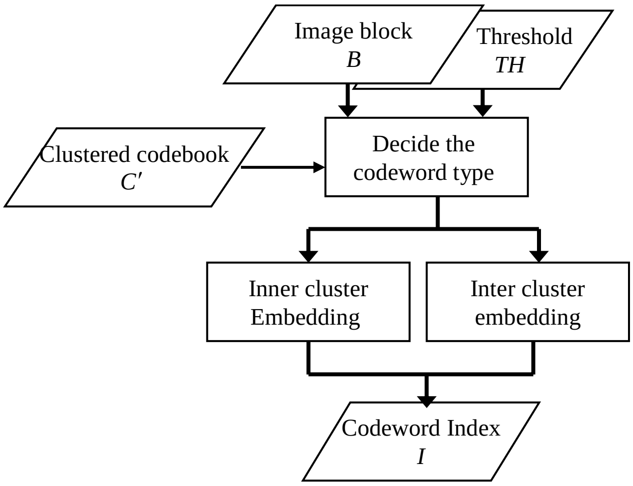


Figure 2. Block diagram of embedding phase

The secret extracting is very straightforward. The threshold *TH* can be seen as a key between sender and receiver for embedding and extracting secret data. The receiver does the same procedure as sender to produce C' and its grouping message. Let an index is k picking from the received index table. We look up the codebook's information about codeword c'_k and find the corresponding label is 10_2, the secret is extracted and the image block is reconstructed using codeword c'_k.

III. EXPERIMENTAL RESULTS

To demonstrated the performance of our proposed scheme. We compare our scheme (DHVC) and other three methods: MGLE, DHPCA, and ACE. A codebook with 512 codewords was generated using the well known LBG algorithm. Six cover images with size of 512×512 are divided into 4×4 non-overlapping blocks. Therefore, each cover image has 16384 blocks. A set of secret bits was generated by a random number generator for our experiments.

In our first experiment, we compare the image quality after embedding 16384 secret bits for six cover images as Table 1. Our scheme is better than other schemes. The VQ column is the compressed only by VQ which not embedded any secret.

Another comparison is given in Table 2. It shows the image quality under various capacities for image "Lena". The visual quality for hiding 48K secret is shown in Figure 6. The MGLE method is a fixed capacity hiding method, only 16K secret can be embedded. Other schemes all can adaptive embedded different secret data. Our image quality is better than MGLE and DHPCA. Besides, our image quality is better than ACE for low hiding capacity, such as 16K and 32K. For the high hiding capacity, out scheme is nearly the same as ACE. Note that, our method for hiding different capacities just need modifying the range threshold in compression procedure using the same clustered codebook. However, ACE must rearrange the clustered group for each variation which is more time-consuming. Therefore, our method has a better performance generally.

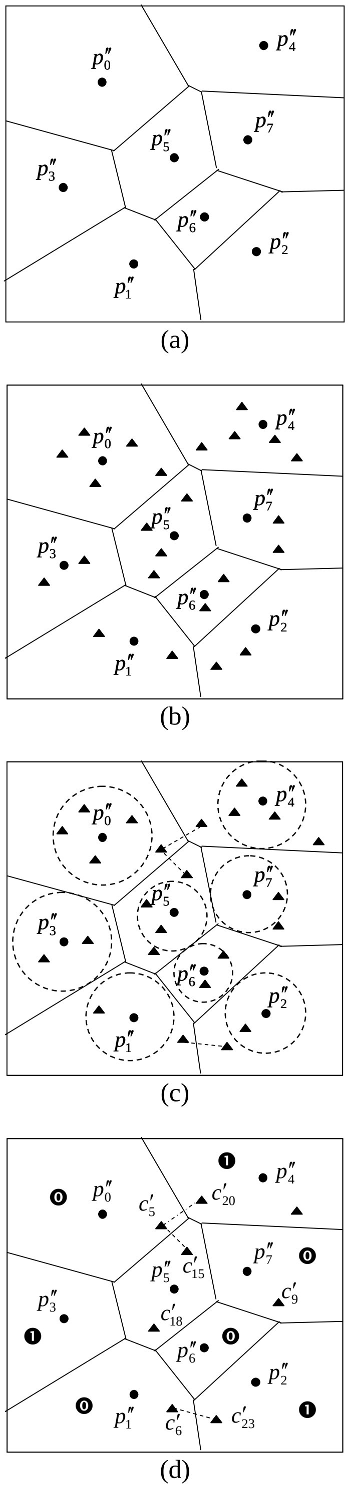


Figure 3. Examples of VD clustering

IV. CONCLUSIONS

In this paper, a VQ-based data hiding scheme is proposed. The clustering process is fast by projecting codewrds into VD space. The clustered codeword can be reused for different cover image, different hiding content, and different hiding capacity only adjusting the range threshold. The experiments show our proposed scheme can hide large amount secret data as DHPCA and ACE schemes. The image quality is better than DHPCA and about the same as ACE. To sum up all the discussion, our proposed scheme had better performance.

ACKONWLEDGMENT

This study was supported by the National Science Council, Taiwan, grant NSC97-2221-E-218-046.

REFERENCES

[1] Aurenhanmmer, F., “Voronoi Diagrams-a Study of a Fundamental Geometric Data Structure,” ACM Computing Surveys, Vol. 23, 1991, pp. 345-405.

[2] Chang, C. C. and Chan, C. S., (2003): “A Watermarking Scheme Based on Principal Component Analysis Technique,” Informatica, Vol. 14, No. 4, 2003, pp. 431-444.

[3] Chang, C. C., and Lin, P. Y., “A Compression Based Data Hiding Scheme Using Vector Quantization and Principle Component Analysis,” Proceedings of the 2004 International Conference on Cyberworlds, Tokyo, Japan, 2004, pp. 369-375.

[4] Du, W. C., and Hsu, W. J., “Adaptive Data Hiding Based on VQ Compressed images,” IEE Proceedings on Vision, Image, and Signal Process, vol. 150, No. 4, 2003, pp.233-238.

[5] Lindy, Y., Buzo, A., and Gray, R. M., “An algorithm for Vector Quantizer Design,” IEEE Transactions on Communications, Vol. 28, No. 1, 1980, pp. 84-95.

[6] Wu, M. N., Lin, C. C., and Chang, C. C., “Hiding Secret Information in VQ Compressed Images Using Voronoi Diagram,” Proceedings of The 2nd International Symposium on Voronoi Diagrams in Science and Engineering, Seoul, Korea, 2005, pp. 55-61.

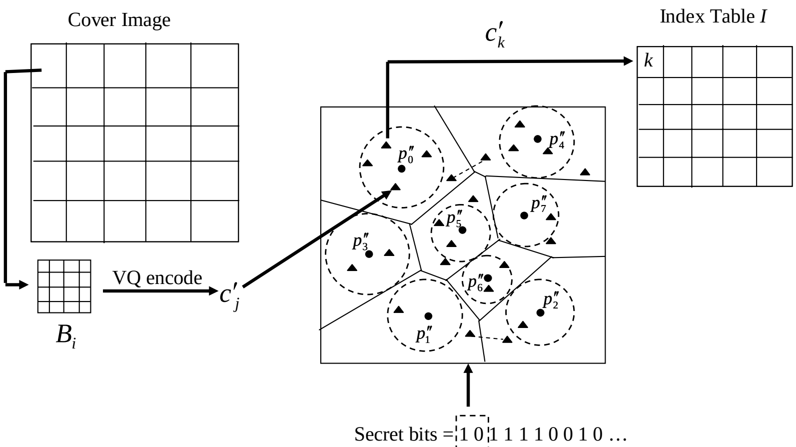


Figure 4. An example of inner clustering embedding

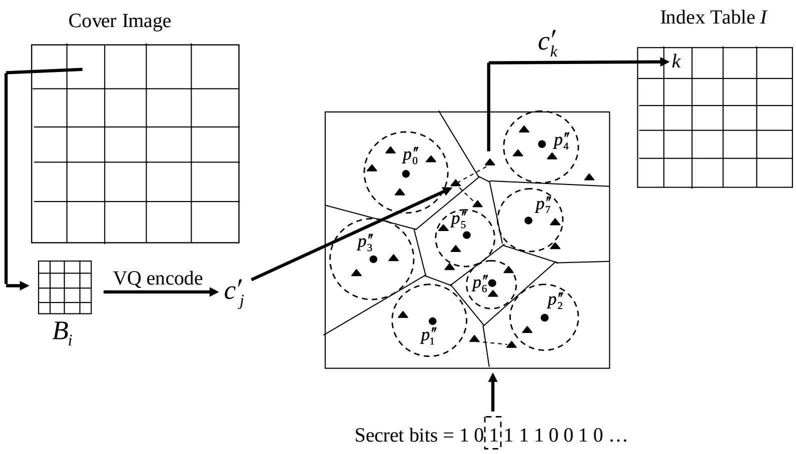


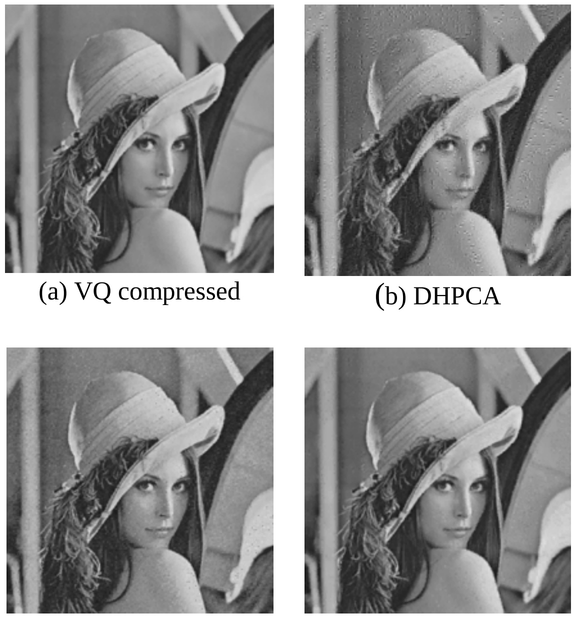
Figure 5. An example of inter clustering embedding

Table 1. IMAGE QUALITY (PSNR) COMPARISONS OF FOUR SCHEMES WITH 16384 SECRET BITS

Images	VQ	MGLE	DHPCA	ACE	DHVC
Lena	32.24	24.46	25.55	29.81	31.12
Airplane	31.58	25.28	26.84	28.30	30.83
Boat	30.16	22.34	26.32	28.43	29.16
Barbara	26.39	21.53	24.64	25.37	26.01
Peppers	31.41	25.28	28.51	28.43	30.32
Toys	31.16	23.78	27.98	28.59	29.66

Table 2. IMAGE QUALITY (PSNR) COMPARISONS UNDER VARIOUS CAPACITIES WITH IMAGE "LENA"

Capacity	MGLE	DHPCA	ACE	DHVC
16K	24.96	29.33	29.81	31.12
32K		26.58	28.59	29.56
48K		25.64	27.61	27.53
64K		23.12	26.40	26.14



(a) VQ compressed (b) DHPCA

(c)ACE (d) DHVC

Figure 6. Embedded results for different hiding schemes