ISSN 2616-5775 Vol. 6, Issue 10: 1-6, DOI: 10.25236/AJCIS.2023.061001

Lossy Compression Approaches Based on Vector Quantization

Shengzhong Zhang[a,*], Lei Yu[b], Yinqian Cheng[c]

Information Network and Data Center, China University of Geosciences (Beijing), Beijing, China
[a]zhsz@cugb.edu.cn, [b]yul@cugb.edu.cn, [c]chengyq@cugb.edu.cn
[*]Corresponding author

Abstract: *Vector Quantization (VQ) is an effective lossy compression technology developed in the late 1970s. Its theoretical basis is Shannon's rate distortion theory. The basic principle of vector quantization is to use the index of the codeword in the codebook that best matches the input vector for transmission and storage, while decoding only requires a simple table lookup operation. Its outstanding advantages are high compression ratio, simple decoding, and the ability to preserve signal details well. In this article, several VQ approaches are introduced for lossy compression.*

Keywords: *Lossy Compression, Vector Quantization (VQ), Codebook, Self-Organizing Feature Mapping (SOFM)*

1. Introduction

Lossy compression is also called entropy compression. Because the information entropy is compressed, the compression efficiency is improved. According to the lossy principle, there are classification, predictive coding, transform coding, quantitative coding, information entropy coding, frequency division coding, structural coding and knowledge-based coding methods for lossy compression [1].

The essence of quantization and vector quantization is to compress statistical redundancy, and VQ is to maximize the retention of image quality within the range of communication capacity and storage requirements. VQ is an effective method for low bit rate image compression. When the vector dimension increases, the coding performance of VQ reaches the approximate rate-distortion limit, but its coding complexity increases exponentially with the dimension. In the VQ algorithm, various relevant information in the image (such as between pixel points, blocks and adjacent coded addresses) can be fully removed by effective codebook design. The primary and core issue in vector quantization design is the design of codebooks. If there is no codebook design, the entire vector quantization system cannot be implemented. The quality of the codebook directly affects the compression efficiency and the quality of the restored signal [2].

After the process of transforming image data into frequency domains, frequency division band coding is dividing the frequency bands according to their frequency, and quantifying them with different quantizers to achieve the optimal combination. Alternatively, progressive coding can be used to start decoding a signal in a certain frequency band and then extend it to all frequency bands. As the decoded data increases, the decoded image becomes clearer. This method is more effective for image fuzzy query and retrieval applications.

The vector quantization method effectively utilizes the correlation properties between the various components of the vector (linear, nonlinear dependencies, probability function shape, and vector dimension) for de-correlation processing, resulting in high compression ratio. VQ is a set quantization of vectors, which only needs to transmit or store vector addresses. The VQ encoding method compares the input vector with the codebook vector, identifies the codebook vector with the smallest error, records its index or sequence number, and then finds the codebook vector in the same vector codebook at the decoding end to replace the input vector, thus restoring the original input vector [3] [4] [5].

Adaptive vector quantization in the wavelet transform domain is to perform vector quantization on each wavelet transform sub image, and when training each sub image with codebook, adaptively select distortion criteria, codebook size, and codebook vector size based on human visual characteristics.

ISSN 2616-5775 Vol. 6, Issue 10: 1-6, DOI: 10.25236/AJCIS.2023.061001

Compared to using a unified codebook for each sub-graph, it increases the complexity of codebook training and requires training multiple codebooks. However, it can effectively reduce the size of the codebook and the encoding programs. Data quantization errors at different scales will have different impacts on the quality of reconstructed images, and the human eyes are more sensitive to large-scale image data information. For sub images with larger scales and larger variances, a larger codebook size should be selected and error compensation should be performed. This can minimize the codebook size, shorten the codebook training time, and improve efficiency while ensuring the subjective and objective quality of the reconstructed image. Currently, effective vector quantization methods based on wavelet decomposition include Pyramidal Lattice Vector Quantization (PLVQ) proposed by M. Baland et al. and Embedded Zerotree Wavelets (EZW) method proposed by J. M. Shapiro. These two methods have high efficiency, but are computationally complex and not suitable for situations with high real-time requirements [3] [4].

Vector quantization maximizes the retention of image quality within the range of communication capacity and storage requirements. VQ is an effective method for low bit rate image compression. When the vector dimension is added, VQ has the coding performance of approximately reaching the rate distortion limit, but its complexity increases with the dimension or Exponential growth.

Image VQ consists of four steps:

(1) Decompose the image into a set of vectors.

(2) An input vector subset is selected as a training set.

(3) Generate codebooks from training sets, typically using iterative clustering methods.

(4) For each input vector, the codeword of the closest code vector in the codebook is found and transmitted.

So the four steps are: vector formation, training set generation, codebook generation, and quantization.

2. LBG Algrithm

The LBG algorithm is developed by Y. Linde, A. Buz, and R. M. Gray based on the Floyd-max algorithm in 1980, and this is the first vector quantizer algorithm published. The idea is to first identify the center of a training sequence and then generate an initial codebook $\hat{A}_0$ using the splitting method. Then group the training sequence according to the elements in codebook $\hat{A}_0$, find a new codebook at the center of each group, and use the new codebook as the initial codebook for the above process. The iterative algorithm for LBG is [6]:

(1) Given the initial book $A_0=\{y_i\}$, i=1, 2, •••, N, the distortion domain value is ε.

(2) Find a K-dimensional vector space partition $\{S_i\}$, and according to this partition, minimize the distortion sum between the training sample vector and its third-class representative vector (codeword) y_i, that is,

$$D = \sum_{x \in s} d(x, y_i) \tag{1}$$

Then correspond the training sample x to the codeword closest to x in codebook A_0 using the nearest neighbor method. That is to say, if d (x, y_i) < d (x, y_1), then the training sample vector x is assigned to the i^{th} group, denoted as $x \in S_i$.

(3) Without changing the spatial partition, only correcting the centers of each group to obtain a new codebook $A_1 = \{y_i^{prime}\}$, which minimizes the total distortion D for the current vector space partition.

Return to step (2) and use the new codebook C_1 as the criterion to partition the vector space again. Repeat the iteration until $(D_{m-1}-D_m) / D_m <= \varepsilon$, then the resulting codebook is the desired one.

ISSN 2616-5775 Vol. 6, Issue 10: 1-6, DOI: 10.25236/AJCIS.2023.061001

There are many methods for designing the initial codebook A_0, and different initial codebooks often have a significant impact on the final codebook.

3. Self Organizing Feature Mapping (SOFM) Algorithm

The code book generated by LBG algorithm gets an unordered arrangement of code vectors. Using this serial number as the encoding output of the vector quantizer is particularly sensitive to channel errors. In order to control the serious performance degradation of the whole vector quantization communication system caused by channel error, the LBG method can be modified by neural network technology.

The SOFM algorithm using Kohonen network belongs to unsupervised learning. The algorithm is:

(1) Provide the number of output nodes N and the number of input nodes (each vector element) K, and initialize the weight values from input node i to output node j. Set the ownership value as a random decimal, and X (t) as the training sequence.

(2) Input mode:

$$X(t) = (x_0(t), x_1(t), \cdots, x_{K-1}(t))^T \tag{2}$$

(3) Calculate the distance between the weight vectors w_j of all output nodes on the input vector X(t):

$$d_j = \sum_{i=0}^{K-1} (x_i(t) - w_{ij}(t))^2 \quad \text{j=0, 1, •••, N-1,} \quad w_j = (w_{0j}, w_{1j}, w_{K-1,j})^T \tag{3}$$

(4) Find the node N_{j^*} with the least moment separation:

$$d_{j^*} = \min_{0 \le j \le N-1} \{d_j\} \tag{4}$$

(5) Adjust the weights connected to N_{j^*} and the weights connected to nodes in the geometric field.

Adjust the weights connected to N_{j^*} and the weights connected to nodes in the $NE_{j^*}(t)$ geometric field.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)(x_i(t) - w_{ij}(t)), \quad N_j \in NE_{j^*}(t), \tag{5}$$

It is a variable learning speed that, like N_{j^*}, decreases over time.

(6) If there are still input samples, then t=t+1, go to (2).

Due to the decrease of d(t) over time, the magnitude of weight adjustment decreases as the training process progresses, so that the weight vector connected to the winning node can represent the essential attributes of the pattern.

4. Fuzzy Vector Quantization (FVQ) Coding [7]

The classic LBG codebook design algorithm is based on the minimum average distortion of training vectors and codebook vectors. Although simple and easy to implement, it strongly relies on the initial codebook selection and is prone to falling into local minima. Some scholars use random relaxation techniques to minimize average distortion in order to search for the global best codebook. The common feature of these technologies is that during each iteration of searching for the minimum average distortion, independent variables included in a certain degree of distortion or a certain degree of distortion are randomly interfered with. Random relaxation technology can generate a global minimum codebook, but the computational complexity is large.

ISSN 2616-5775 Vol. 6, Issue 10: 1-6, DOI: 10.25236/AJCIS.2023.061001

The above code book design calculation method is based on hard decision, where each training vector is assigned to a single cluster according to some criteria, ignoring the possibility that the training vector belongs to other clusters. The Fuzzy K-Means (FKM) clustering algorithm assigns a membership value between the numerator 0 and 1 to each element of the training set, indicating the degree to which the training vector belongs to a certain cluster. FKM performs better than LBG, but it requires a lot of computation. Nicolas B. K et al. proposed the FVQ algorithm. FVQ regards each cluster as a Fuzzy set, and uses membership function to indicate that the training vector belongs to a certain cluster degree. In this way, each training vector is assigned to multiple clusters based on the measure of the membership function. If the training vector being considered is the center of a hypersphere, ensuring the participation of all training vectors in the centers of overlapping hyperspheres effectively reduces the dependence of the design results on the initial codebook. FVQ generates the globally optimal final codebook vector based on ensuring no local minimum constraints.

FVQ design follows three essential factors. First set the training vector set χ be n-dimensional, with a value of M, $\chi = \{x_1, \bullet\bullet\bullet\bullet, x_M\}, x_i \in R^n \forall i = 1,\cdots, M$. The codebook vector set is y, with the value of k, $y = \{y_1, \bullet\bullet\bullet, y_k\}$, $y_j \in R^n, \forall j = 1,2,\cdots, k$.

(1) The principle of transforming the training vector from soft decision to hard decision is that the transformation speed can be adjusted using the shrinkage scheme of the hypersphere in clustering.

(2) The selection of membership function $u_j(x_i)$ depends on the distance x_i and $y_i \in I_i^{(v)}$, Satisfying:

(a) $u_j(x_i)$ is a decreasing function of distance $d(x_i, y_j)$;

(b) When $d(x_i, y_j) \to 0$, $u_j(x_i) \to 1$;

(c) When $d(x_i, y_j) \to 1$, $u_j(x_i) \to 0$.

(3) Select the threshold values ε.

Jihong Zhang et al. [7] proposed exponential type membership functions and exponential type fuzzy vector quantization (FVQE) algorithm based on the overlapping conditions of membership functions in the second factor. The coding performance is equivalent to FVQ, and the rate of convergence is slightly faster.

5. A Fast Vector Quantization Encoding Method [8]

In order to accelerate the VQ process, people search for fast encoding algorithms, which can be divided into two categories:

(1) Not solving the nearest neighbor problem itself, but finding a suboptimal solution that is almost as good as Mean Square Error (MSE). Generally, they rely on the use of data structures that facilitate rapid search.

(2) A simple and effective method to solve the nearest neighbor coding problem is the Partial Distortion Search (PDS) method, which does not require memory overhead but only has a moderate acceleration ratio. Increasing memory overhead, fast nearest neighbor search (FNNS) algorithm and projection method can save a lot of time. FNNS can skip many impossible codewords with Triangle inequality. But it requires N*(N-1)/2 to store the distance of all codewords' time. Projection methods, such as the Equal Mean Nearest Neighbor Search (ENNS) algorithm, use the mean of the input vector to cancel impossible codewords. Compared with traditional global search algorithms, it saves a lot of computational time and only increases N memory. The improved algorithm is called the Equal Mean Equal Variance Nearest Neighbor Search (EENNS) algorithm, which uses the variance of the input vector to save more time and increase memory by 2N.

Seong Joon Beek et al. [8] proposed a new inequality between the mean, variance, and distance of input vectors. Applying this inequality to the VQ encoding algorithm can reduce the search range of codewords. Experiments have shown that its performance is superior to ENNS and EENNS. And compared to EENNS, it does not require additional memory.

ISSN 2616-5775 Vol. 6, Issue 10: 1-6, DOI: 10.25236/AJCIS.2023.061001

6. Model Based Vector Quantization (MVQ) Encoding [6]

There are two key factors in the generation of codebooks. One is to choose an appropriate codebook generation algorithm, and the other is to select data as the training set. Codebooks can be divided into local and global codebooks. Due to the inclusion of compressed image statistics in the local codebook, it provides good rate distortion performance, but there are also drawbacks.

(1) Generating a codebook for each image requires too much computation;

(2) For a given distortion, the codebook has to be transmitted as side information to the decoder, reducing compression efficiency.

For global codebooks, the compression quality of images that are not similar to the images in the training set will decrease, and selecting an appropriate global codebook requires knowledge of the compressed image. To compress images well enough, several sets of training sets must be used to represent the classes of image data. And the codebook for each image class must be sent separately to the decoder.

Manohar et al. designed a VQ method that does not involve the computation of local codebooks or the storage of global codebooks. It can also have promising asymmetric computational properties. MVQ can achieve this goal.

MVQ does not require codebook training and storage, transmitting codebooks. In MVQ, codebook generation is based on the error model and Human Visual System (HVS) model, which eliminates the need for selecting training sets and avoids the significant calculation of the training process caused by codebook generation.

In MVQ, the image is divided into K pixel non overlapping square blocks, and the block mean value (m_0... m_{N-1}) is calculated and stored or transmitted, which can be completed after lossless compression. Then, the block mean is subtracted from the image vector, and the image residual vector is vectorized using a model codebook. This model codebook is constructed using residuals as Laplacian distribution modeling, generating a random codebook based on this model, and then adding relevant structures to the codebook elements based on HVS. In order to reconcile the codebook with the original image characteristics, it is necessary to add relevant structures to the vector elements so that they have the same covariance as the source residual image. After generating the model codebook, VQ is used to compress the residual vectors.

At the decoding end, the codebook is regenerated using the Laplacian parameter λ.

MVQ has stronger performance on distortion than VQ, and less decoding time than JPEG, so it gets better performance.

7. Summary

Vector quantizers are the best quantizers to achieve minimum distortion for a given bitrate and vector dimension. The goal of a vector quantizes based system is to reduce the bitrate and minimize communication channel capacity or digital storage memory requirements while maintaining the necessary fidelity of the data. Vector quantization provides many attractive features for image coding applications with high compression ratios [9].

Vector quantization gets excellent performance in lossy compression and will reduce transmission and storage cost significantly. Vector quantization has a wide range of applications in the field of image compression, such as compression and real-time transmission of satellite remote sensing photos, video compression of digital TV and DVD, compression and storage of medical images, and image recognition. Therefore, vector quantization has become one of the important technologies in image compression coding.

Acknowledgements

This research is supported by China Earthquake Science Experiment project, China Earthquake Administration (2019CSES0113).

ISSN 2616-5775 Vol. 6, Issue 10: 1-6, DOI: 10.25236/AJCIS.2023.061001

References

[1] Zhengxing Cheng, Wavelet Analysis Algorithms and Applications, Xi'an Jiaotong University Press, 1998. 8
[2] Wenji Xu, Research on Fast Codeword Search Algorithms for Vector Quantization, Master's Thesis at Suzhou University, 2008. 5
[3] Qiang Li, Zhengzhi Wang. Research on High Fidelity Compression Methods for Remote Sensing Images Based on Wavelet Theory, Chinese Journal of Image and Graphics, 1999, 3 (1): 31-36
[4] Wen Jiang, Zhongxin Le. Video Coding Algorithm Based on Wavelet Transform, Chinese Journal of Image and Graphics, 1997, 2 (10), 721-725
[5] An Li et al. A VQ image encoding method based on wavelet transform combined with reverse selection correction, Chinese Journal of Image and Graphics, 1997, 2 (7): 488-490
[6] Mareboyanna Manohar et al. Model-Based Vector Quantization with Application to Remotely Sensed Image Data, IEEE Trans. on Image Processing, 1999, 8(1): 15-21
[7] Zhang Jihong, Wang Hui, et al. Image Coding Research Based on Fuzzy Vector Quantization, Chinese Journal of Image and Graphics, 1998, 3 (4): 295-298
[8] Seong Joon Back et al. A Fast Encoding Algorithm for Vector Quantization, IEEE Signal Processing Letters, 1997, 4(20): 325-327
[9] A. J. Hussain et al. , Image compression techniques: A survey in lossless and lossy algorithms, Neurocomputing, 2018, 300: 44–69