Image Compression Schemes A Literature Survey

Komal Kant[1], Prof. Amarjeet Ghosh[2]

[1]M.Tech Scholar, [2]Research Guide

Department of Electronics and Communication Engineering, VITS, Bhopal

Abstract-The expanding interest for media substance, for example, digital images and video has prompted extraordinary enthusiasm for examination into compression methods. The advancement of higher quality and more affordable image procurement gadgets has delivered consistent increments in both image size and resolution, and a more noteworthy resulting for the plan of proficient compression frameworks. In spite of the fact that capacity limit and exchange bandwidth has developed appropriately as of late, numerous applications still require compression. The taking off number of images and recordings requires productive compression. Deep Compression comprises of pruning, prepared quantization, and variable-length coding, and it can pack deep neural systems by a request of greatness without losing the expectation exactness. This huge compression empowers machine learning applications to keep running on cell phones..

Keywords- Deep Image Compression, Non Uniform Quantization,Entropy Coding,

I. INTRODUCTION

In general, this study investigates still image compression in the transform domain. Multidimensional, multispectral and volumetric digital images are the main topics for analysis. The main objective is to design a compression system suitable for processing, storage and transmission, as well as providing acceptable computational complexity suitable for practical implementation. The basic rule of compression is to reduce the numbers of bits needed to represent an image. In a computer an image is represented as an array of numbers, integers to be more specific, that is called a digital image. The image array is usually two dimensional (2D), If it is black and white (BW) and three dimensional (3D) if it is color image. Digital image compression algorithms exploit the redundancy in an image so that it can be represented using a smaller number of bits while still maintaining acceptable visual quality. Factors related to the need for image compression include:

In the array each number represents an intensity value at a particular location in the image and is called as a picture element or pixel. Pixel values are usually positive integers and can range between 0 to 255. This means that each pixel of a BW image occupies 1byte in a computer memory. In other words, we say that the image has a grayscale resolution of 8 bits per pixel (bpp) . On the other hand, a color image has a triplet of values for each pixel one each for the red, green and blue primary colors. Hence, it will need 3 bytes of storage space for each pixel. The captured images are rectangular in shape.

- The large storage requirements for multimedia data
- Low power devices such as handheld phones have small storage capacity
- Network bandwidths currently available for transmission
- The effect of computational complexity on practical implementation.

The main advantage of compression is that it reduces the data storage requirements.. Offered by the compression techniques, computer network and Internet usage is becoming more and more image and graphic friendly, rather than being just data and text-centric phenomena. In short, high-performance compression has created new opportunities of creative applications such as digital library, digital archiving, video teleconferencing, telemedicine and digital entertainment to name a few. There are many other secondary advantages in data compression.

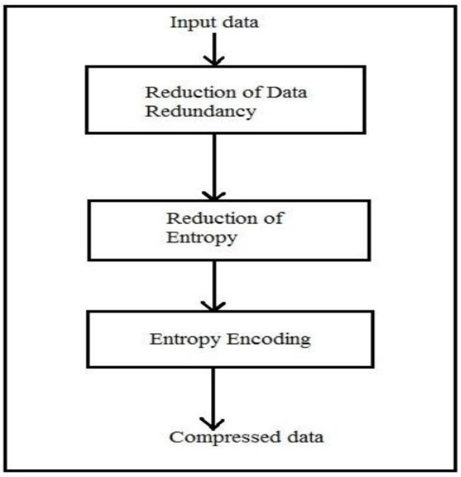

Figure: 1.1 A Data Compression Model.

For Example it has great implications in database access. Data compression may enhance the database performance because more compressed records can be packed in a given buffer space in a traditional computer implementation. This potentially increases the probability that a record being searched will be found in the main memory. Data security can also be greatly enhanced by encrypting the decoding parameters and transmitting them separately from the compressed database file to restrict access of proprietary information. An extra level of security can be achieved by making the compression and decompression processes totally transparent to unauthorized users. The rate of input-output operations in a computing device can be greatly increased due to shorter representation of data. Data compression obviously reduces the cost of backup and recovery of data in computer systems by storing the backup of large database files in compressed form. The advantages of data compression will enable more multimedia applications with reduced cost.

Various coding techniques are used to facilitate compression. Here are pixel coding, predictive coding and transform coding will be discussed:

- *Pixel Coding*

In this type of coding, each pixel in the image is coded separately. The pixel values that occurs most frequently are assigned shorter code words (i.e. fewer bits), and those pixel values that are more fare (i.e. less probable) are assigned longer codes. That makes the average code word length decrease.

- *Predictive Coding*

As images are highly correlated from sample to sample, predictive coding technique is relatively simple to implement. Predictive coding predicts the present values of the sample based on the past values and only encodes and transmits the difference between the predicted and the sample value.

- *Transform Coding*

In transform coding, an image is transformed from one domain (usually spatial or temporal) to a different type of representation, using some well –known transform. Then the transformed values are coded and thus provide greater data compression.

II. NON UNIFORM QUANTIZATION AND DEEP COMPRESSION

Deep Compression" is a three-stage pipeline to reduce the model size of deep neural networks in a manner that preserves the original accuracy. First, we prune the network by removing the redundant connections, keeping only the most informative connections. Next, the weights are quantized and multiple connections share the same weight. Thus, only the codebook (effective weights) and the indices need to be stored; each parameter can be represented with much fewer bits. Finally, we apply variable-length coding (Huffman coding) to take advantage of the non-uniform distribution of effective weights and use variable length encoding to represent the weights in a lossless manner.

Our most important insight is that pruning and trained quantization can compress the network without interfering with each other, thus leading to a surprisingly high compression rate. Deep Compression makes the required storage so small (a few megabytes) that all weights can be cached on- chip instead of going to off-chip DRAM, which is slow and energy consuming.

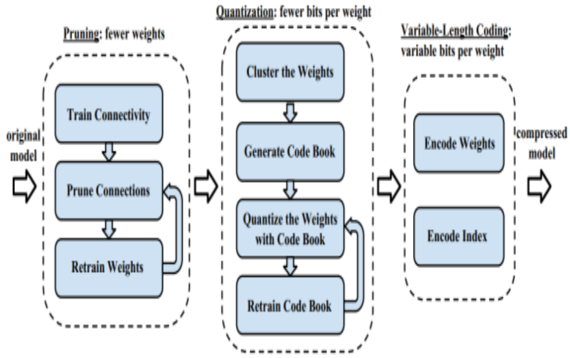

Figure 2.1 Deep Compression Pipeline.

An NU quantizer can be used to reduce overall quantization noise power. To take advantage of NU-quantization in Sigma-Delta converters, the quantization steps should be arranged in such a way that small-amplitude input signal is quantized finely by smaller quantization steps and larger-amplitude signals are quantized coarsely.

$$Q_n = X_{n-2} + (-2e_{n-1} + e_{n-2})$$

The characteristics of quantizer input distribution in a sigma-modulator can also be explained on the basis of the delta operation. Because of the delta operation, the quantizer input or the unquantized modulator output oscillates across the analog-ground, in the form of a ramp, by a factor which is proportional to the input amplitude. Because of this reason, the modulator output remains near the analog-ground more often. So, there is a higher density of smaller values.

These quantizers require a very peculiar set of thresholds as the largest quantization step could be ten times larger than the smallest step. Also, because the different quantization steps are not linearly changing, therefore it is not possible to use unit elements to match these steps.

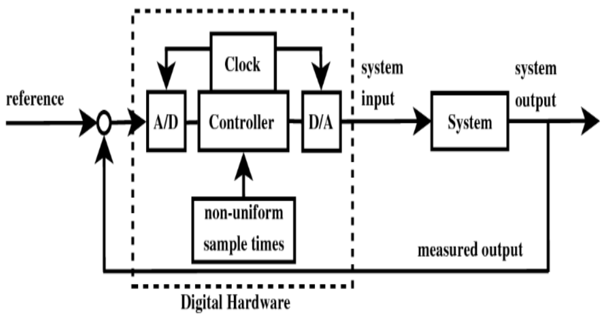

Figure 2.2 Non uniform Quantization Block Diagram.

a) Quantizer

Quantizer is a key component in the transform compression framework that introduces non-linearity in the system. It maps the transformed digital image to a discrete set of levels or discrete number. it is a lossy compression technique as it introduces an error in the image compression process. It converts sequence of floating point numbers to sequence of integers. In other words, If the data symbols are real numbers, quantization may round each to the nearest integer.

(b) Scalar Quantization

Scalar Quantization is the simplest quantization because each input is treated separately in producing the output. In other words, ―If the data symbols are numbers, then each is quantized to another number in a process referred to as scalar quantization‖. Many image compression methods are lossy, but scalar quantization is not suitable for image compression because it creates annoying artifacts in the decompressed image.

(c) Vector Quantization

In vector quantization, the input samples are clubbed together in groups called vectors, and then processed to give the output. If each data symbol is a vector, then vector quantization converts a data symbol to another vector. The idea of representing groups of samples rather than individual samples is the main concept of vector quantization. Vector quantization is based on the fact that adjacent data symbols in image and audio files are correlated.

(d) Predictive Quantization

The quantization of the difference between the predicted value and the past samples is called predictive quantization. A good quantizer is one, which represents the original signal with minimum loss or distortion.

III. LITERATURE SURVEY

Sr. No.	Title	Author	Year	Approach
1	Deep Image Compression with Iterative Non-Uniform Quantization	J. Cai and L. Zhang	2018	In this paper, we present an iterative nonuniform quantization scheme for deep image compression. More specifically, we alternatively optimize the quantizer and encoder-decoder.
2	Deep network-based image coding for simultaneous compression and retrieval	Q. Zhang, D. Liu and H. Li	2017	Images on the Internet are usually in the form of compressed bitstream to save storage. To fulfill content-based image retrieval (CBIR), image features are also required to be stored in binary form.
3	Non-uniform quantization scheme for the decoding of low-density parity-check codes with the sum-product algorithm	X. Qu and L. Yin	2016	In this paper, a non-uniform quantization scheme is proposed for the decoding of Low-Density Parity-Check (LDPC) codes with the Sum-Product Algorithm (SPA).
4	Deep convolutional network based image quality enhancement for low bit rate image compression	C. Jia, X. Zhang, J. Zhang, S. Wang and S. Ma	2016	In this contribution, a novel image quality enhancement algorithm based on convolutional network is proposed for low bit rate image compression. Specifically, a downsample procedure is performed to generate lower resolution image for low bit rate compression.
5	Exploiting deep neural networks for digital image compression	F. Hussain and J. Jeong	2015	In this paper we address the problem of digital image compression using DNNs. We use two different DNN architectures for image compression i.e.
6	The modified reliability-based iterative majority-logic decoding algorithm with non-uniform quantization	Haiqiang Chen, Lingshan Luo, Youming Sun, Tuanfa Qin and Yunyi Liu,	2013	This paper investigates the performances of the modified reliability-based iterative majority-logic decoding (MRBI-MLGD) algorithm under different quantization schemes. An efficient non-uniform quantization method is presented for the MRBI-MLGD algorithm.

J. Cai and L. Zhang [1] Image compression, which aims to represent an image with less storage space, is a classical problem in image processing. Recently, by training an encoder-quantizer-decoder network, deep convolutional neural networks (CNNs) have achieved promising results in image compression. As a nondifferentiable part of the compression system, quantizer is hard to be updated during the network training. Most of existing deep image compression methods adopt a uniform rounding function as the quantizer, which however restricts the capability and flexibility of CNNs in compressing complex image structures. In this paper, we present an iterative nonuniform quantization scheme for deep image compression. More specifically, we alternatively optimize the quantizer and encoder-decoder. When the encoder-decoder is fixed, a non-uniform quantizer is optimized based on the distribution of representation features. The encoder-decoder network is then updated by fixing the quantizer. Extensive experiments demonstrate the superior PSNR index of the proposed method to existing deep compressors and JPEG2000.

Q. Zhang, D. Liu and H. Li [2] Images on the Internet are usually in the form of compressed bitstream to save storage. To fulfill content-based image retrieval (CBIR), image features are also required to be stored in binary form. Can the bitstream of images and image features be unified and further condensed? Is it possible that the same binary code serves for compression and retrieval simultaneously? To address this problem, we make preliminary studies on a deep network-based image coding scheme in this paper. We first train a deep network for compressing images into bitstream, and then train another deep network for extracting image features as binary vector. We then combine the above two networks, and finetune the combined network using triplets of images for the task of CBIR. Our experimental results show that the proposed scheme achieves a compression ratio of 5.3 for 32×32 thumbnails, outperforms JPEG at similar compression ratios, and the resulting code is directly available for CBIR. Our work indicates a promising direction of simultaneous image compression and retrieval.

X. Qu and L. Yin [3] In this paper, a non-uniform quantization scheme is proposed for the decoding of Low-Density Parity-Check (LDPC) codes with the Sum-Product Algorithm (SPA). With the proposed scheme, the quantization-range and quantization-interval for the soft Log-Likelihood-Ratio (LLR) values of the decoder input, the variable nodes, and the check nodes are set separately, and are also adjusted according to the evolution of the probability density distribution of the LLR values in the iterative decoding procedure to minimize the difference between the decoding values before quantization and that after quantization. Simulation results show that decoding with 3-bit non-uniform quantization of the proposed scheme may obtain a BER performance slightly better than that of decoding with 6-bits quantization of the traditional uniform quantization method.

C. Jia, X. Zhang, J. Zhang, S. Wang and S. Ma [4] In this contribution, a novel image quality enhancement algorithm based on convolutional network is proposed for low bit rate image compression. Specifically, a downsample procedure is performed to generate lower resolution image for low bit rate compression. While the decoder side, upsample is to be performed firstly to the original resolution. Image quality is further enhanced by the proposed convolutional deep network. In particular, an optional image quality improvement network can be utilized for further enhancement after the first network. With the help of deep network, more detailed and high-frequency information can be recovered while maintaining the consistency of contour area, leading to better visual quality. Another benefit of this approach lies in that the proposed approach is fully compatible with all third-party image codec pipeline. Experimental result shows that the proposed scheme significantly outperforms JPEG in low bit rate image compression.

F. Hussain and J. Jeong [5] Deep neural networks (DNNs) are increasingly being researched and employed as a solution to various image and video processing tasks. In this paper we address the problem of digital image compression using DNNs. We use two different DNN architectures for image compression i.e. one employing the logistic sigmoid neurons and the other engaging the hyperbolic tangent neurons. Experiments show that the network employing the hyperbolic tangent neurons out performs the one with the sigmoid neurons. Results indicate that the hyperbolic tangent neurons not only improve the PSNR of the reconstructed images by a significant 2~5dB on average but they also converge several order of magnitude faster than the logistic sigmoid neurons.

Haiqiang Chen, Lingshan Luo, Youming Sun, Tuanfa Qin and Yunyi Liu [6] This paper investigates the performances of the modified reliability-based iterative majority-logic decoding (MRBI-MLGD) algorithm under different quantization schemes. An efficient non-uniform quantization method is presented for the MRBI-MLGD algorithm. With the non-uniform strategy, the real received symbols will be quantized into integer values before passing to the decoder. The quantized reliability measures exhibit a non-uniform distribution with respect to their original received signals. Furthermore, the quantization resolution can be modified as required. Simulation results show that, when combined with factor correction techniques and a proper quantization parameter r, the MRBI-MLGD with nonuniform quantization performs well with very small quantized bits (3-4 bits) and has very fast decoding convergence speed.

IV. PROBLEM IDENTIFICATION

Digital images and videos are in an exponential increase due to the proliferation of Internet, digital cameras, and image and video applications. With the immersive display and imaging technologies, increasing penetration of high-speed broadband, and availability of computing power, image and video compression and copyright protection technology have come of age, enabling images and videos as a driving force for growth and innovation over the years. In many applications and end equipments, images and videos play a key role as a mechanism of information exchange, transmission and storage, and play an important part in human lives.

V. CONCLUSION

It additionally offers an appealing way to deal with lessen the communication cost in transmitting high volumes of information over whole deal joins by means of higher powerful usage of the accessible bandwidth in the information joins. This altogether helps in lessening the expense of communication because of the information rate decrease. Because of the information rate decrease, information compression additionally expands the nature of interactive media introduction through constrained bandwidth communication channels, In light of the diminished information rate. Transmission of extremely touchy packed information through a noisy communication channel, (for example, remote media) is hazardous in light of the fact that the burst blunders presented by the noisy channel can wreck the transmitted information. Another issue of information compression is the disturbance of data properties, since the compacted information is not quite the same as the first data.

REFERENCES

[1]. J. Cai and L. Zhang, "Deep Image Compression with Iterative Non-Uniform Quantization," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 451-455.

[2]. Q. Zhang, D. Liu and H. Li, "Deep network-based image coding for simultaneous compression and retrieval," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 405-409.

[3]. X. Qu and L. Yin, "Non-uniform quantization scheme for the decoding of low-density parity-check codes with the sum-product algorithm," 2016 6th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, 2016, pp. 121-125.

[4]. C. Jia, X. Zhang, J. Zhang, S. Wang and S. Ma, "Deep convolutional network based image quality enhancement for low bit rate image compression," 2016 Visual Communications and Image Processing (VCIP), Chengdu, 2016, pp. 1-4.

[5]. F. Hussain and J. Jeong, "Exploiting deep neural networks for digital image compression," 2015 2nd World Symposium on Web Applications and Networking (WSWAN), Sousse, 2015, pp. 1-6.

[6]. Haiqiang Chen, Lingshan Luo, Youming Sun, Tuanfa Qin and Yunyi Liu, "The modified reliability-based iterative majority-logic decoding algorithm with non-uniform quantization," 2013 8th International Conference on Communications and Networking in China (CHINACOM), Guilin, 2013, pp. 882-885.