VQ BASED DATA HIDING METHOD FOR STILL IMAGES BY TREE-STRUCTURED LINKS

Hisashi IGARASHI, Yuichi TANAKA, Madoka HASEGAWA, and Shigeo KATO

Graduate School of Engineering, Utsunomiya University
7-1-2 Yoto, Utsunomiya, Tochigi, 321-8585 Japan
e-mail : (igarashi@mclaren., tanaka@, madoka@, kato@)is.utsunomiya-u.ac.jp

ABSTRACT

In this paper, we propose a data embedding method into still images based on Vector Quantization (VQ). In recent years, several VQ-based data embedding methods have been proposed. For examle, '*Mean Gray-Level Embedding method* (*MGLE*)' are '*Pair wise Nearest-Neighbor Embedding method* (*PNNE*)' are simple, but not sufficiently effective. Meanwhile, an efficient adaptive data hiding method called '*Adaptive Clustering Embedding method* (*ACE*)' was proposed, but is somewhat complicated because the VQ indices have to be adaptively clustered in the embedding process. In our proposed method, output vectors are considered as nodes, and nodes are linked as a tree structure and information is embedded by using some of linked vectors. The simulation results show that our proposed method indicates higher SNR than the conventional methods under the same amounts of embedded data.

Index Terms— *Data Hiding, Vector Quantization, Tree-Structured Links*

1. INTRODUCTION

In a vulnerable communication network, there are potential threats of wiretapping or tampering of messages. An enciphered communication system is one of the most effective methods to be free from such threat, but enciphered texts are usually randomly arranged or meaningless bit streams. Consequently, the network attacker can easily notice that monitored messages are enciphered. Once they notice the messages are enciphered, skillful attackers may challenge to decrypt these enciphered messages.

On the contrary, data-hiding techniques have been developed to transmit the information in secret [1], [2]. In data-hiding techniques, the information is embedded into decoy data and then transmitted. In general, the amount of decoy data is much larger than that of the information data, so that the attacker does not notice this fact.

In recent years, several data-hiding techniques using Vector Quantization (VQ) have been proposed [3]-[5]. In the common VQ-based data embedding methods, embedding is performed by changing the vector index corresponding to the embedding information bit. However, these methods are not sufficient enough to obtain high quality embedded images. Embedded images are sometimes severely degraded because vectors which are not close to the original vectors are selected for embedding.

In our method, vectors used for embedding are constructed by selecting a pair of the nearest vectors and linking each other. Embedding is performed by choosing output vectors from the original one or the one nearest to it. In our method, it is expected that the highest SNR is obtained among embedding methods by changing the VQ indices. The simulation results show that our proposed method indicates higher SNR than the conventional method under the same amounts of embedded data.

2. CONVENTIONAL VQ-BASED DATA EMBEDDING METHODS

In this section, we introduce two representative data hiding methods based on VQ [5]. The simplest is the mean gray-level embedding method (MGLE), which permutes the code vectors according to their mean gray-scale values. Although the mean gray levels for adjacent code vectors are similar, individual components may differ greatly. Another method is the pair-wise nearest neighbor embedding method (PNNE), which repeatedly applies the nearest neighbor rule to select a pair of nearest code vectors, until all code vectors have been chosen. However, both of these embedding methods bring out the large distortion in the embedded images, because the distribution of all code vectors is not uniform, and therefore the merging of two code vectors into one group is not suitable for effective data hiding.

For example, Fig. 1 shows linked results of output vectors using the MGLE and PNNE methods in 2D vector space. If output vector c is used for embedding, it causes large distortion because the counterpart of the vector c is f in the MGLE and d in the PNNE, respectively, and these norms are so large.

In the next section, we propose a new method to reduce image quality degradation.
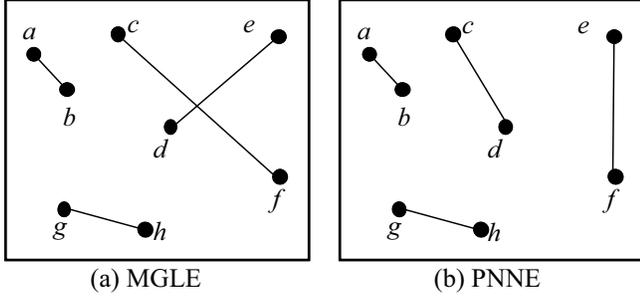

(a) MGLE (b) PNNE

Fig.1 Pairwise relationships of VQ in conventional methods

3. PROPOSED DATA EMBEDDING METHOD

In our method, the output vectors used in the original image (called *the first approximation vector: FAV*) and the nearest neighbor vector of each FAV (called *the second approximation vector: SAV*) are paired up, then the information bits are embedded using the paired vectors with small norm. Our method looks similar to the conventional PNNE method. However, in our method, link process is performed by using one of the whole vectors, or including the already linked vectors at preceding steps, as companion of pair for the residual vectors. This is the constitutive difference from PNNE. Even if the output vector is changed to embed the information, the nearest vector to the original FAV is always used, and therefore, our method obtains the highest SNR among embedding methods by changing the VQ index. Details of our method are explained in the following section.

3.1. Selection of the candidates of the embedding vectors

Fig. 2 shows the selection process of the candidates from output vectors to embed the information bits. The black dots *a* to *h* mean the output vectors in 2D vector space. Assuming that the original VQ image is I, embedding information bit sequence is $B = \{b_1, b_2, \ldots, b_w\}$, where $b_i = 0$ or 1 and the output vectors (*VQ code book*) is $Y = \{y_1, y_2, \ldots, y_p\}$. In example of Fig. 2, p equals to 8. At first, the pair of minimum norm (y_i, y_j) is searched in the VQ code book and linked. In Fig. 2(a), $y_i = a$ and $y_j = b$, and connected each other with the solid line.

The selection process explained in the following is repeated until satisfying the equation (1), where $|y_i|$ and $|y_j|$ are frequencies of y_i and y_j in the image, respectively.

$$T = |y_i| + |y_j| \geq w \qquad (1)$$

The equation (1) is the termination condition of the link generating process and T means its parameter. The link generation process is continued until satisfying the equation (1). In this process, two vectors with minimum norm are searched among the unlinked vectors, and linked. The com-

panion vector is selected from the whole vectors, or including the already linked vectors at preceding steps. In Fig. 2(a), after vector *a* was linked to vector *b*, the minimum norm vector to unlinked vectors *c*, *d*, *e*, *f*, *g* and *h* is searched from the VQ code book. Assuming that y_k and y_l are just searched vectors, y_k is unlinked yet, but y_l is either unlinked yet or linked already.

Fig. 2(b) shows the case of $y_k = c$ and $y_l = b$, where $y_l (= b)$ is already linked and $y_k (= c)$ is not linked yet in the preceding step.

If the y_k would not be linked, parameter T is recalculated in equation (2).

$$T \Leftarrow T + |y_k| + |y_l| \qquad (2)$$
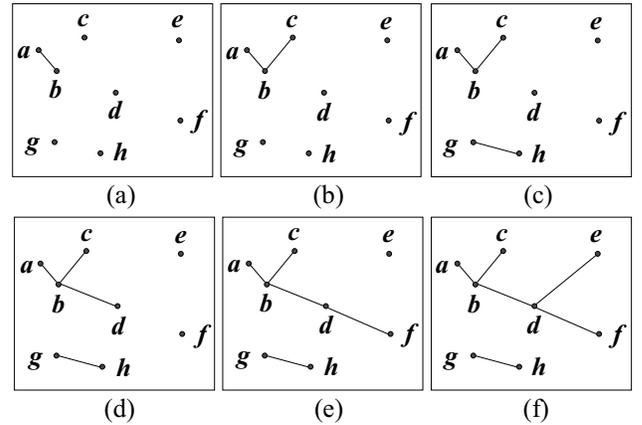

(a) (b) (c)

(d) (e) (f)

Fig. 2 The selection process of the candidates from output vectors

On the other hand, y_k is already linked, T is recalculated in equation (3).

$$T \Leftarrow T + |y_k| \qquad (3)$$

This process is repeated until satisfying the equation (4).

$$T \geq w \qquad (4)$$

Thus generated linked vectors are used to embed the information bits.

Fig. 2(c) to (f) shows the link generating process. In the Fig. 2(c), output vectors *g* and *h* are newly linked by judging the size of the norm. Fig. 2(d) shows the case that three vectors are connected to vector *b*. Even in such case, there causes no closed loop as long as complying with the link generation rule as mentioned above, therefore, the information bit assignment for the output vectors can be uniquely determined. The link generating process is repeated to satisfy the equation (4). Fig. 2(f) shows the state that the equation (4) is satisfied.

3.2. Embedding the information bits

The linked vectors obtained by above mentioned process are used to embed the information bits. At first, assign linked vectors to information bits. Bit 0 or 1 is assigned alternatively to the linked vectors in each link as shown in Fig. 3. In Fig. 3, 0 or 1 given in the side of each vector corresponds to the embedding bit 0 or 1 and labeled numbers i, ii, …, vi in the side of each link mean the ascending order of the norm size of corresponding two vectors.

Again, the vectors to embed the information have tree structures and do not have any loop, so that the embedding bit assignments cause no contradiction.

Embedding is performed by the following process. Linked vectors are only used to embed the information bits and unlinked vectors are not used and not changed. At first, vector indices in the original VQ image are scanned from the upper left corner of the image and search the linked vector. The first bit in the embedding information sequence is embedded in the link. In our method, since VQ indices are changed according to the information bits, it is hard to embed in ascending order of the norm of two vectors. However, vectors with larger norm are never used for embedding. It is because the VQ indices corresponding amounts of the embedding information are selected in ascending order of the norm and embedding is performed by using the candidate vectors in the predetermined order of the original VQ image. Consequently, SNR of the embedded images by this method is the highest among the VQ based data hiding methods which embed information by changing vector indices.
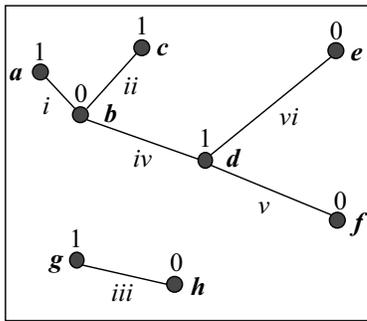

Fig. 3 Vector assignment for embedding bits

4. SIMULATION RESULTS

This section presents some experimental results concerning data embedding. Experimental fidelity is maintained by embedding random data into standard images, such as '*Lena*' and '*Goldhill*', as depicted in Fig. 4(a) and Fig. 5(a), respectively, each of which is 512×512 pixels, with 8 bit/pixel. Both images are divided into 16,384 blocks of 4×4 pixels for VQ encoding. A codebook of size 512 is determined using the well-known LBG algorithm [6]. For comparison, the following two algorithms, MGLE and PNNE methods as described briefly in Section 2, were implemented for the comparison. Fig. 4 and Fig. 5 present the embedding results

with hiding capacity of 16 kbits. Figure captions in Fig. 4 and Fig. 5 mean that (b) embedded images by MGLE, (c) embedded images by PNNE, and (d) embedded images by the proposed method, respectively.
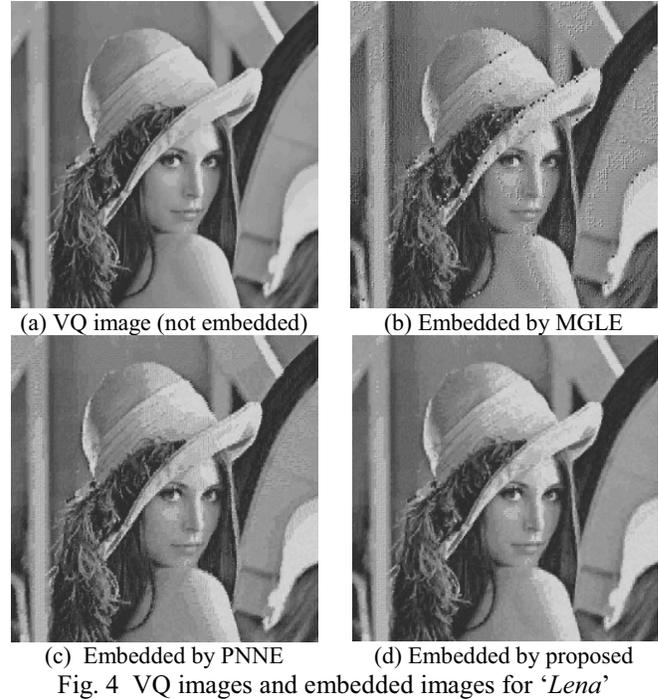

(a) VQ image (not embedded)   (b) Embedded by MGLE

(c) Embedded by PNNE   (d) Embedded by proposed
Fig. 4 VQ images and embedded images for '*Lena*'

(a) VQ image (not embedded)   (b) Embedded by MGLE

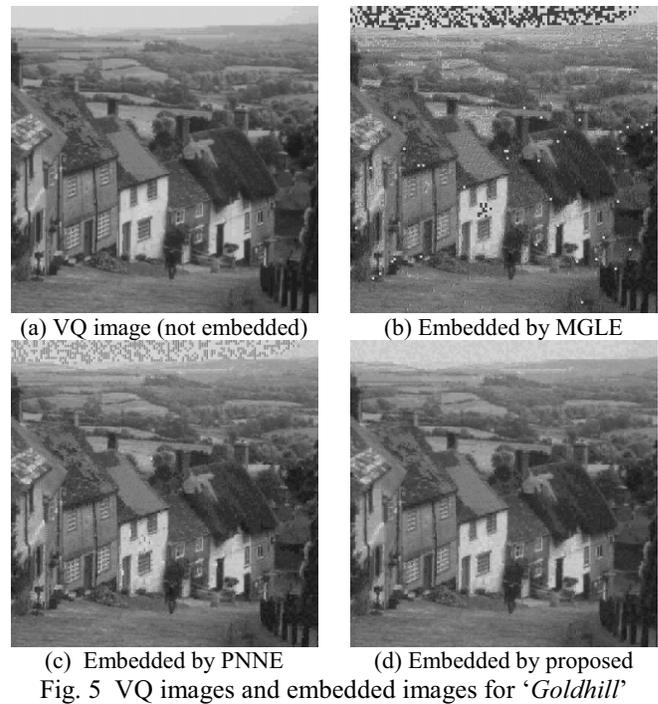(c) Embedded by PNNE   (d) Embedded by proposed
Fig. 5 VQ images and embedded images for '*Goldhill*'

These experiments indicate that the MGLE algorithm yields poor results since its algorithm causes larger distortion than those of PNNE and our proposed method.

To consider difference of image quality among these methods, the enlarged parts of processed images '*Lena*' by each method are shown in Fig. 6 at the embedding rate of 16 kbits. In Fig. 6, image degradation can be observed at the cheek part of '*Lena*' in PNNE method. In contrast, there can be seen little image degradation in our method.

(a) VQ image (not embedded)   (b) Embedded by MGLE

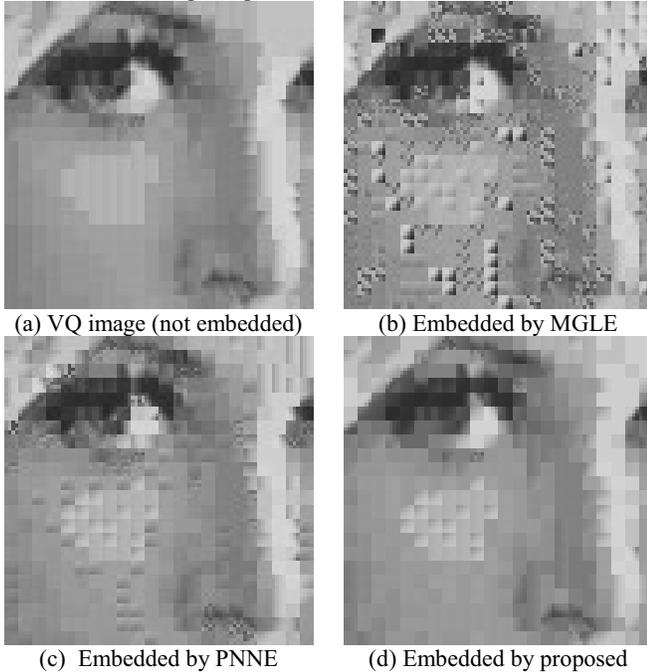(c) Embedded by PNNE   (d) Embedded by proposed
Fig. 6 Enlarged part of '*Lena*'

Table 1 and Table 2 show the SNR of the image embedded by each algorithm via VQ image before embedding. Difference of SNR between PNNE and the proposed method for '*Lena*' is very slight up to 8 kbits of embedding rate but our proposed method shows higher SNR than PNNE over 8 kbits of the embedding rate.

5. CONCLUSIONS

In this paper, we proposed a VQ based effective data hiding method. Our proposed method can obtain the highest SNR among the VQ based data hiding methods which embed information by changing vector indices. The simulation results show that our proposed method indicates higher SNR than conventional methods under the same amounts of embedded data. Our method takes much time to link the output vectors compared to MGLE and PNNE. That is, the numbers to calculate the norm of vectors decrease two times a linking process in MGLE and PNNE, meanwhile in our method these numbers does not decrease as well as the conventional ones. However, according to our supplementary simulation increase of the processed time is not so much, and the improvement effect of SNR outweighs this drawback.

Table 1 SNR of embedded image '*Lena*'

[dB]

Embedded Bits [Kbit]	MGLE	PNNE	Proposed
1	36.49	50.67	50.67
2	31.15	47.30	47.30
4	27.32	43.40	43.66
8	25.33	38.29	39.46
12	24.30	33.81	36.14
16	22.85	29.11	31.69

Table2 SNR of embedded image '*Goldhill*'

[dB]

Embedded Bits [Kbit]	MGLE	PNNE	Proposed
1	37.71	50.73	51.27
2	35.58	47.20	47.61
4	33.26	43.52	43.61
8	29.50	38.56	39.42
12	27.57	35.26	36.17
16	18.73	26.14	32.62

REFERENCES

[1] K. Matsui and M. Iwakiri, "The fundamentals of Information Hiding -Techniques of Information security in Ubiquitous Age-," Morikita Pub., 2004. (*in Japanese*).

[2] H. Harashima (*Ed.*) "Image Information Compression," Ohmsha Ltd., 1991. (*in Japanese*).

[3] Y. C. Lin and C. C. Wang, "Digital images watermarking by vector quantization," Proc. 9th National Computer Symposium, Taichung, Taiwan, pp. 76-87, 1999.

[4] Z. M. Lu and S. H. Sun "Digital image watermarking technique based on vector quantization," Electron. Letter, vol.36, no.4, pp. 303-305, 2000.

[5] W. C. Du and W. J. Hsu, "Adaptive data hiding based on VQ compressed images," IEE Proc.-Vis. Image Signal Process., vol.150, no.4, pp.234-238, Aug. 2003.

[6] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantization design," IEEE Trans. Commun., vol.28, no.1, pp. 84-95,1980.