Hierarchical Finite-State Vector Quantization for Image Coding

Ping Yu and Anastasios N. Venetsanopoulos, Fellow, IEEE

Abstract—The hierarchical finite-state vector quantization (HFSVQ) introduced in this paper is an improvement of the finite state vector quantization combined with hierarchical multirate image coding. Based on an understanding of the perception of human eye and the structural features of images, the HFSVQ technique employs different coding rates and different numbers of the predictive states for representative vector selection. The bit rate used to encode images is very low while the reconstructed images can still achieve a satisfactory perceptual quality.

I. INTRODUCTION

VQ (vector quantization) [1] is an efficient spatial domain image coding technique at rather low bit rate, say, less than 1 b/pixel (bpp). A vector quantizer is a mapping of k-dimensional Euclidean space $R^k$ into a finite subset (codebook) of $R^k$. An image block of k pixels $X_k$ will be replaced by a representative vector (codeword) in the codebook. Since only the codeword addresses are going to be transmitted or stored, the bit rate $R$ for the image coding with VQ can be calculated by

$$R = \frac{\log_2 N_c}{k} \text{ bpp}$$

where $N_c$ is the codebook size.

Various improvements of the VQ techniques have been developed in order to achieve further successful compression as well as keeping satisfactory qualities of the reconstructed image. The main focus of these improvements is based on the following two approaches. One approach attempts to increase the denominator of (1) by emphasizing the features of the different regions in an image, such as the variable block vector quantization [2] and the hierarchical multirate vector quantization (HMVQ) [3] where the blocks with different gray-scale transition features are assigned into different layers. Variant bit rates are applied for the different layer coding. By using this technique, the bit rate for smooth region coding can be considerably reduced. The other approach attempts to reduce the address bits, i.e., the numerator of (1) by forming a subset of the codebook for representative vector selection. A typical techniques is the finite state vector quantization (FSVQ) [4], [5], which uses the previously encoded blocks to build a candidate category for the optimal representative vector by evaluating the correlation of the adjacent pixels across block boundaries. FSVQ also alleviates the annoying "blocky" structure effect in memoryless VQ.

In this paper, a new coding technique, the hierarchical finite-state vector quantization (HFSVQ) is introduced, which is developed on the basis of the above VQ techniques. In the HSFVQ, an original image is decomposed into blocks of different sizes, which are then assigned into different layers, according to their grey-scale contrast (the largest grey-scale difference within a block). The blocks with low contrast, which are located in a smooth region of the image, will have large sizes and be assigned into higher layers. Fewer codewords are chosen to be candidates for the representative vectors in the higher layers, since there are strong correlations between adjacent pixels in smooth regions. For blocks located in edge regions where the grey-scales vary dramatically from pixel to pixel, all the codewords of the corresponding subcodebook are to be involved for the representative vector selection, in case that there may exist edges on block boundaries. The codes for the image coding with the HFSVQ consist of two parts. One part consists of the structure codes which provide the information of layer assignment of the image blocks. A related structure map is constructed based on this information. The other part consists of the local address codes or indexes of the codewords in the codebook. The address of the codewords for each block coding is rearranged on the basis of their occurrence probabilities. It is proved that this coding scheme is more efficient than conventional VQ and FSVQ techniques as we can adjust the reconstruction accuracy in different regions and achieve a further bit rate reduction.

This paper is organized as follows. Section II presents the structure map construction and layer assignment. Section III introduces the scheme of the HFSVQ for the block coding based on the layer information. In Section IV, the HFSVQ for color image coding is discussed. Section V presents experimental results and the conclusions are given in Section VI.

II. STRUCTURE MAP CONSTRUCTION

Any monochrom image can be divided into several regions according to its gray-scale contrast. The regions with low contrast correspond to the smooth area of the image and the regions which have edges with different sharpness demonstrate relatively high gray-scale contrast. VQ techniques generally result high-frequency quantization errors around edges and low-frequency ones in smooth areas. The frequency errors may cause the sharp edges blurred, which seriously damages
the perceptual quality in the reconstructed image, while the distortions in the smooth area, if there is no block effect, may not be apparently observed by human eye. On the other hand, the adjacent pixels in the smooth area are strongly correlated to each other as their gray-scales vary very slowly. According to the individual features of the different regions, we segment the original image with blocks of different sizes and construct a corresponding multilayer structure map.

Several layers are constructed to collect the blocks located in the different image regions. The higher layers have the blocks in smooth regions and the lower ones have the blocks in fluctuating regions. Different block sizes are designed to increase coding efficiency. The layer assignment can be conducted as follows.

First, we decompose the whole image into a group of blocks of size 16 × 16. If the gray-scale contrast in a block is lower than a given threshold, the block is assigned into layer 1 (L1), which collects the blocks located in the smoothest regions of the image. Then we decompose the rest of the blocks into subblocks of size 8 × 8. The same threshold is used and blocks located in a fairly smooth regions are assigned into layer 2 (L2). Finally, we further decompose the remainder into blocks of size 4 × 4 and assign them to the edge layer, layer 3 (L3).

In order to obtain a more satisfactory perceptual result, we need to provide a more accurate edge reconstruction. Therefore a special layer for edges, layer 4 (L4) is constructed from layer 3. Layer 4 collects the blocks with large gray-scale transition by using a threshold with higher value, and then excludes the isolated blocks by concerning them as trivial textures.

We describe the layer assignment procedure by a structure tree shown in Fig. 1. The assignment result can be demonstrated as a structure map shown in Fig. 2. A group of overhead codes, structure codes, are needed to record the structure map based on the branch distribution of the structure tree. The symbol "*" in Fig. 2 indicates that there will be a local address code inserted between the structure codes. It can be calculated that a block of size 16 × 16, which belongs to layer 1, needs 1 b structure code; a block belonging to layer 2 needs 1.25 b; an 8 × 8 block which consists of four 4 × 4 subblocks all belonging to L3 needs 2.25 b and another kind of 8 × 8 block, which consists of four 4 × 4 subblocks and at least one of them belonging to L4, is encoded with 6.25 b. The bit rate for the structure map can be calculated by the following expression:

$$R_{ST} = N_1 + 1.25N_2 + \frac{2.25(N_3-N_4)}{4} + \frac{6.25(N_4+N_2)}{4} \text{ bpp}$$

(4)

where \(N\) is number of the total pixels of the image; \(N_i, i = 1, 2, 3, 4\), is the number of the blocks in layer \(i\); \(N_0\) is the number of the blocks of layer 3 which form an 8 × 8 blocks with the participation of the blocks of layer 4. It is obvious that the larger number of the blocks located in higher layers, the lower the bit rate that can be obtained. On the other hand, more details will be neglected as more blocks are assigned into higher layers, and the quality of the reconstructed image may decline. Therefore, an appropriate threshold has to be selected to compromise an acceptable perceptual quality and a low coding rate. A structure map can be considered as a classification: the blocks in the same layer have similar gray-scale contrast.

The codebook is generated layer by layer. Each layer has its own subcodebook. There are many advantages for this subcodebook generation strategy. 1) Only layer members are involved in training for generation of the optimal subcodebook. Both the iteration times and the number of comparisons in each iteration are considerably reduced. As the result, the total training time can be shortened, though several training procedures are required for the whole codebook construction. 2) By applying layer assignment, we can adjust the size of each subcodebook according to the accuracy requirement for reconstruction of the different regions of an image by either choosing the threshold values or changing the sizes of the subcodebooks for the specific layers. 3) The blocks in layers 1 and 2 can be subsampled due to the strong correlations of the gray-scales among the adjacent pixels. We can use 16-D vectors to represent the blocks of size 16 × 16 in layer 1 and 8 × 8 in layer 2. In the reconstruction period, the missed pixels in these two layers can be reconstructed by linear interpolation at the receiving end. No significant deterioration of the perceptual quality will occur.

**III. Hierarchical Finite-State Coding**

With the HFSVQ, the size of the candidate codeword category for a coding block is variable based on the block's layer location. Except the edge region, the gray-scales do not
demonstrate the occurrence probability distributions of the optimal representative vectors ordered by the increment of the distortions before its representative vector decision. That is, we take some encoded blocks as the references to predict the block X as shown in Fig. 3. We first assume that the block above X° and the block on the left side X' have the same size of the block X. The prediction performs by measuring the distances between the encoded (or vector quantized) pixel intensities of the blocks X, X°, and X' respectively.

If a block is located on neither top nor left side of the original image, we can rearrange the codewords of the related subcodebook in an increasing order of their sidematch distortions before its representative vector decision. That is,

\[
W_l \in C_i \quad 1 \leq l \leq N_i
\]

where \(W_l\) is the \(l\)th element in the subcodebook \(C_i\) after rearrangement and \(N_i\) the size of the subcodebook \(C_i\). Fig. 4 demonstrate the occurrence probability distributions of the optimal representative vectors ordered by the increment of the \(e_r\). It can be seen that the occurrence distributions demonstrate a rapidly decreasing trend, i.e., the optimal representative vectors most probably occur among the codewords which have small side-match distortions, especially in the higher layers.

For any block located on neither top nor left side of the image frame, its representative vector can be chosen merely from a few of codewords having the smallest side-match distortions with the encoded blocks \(X^\circ\) and \(X'\). Therefore, for blocks located in layers 1 to 3, about 1/8 of the codewords in the subcodebooks, which have the smallest \(e_r\), are selected to form candidate categories for representative vectors. There are only a few of the blocks which are encoded without the best representative vectors. Though the SNR of the reconstructed image by HFSVQ may be a little bit lower, the perceptual quality almost is not affected by this slight deterioration due to the insensitivity of human eye on the grey-scale variance in smooth areas.

The bit rate for the address codes \(R_{AD}\) is given by

\[
R_{AD} = \frac{\sum_{i=1}^{n} N_i \log_2(n_i)}{N}
\]

where \(n_i, i = 1, 2, 3, 4\), is the size of the candidate categories of the subcodebook for layer \(i\). If Huffman codes, which are represented by variable word-length codes based on the statistics of the codeword occurrence probability, are employed for layers 1 to 3 coding, the bit rate can be further reduced by 20% since the candidate category size for layer 1, 2, and 3 is rather small (8–32) and little computational complexity is added. It is more rigorous to encode the blocks of the edge layers \(L_4\), as the corresponding occurrence trend of the representative vector is not as steep as the trends in other layers, as shown in Fig. 4(d), and slight edge distortions may thoroughly damage the perceptual result of the reconstructed image. Considering the possibility that there are sharp edges occurring on the block boundaries, we cannot neglect the possibly optimal vectors even though they may have large \(e_r\). Hence the sizes of the candidate categories in layer 4 will be larger than those in the other layers.

If the block \(X^\circ\) or \(X'\) are not located in the layer \(X\) located, their sizes may be different from that of \(X\). In this case \(X^\circ\) or \(X'\) have to be adjusted to adapt the size of \(X\).

A color image compression can be conducted by encoding its luminance image \(Y\) and two chrominance images \(I\) and \(Q\) separately. Although the three images have their own characteristics, their basic structures are the same. Therefore, all the structure information can be recorded in one structure map.

The luminance image \(Y\) contains most energy of the three components. It can be considered as the corresponding grey
image and can be processed with the same technique used in monochrome image coding. The construction of the structure map for the luminance image is the same as the map for monochrome images.

As for the two chrominance images, a further compression can be achieved by making use of the characteristics of their limited energy. Compared to the luminance image, chrominance images have smaller dynamic ranges of the gray-scales. Fewer representative codewords are needed for image reconstruction with acceptable perceptual quality. Only half of the sizes of the codebooks for the image $Y$ are sufficient for the codebooks of the images $I$ and $Q$. The bits of the address codes then are much fewer than those for the luminance image. At the same time, the edges in the chrominance images are usually more vague and the human eye is less sensitive in observing their minor distortions. The high frequency components can be removed and a relatively coarse reconstruction of chrominance images is acceptable. A low-pass filtering is performed by simplifying the structure map and sub-sampling. The structure map for the luminance image is simplified by enlarging the blocks in layers 3 and 4 to the blocks of size $8 \times 8$ and then forming layers $3'$ and $4'$. Layer $4'$ collects the blocks in which at least one of the four subblocks of size $4 \times 4$ is located in layer 4 in the original map. The blocks, whose four subblocks originally are all in layer 3, are assigned into layer $3'$. No additional information is needed for building the structure map of the chrominance images. By enlarging the block sizes, the bits of the address codes for the blocks in layers 3 and 4, which are the main parts of the codes, can be reduced to $1/4$ of the bits for the luminance image.

If the HFSVQ technique is employed to encode chrominance images, the members of the candidate categories for prediction can also be fewer than those required for luminance image.
image coding due to the smaller sizes of their codebooks. Usually 1 or 2 b can be reduced in each address code. The bit rate of the address codes for a chrominance image \( R_{AD}^{ch} \) can be obtained by modifying \( R_{AD} \) in (8) to the follows:

\[
R_{AD}^{ch} = \frac{(N_1 + N_2)(\log_2 n_1 - \delta) + \frac{N_3 + N_4}{4}(\log_2 n_3 - \delta)}{N} \text{ bpp}
\]  

(9)

where \( \delta \) may be 1 or 2 according to the requirement of the reconstruction accuracy. With equal perceptual quality of the reconstructed images, \( R_{AD}^{ch} \) is less than 1/3 of \( R_{AD} \) of the luminance images.

V. EXPERIMENTAL RESULTS

Several monochrome and color images are encoded by using the HFSVQ technique. Since there is no measurement equivalent to the satisfactory perception of human eye, we can only calculate the SNR of each image for rough evaluation of its reconstruction quality. The SNR is defined by

\[
\text{SNR} = -10 \log_{10} \frac{1}{N} \sum_{ij} \frac{(x_{ij} - \hat{x}_{ij})^2}{255^2} \text{ dB}
\]  

(10)

where \( x_{ij} \) and \( \hat{x}_{ij} \) are original and encoded pixel intensities, respectively.

"Lena" is an 8 bpp, 512 x 512 monochrome image as shown in Fig. 5. By using the HFSVQ, the image is decomposed into the blocks with different sizes and the blocks are assigned into the four layers as shown in Fig. 6 where different grey-scale shades indicate different layers. The lighter a block is, the higher layer it belongs and the smoother region it is located. The super codebook consists of 3 subcodebooks and there are 256 codewords in the codebook: subcodebooks 1 and 2 have 64 codewords, used for layers 1 and 2 block coding, respectively; both subcodebooks 1 and 2 are also for layer 3 block coding; subcodebook 3 has 128 codewords which are used for layer 4 block coding. Before we encode each image block, we arrange the codewords in the related subcodebook in the increasing order of side-match distortion \( \epsilon_{ss} \), only first few codewords of the subcodebooks, which have the smallest \( \epsilon_{ss} \), are selected as the representative vector candidates for the current block coding. Based on the correlation of the adjacent pixels in the different regions, the numbers of the candidates are variable. Eight codewords are used in layers 1 and 2, 32
in layer 3. Hence, we only use 3 b to code a 16 × 16 block in layer 1 or an 8 × 8 block in layer 2; 5 b to code a 4 × 4 block in layer 3. No finite-state prediction is applied in layer 4 block coding in case that there are edges located on the block boundaries, so that each block in this layer is encoded by 7 b. Because of the steep descending trend of the probability of the optimal representative coding, an entropy coding, such as a Huffman coding scheme is applied to achieve further data compression. The reconstructed image, encoded by a codebook generated by inside sequence training, is shown in Fig.7 with the \( \text{SNR} = 30.645 \text{ dB} \). The total bit rate is 0.212 bpp without the Huffman coding and 0.174 bpp with it, in which \( R_{ST} = 0.029 \text{ bpp} \), \( R_{AD}(Y) = 0.125 \text{ bpp} \), \( R_{AD}(I) = 0.035 \text{ bpp} \) and \( R_{AD}(Q) = 0.035 \text{ bpp} \). The total bit rate is 0.224 bit per color pixel, less than 1% of the bit rate of the original image (24 bits per color pixel). The SNR of the reconstructed image is 30.02 dB where \( \text{SNR}(Y) = 29.97 \text{ dB} \), \( \text{SNR}(I) = 29.87 \text{ dB} \) and \( \text{SNR}(Q) = 30.22 \text{ dB} \).

VI. CONCLUSIONS

We have presented an improved VQ, HFSVQ in this paper, which has advantages of both hierarchical VQ and finite-state VQ techniques. By using hierarchical structure, the bit rate for the information in smooth areas is considerably reduced by large block coding. The layer assignment also enable us to choose different candidate category size for different layers: fewer candidates for the higher layer block coding and more for lower layers one. Hence, we can obtain a further compression of the smooth area information as well as providing more accurate edge reconstruction by guaranteeing the optimal representative vector selection.

This technique also encourages us to explore deeper relations among the sensitivity of human eye, the transitions of grey-scales in images and the various features of the different parts of an image, so that further effective compression can be achieved.

REFERENCES

Ping Yu received the B.E. and M.A.Sc. from Beijing University of Posts and Telecommunications, China, in 1982 and 1984. She is currently working toward the Ph.D. degree at the University of Toronto. From 1984 to 1989, she was a lecturer in Beijing University of Posts and Telecommunications. Her research interest includes image processing, pattern recognition and neural networks.

Anastasios N. Venetsanopoulos (S'66-M'69-SM'79-F'88) received the B.S. degree from the National Technical University of Athens (NTU), Greece, in 1965, and the M.S., M.Phil., and Ph.D. degrees in electrical engineering, all from Yale University, in 1966, 1968, and 1969, respectively. He joined the University of Toronto, Canada, in September 1968, where he is now Professor in the Department of Electrical Engineering. He also served as Chairman of the Communications Group (1974–1978 and 1981–1986), and as Associate Chairman of the Department of Electrical Engineering (1978–1979). He was on research leave at the Swiss Federal Institute of Technology, the University of Florence, the Federal University of Rio de Janeiro, the National Technical University of Athens, and the Imperial College of Science and Technology, and was Adjunct Professor at Concordia University. He served as Lecturer of numerous short courses to industry and continuing education programs; he is a contributor to eleven books and has published over 300 papers in digital signal and image processing, and digital communications; he also served as consultant to several organizations, and as Editor of the Canadian Electrical Engineering Journal (1981–1983).

Dr. Venetsanopoulos has served the IEEE in the following capacities. 