CONTEXT BASED ARITHMETIC CODING FOR VIDEO COMPRESSION

BY

Ch.S.Pratheek

Arun Balajee V

Srikanth M

CONTEXT BASED ARITHMETIC CODING FOR VIDEO COMPRESSION

Chinthapalli Shiva Pratheek, M Srikanth, Arun Balajee V

INTRODUCTION

The image which is formed on the plane of our camera is basically a 2-Dimensional, time-dependent and continuous distribution of light energy. In order to obtain a digital snapshot of the above distribution, certain steps have to be followed which are listed below:

The continuous distribution of light energy must be sampled spatially. It generally depends on the geometry of the sensor elements of camera or any other photo acquisition device.

The resulting discrete function must then be sampled in time domain, to obtain a still image. This is done by a CCD or CMOS sensor present inside a digital camera

The resulting sets of values are quantized to a finite set of numeric values, so that they can be represented within the computer.

After performing these steps, we obtain a 2-D ordered matrix of integers. So, in other words, a digital image is a 2-D function of integer co-ordinates that maps to possible range of values which are called pixels. The digital photographs are made of large array of pixels. These have varying amounts of primary colours mixed to give the final desired colour. The pixel values are practically always made of binary words of length k, which is termed as bit-depth of an image. All the image formats used today, are simply ways of storing the image information efficiently in a computer file until the original image is reconstituted for displaying /printing purposes. Image compression is very essential in transmission and storage. The pixels in most of the images are correlated and hence, contain redundant information. The main objective of image compression is to obtain a less correlated representation of image i.e. redundancy and irrelevancy reduction. Redundancy reduction aims at removing duplication from signal source. Irrelevancy reduction aims at removing the parts of signal which are not recognized by receiver. The techniques used to compress images can be classified into two ways[1]:

Lossy v/s Lossless compression: In lossy compression, the quality of reconstructed image is less as compared to lossless compression as the former discards redundant information. It is because of this reason that lossy compression achieves higher compression ratio compared to lossless compression.

Predictive v/s Transform coding: In predictive coding, the sent/available information is utilized to predict future values, and the difference is coded in image or spatial domain. It is easy to implement and is readily adapted to local image characteristics. **DPCM** is an example of Predictive coding. Transform coding first transforms the image its spatial domain representation to different type of representation using some well-known techniques and is then coded. Higher compression ratio is achieved in this method at the expense of greater computing power.

When an image is stored on a computer, it is stored in a compressed format. Depending on the type of compression technique used viz. DCT, Wavelet etc. and algorithm used; different file-extensions are defined. These are explained below:

TIFF: It stands for Tagged Image File Format. It supports a wide range of images including special types of images which have large-depth integer and floating point elements. It is possible to store a number of variations of image in a single TIFF file since it supports various compression techniques like LZW, JPEG etc. the strength of this image format lies in its architecture which enables users to create new

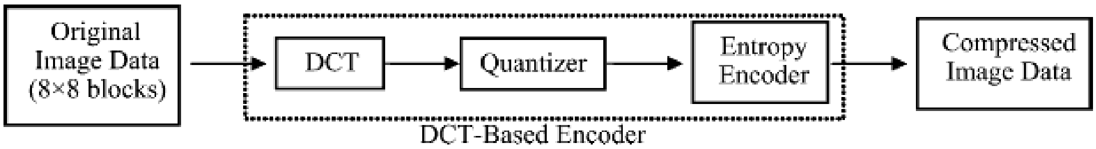


Fig. 1. Block diagram of JPEG compression process.

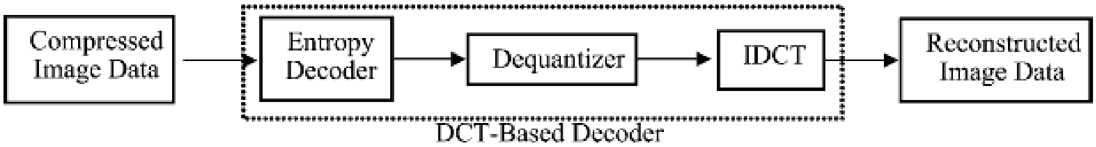


Fig. 2. Block diagram of JPEG decompression process.

BMP: The Windows Bitmap format is simple and in Windows, is most widely used file format supporting greyscale, indexed and true colour images. It also supports binary images, but not in an efficient manner since each pixel is stored using an entire byte. It is an uncompressed format. It is an example of lossless image representation format.

PNG: The Portable Network Graphics format is a bitmap image format that employs lossless data compression. The PNG format was created to improve and replace GIF format with a new format that doesn't require a patent license to use. It uses DEFLATE Compression algorithm which consists of LZ77 Algorithm and Huffman Coding**.** PNG includes an alpha channel for transparency with a maximum bit depth of 16 bits. PNG meets/exceeds the capabilities of GIF format except that unlike GIF format; it cannot include multiple images in single file to create animations.

Comparison with JPEG:

JPEG has a big compressing ration, reducing the quality of the image, it is ideal for big images and photographs.

PNG is a lossless compression algorithm, very good for images with big areas of one unique color, or with small variations of color.

PNG is a better choice than JPEG for storing images that contain text, line art, or other images with sharp transitions that do not transform well into the frequency domain.

Comparison with TIFF:

TIFF is a complicated format that incorporates an extremely wide range of options. While this makes it useful as a generic format for interchange between professional image editing applications, it makes supporting it in more general applications such as Web browsers difficult.

The most common general-purpose lossless compression algorithm used with TIFF is LZW, which is inferior to PNG and until expiration in 2003 suffered from the same patent issues that GIF did.

GIF: The **G**raphics **I**nterchange **F**ormat is one of the most widely used formats in Internet, largely due to its early support for indexed colour at multiple bit-depths, LZW Compression and its ability in creating animations. GIF can encode two 4-bit pixels into each 8-bit byte. This results in 50% reduction in size over a standard 8-bit indexed bitmap format.

JPEG: The JPEG standard defines a compression method for continuous greyscale and colour [2] images. It was developed by Joint Photographic Experts Group. Depending on the application, JPEG achieves a compression ratio of 1:25, when compressing 24-bit colour images to acceptable quality for viewing. JPEG uses Discrete Cosine Transform based algorithm. The compression technique used in this format makes the co-efficient of quantization matrix bigger when we want more compression and smaller when we want less compression. The algorithm is based on two visual effects of the human visual system. First, humans are more sensitive to the luminance than to the chrominance. Second, humans are more sensitive to changes in homogeneous areas, than in areas where there is more variation (higher frequencies). JPEG is the most used format for storing and transmitting images in Internet. The drawbacks include its limitation to 8-bit images, poor performance on non-photographic images such as line art, poor handling of abrupt transitions in an 8-bit image etc.

JPEG 2000[3]: It is a wavelet-based image compression standard. It was created by the Joint Photographic Experts Group committee with the intention of superseding their original discrete cosine transform-based JPEG standard. It has higher compression ratios than JPEG. It does not suffer from the uniform blocks, so characteristics of JPEG images with very high compression rates. But it usually makes the image more blurred that JPEG.

Despite common usage, JPEG is not a file format, it only provides us a method of compressing and decompressing. The image file extension which we generally find is an instance of JFIF(JPEG File Interchange Format".

VIDEO COMPRESSION

A wide variety of application of digital video applications currently exist from simple low resolution and low bandwidth applications like multimedia, Picture phone to very high resolution and high bandwidth demands such as HDTV. These HDTV has 700 horizontal lines each having 1200 picture elements producing 25-30 interlaced pictures per second. Video signals needs the compression as there is requirement of producing, transmitting and receiving in a given period of time. It basically involves an encoder that receives digital video signals representing the characteristics of a sequence of picture elements. The ultimate goal of video source coding is the bit rate reduction for storage and transmission by encoding a minimum set of information through entropy coding techniques. With practical schemes, a tradeoff between coding performance and implementation complexity is sought. Implemented coding technique depends on VLSI present day technology, bandwidth limit and purpose involved. Video signals can be sent both as lossy and lossless way depending on the image quality.

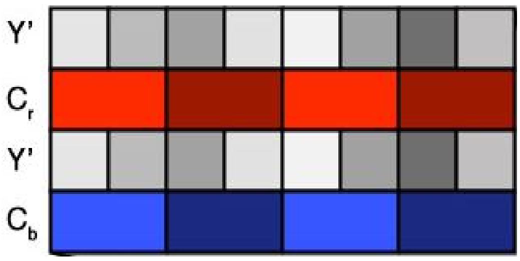
GENERAL CODING ALGORITHM

Video sequences usually have statistical redundancies in both temporal and spatial directions. The magnitude of the particular picture element or pixel can be predicted from nearby pixels within the frame i.e. intraframe coding or from nearby frame, Interframe techniques. Intraframe coding techniques are appropriate to explore spatial correlation to achieve efficient data compression where temporal correlation between nearby frames is small. The prediction error signal is processed using transform coding for exploiting spatial redundancy. The transform coefficients that are obtained by applying a decorrelating transform to the input signal are quantized and entropy coded together with side information like coding modes and motion parameters. In case of nearby frames having high correlation, we use Interframe coding. So, in practical video coding schemes, an adaptive combination of both methods is used to achieve data compression. In entropy coding, the pixel color values that are prequantised to fixed length words are quantized to lower code words for higher probability and higher for low probability. In predictive coding, an approximate prediction of pixel to be coded is made from previously coded information that has been transmitted. The difference between actual and predicted is usually quantized and entropy coded (DPCM) [4]. Prediction error and motion vector are sent to the receiver encoded.

Although all considered standards follow the same basic design, they differ in various aspects that finally results in an improved coding efficiency from one generation of standard to the next.

MPEG-2

In 1994, it was developed and is used for digital television and DVD video optical disc format [5]. Each picture of a video sequence is partitioned into macro blocks that consist of 16x16 luma block and in the 4:2:0 chroma sampling format, 2 associated 8x8 chroma blocks. Chrominance is represented as two color difference [10]



components: U=B-Y and V= R-Y where Y= 0.2126R+ 0.7152G+ 0.0722B and color components are gamma compressed. Here, there are picture types: I, P, and B pictures. In I pictures, all macro blocks are coded in intra coding mode, without referencing other pictures in video sequence. P is transmitted in intra or inter mode using previously coded I or P picture as reference picture. The displacement of P picture is specified by motion vector that is differentially coded. B pictures produced in a coding mode for which prediction signal is obtained from reference pictures and motion vectors are transmitted. For transform coding of intra macro blocks and prediction errors of inter macro blocks, a DCT is applied to blocks of 8x8 samples. Intra DC coefficients are differentially coded using the intra DC coefficient of the block to their left as their predicted value. Quantized Transform coefficients of a block are scanned and transmitted using Huffman coding in combination with run level coding.

MPEG-4

Mpeg-4 AVC still uses the concept of 16x16 macro blocks that can be partitioned into square and rectangular block shapes ranging from 4x4 to 16x16 luma samples. The sizes of the blocks are diluted here for various block types can be chosen for a macro block representation according to which part in the image. Compared to older standards, additional transform is applied to DC coefficients of each chroma and luma blocks and inverse transforms are specified by exact integer operations. For entropy coding it uses 2 methods- context adaptive variable length coding and context adaptive binary arithmetic coding. Latter uses the statistics of previously coded symbols to estimate conditional probabilities for symbols that are transmitted using arithmetic coding.

Interesting feature of mpeg 4 is the object based coding that helps the end users to access the video content rather than frames and its use comes in real time interactive visual communication system [6]. Mpeg 4 uses Video Object Plane (VOP) concepts unlike mpeg 2 i.e. frame based coding. Mpeg 4 is quite complicated compared to mpeg 2 compression as it is designed to achieve high quality videos for multimedia applications at a relatively low bit rate. Mpeg 4 has bandwidth requirement much lesser than mpeg 2. Mpeg 2 is used mainly on DVD and digital television while Mpeg 4 is designed for network or multimedia application.

High Efficiency Video Coding (HEVC) is the current joint standardization project to be finalized in 2013.

ARITHMETIC CODING

Arithmetic coding is a form of entropy encoding used in lossless data compression where fewer bits used for high probable object and higher bits for less occurring object. Unlike Huffman coding, Arithmetic coding encodes the entire message into a single number. The optimal value is -$\log_2 p$ for each symbol of probability P, according to source coding theorem. The encoder divides the interval [0, 1) into sub-intervals, each representing a fraction of the current interval proportional to the current proportional to the probability of that symbol in the current context. Whichever interval corresponds to the actual symbol that is next to be encoded becomes the interval used in the next step. When all symbols have been encoded, the resulting interval identifies the sequence of symbols that produced it. Anything that has the same final interval that is being used can reconstruct the symbol sequence that must have entered the encoder to result in that final interval. It is only necessary to transmit one fraction that lies within that interval. In particular, it is only necessary to transmit enough digits of the fraction so that all fractions. The final interval from which selection has to be made will have infinite numbers. The fraction that gives minimum binary digits has to be chosen. The use of decimal instead of binary leads to inefficiency. Initially calculated probabilities are converted to equivalent low binary digit number interval and then subsequent divisions of intervals are made into lower bit precision.

DCT[8]

DCT is the technique used for compressing the images and videos. The expression for 2D DCT of a 2D Discrete function (in our case it is the image) of block size N_cxN_r is given by

$$S(v,u) = \alpha_{2D}(v,u)\sum_{y=0}^{y=N_r-1}\sum_{x=0}^{x=N_c-1} s(y,x)\cos(t1)\ \cos(t2)$$

$$\text{where } t1 = \frac{(2x+1)u\pi}{2N_c},$$

$$t2 = \frac{(2y+1)v\pi}{2N_c}$$

$$\alpha_{2D}(v,u) = \sqrt{2/N_r}\sqrt{2/N_c}C(v)C(u)$$

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}}, k \neq 0 \\ 1 \ \text{Otherwise} \end{cases}$$

Inverse 2D DCT is given by:

$$s(y,x) = \sum_{y=0}^{y=N_r-1}\sum_{x=0}^{x=N_c-1} \alpha_{2D}(v,u)S(v,u)\cos(t1)\ \cos(t2)$$

The computation of DCT coefficients can be made faster using algorithms like Fast Fourier Transform that is used in case of DCT, similarly in this case. In this case, the 3D DCT is separated into three 1D transforms, computed for rows and column at a particular time, then DCT is applied to all the time frames of the block used. However, it is the DCT coding process that is importantly needed to be done in case of image ad video compression. DCT coding has advantage over Fourier transform that it has good energy compaction, i.e. in the frequency domain, energy concentrated in lesser range than that of the Fourier Transform. This property is essential part of image compression where we aim to minimize as much information needed to be stored. One can use the technique of context based arithmetic coding in DCT block to make DCT video compression more effective because a bulk of bit budget is spent on DCT coefficients. Consequently, how efficiently the DCT coefficients are entropy coded will ultimately determine the compression performance. Statistically, DCT coefficients exhibit diverse behaviours in different types of scene contents and video formats. Learning the local statistics based on contextual information is of great importance for higher coding efficiency. Hence the technique of context based adaptive Arithmetic binary coding technique can be used in this case for efficient coding of DCT coefficients. In this section, we focus towards Image and Video Compression using this technique.

STEPS IN IMAGE AND VIDEO COMPRESSION [7,9]

A video compression has three parts: a forward DCT, a Quantizer, and an entropy encoder. The input is the set of some blocks of the input, say 8x8x8. Forward DCT will result in a set of DCT coefficients. Then a set of redundant data is removed by the next two blocks. The result might be either a lossy compression when the quantizer is used to intentionally reduce the number of levels in such a way that the original information is not completely stored in the compressed format, or a lossless compression, where the original information is completely restored in the compressed output. Each of the step eliminates the bits of the input which doesn't contribute to the information about it. But in this process, they might consume some bits. The bits that are thrown away are the ones that have high correlation. This similar set of information is stored just once. This rate of redundancy depends on the information content of the input image. If the image surface is very smooth, as it is in most of the cases, this redundancy will be very high and the compression will result in greater reduction in number of bits than the case where image surface has many edges.

The DCT is a very efficient coding technique that it does orthogonal transformation that then converts the input into highly unrelated coefficients. Also this enables us to work in frequency domain. Because the edge details can be compressed. Quantizer converts the raw output of CT coefficients into discrete coefficients. Sometimes, it can be used to compress data by saving the number of bits drastically. As mentioned, the adaptive context based arithmetic binary code is used ssto describe the technique of compression here (CADAC). The significance map is first encoded to mark the significant coefficients inside the block of quantised DCT. Then the magnitudes of all non-zero coefficients (Level) are encoded in reverse scanning order. Specific context models are assigned to the significance map according to its position, while the contexts for Level magnitudes are classified according to the successive coefficients (in reverse scanning order). These contexts take full advantage of the localization property of DCT coefficient in a block.

Encoding scheme is as follows:

The DCT quantised output is sent to the system where it is converted to a set of (level, run) pairs. The *level* indicates the non-zero coefficient and the *run* indicates the number of successive zeroes before *Level.* These instances are coded in the reverse coding order until all pairs are coded. At last, the flag EOB (0, 0) is coded to indicate the end of block. For each pair, *Level* is coded followed by run. They are unary binarized into several bits. The *abslevel* is stored in first few blocks of secondary context and *run* is stored in next few bits of secondary context. The End of Block flag is stored in first bit of secondary block. The primary context stores the quantised value of the parameter *LMax* and not the exact value because it might be too high and may consume large number of bits. The value of *accompanying* context *ReverseP* which represents the position of the current (*Level, Run*) pair in the existing sequence is also encoded in primary context. Level, run pairs are encoded by the technique of arithmetic coding. Context weighting is the technique used to encode with *LMaxl* and *ReverseP* taken into consideration.

Thus each block is represented by this set of parameters and these comprise the compressed data of the source image. Thus the context based arithmetic coding is illustrated on DCT coefficients for the Image compression. Below is a flow chart of arithmetic coding[7]:

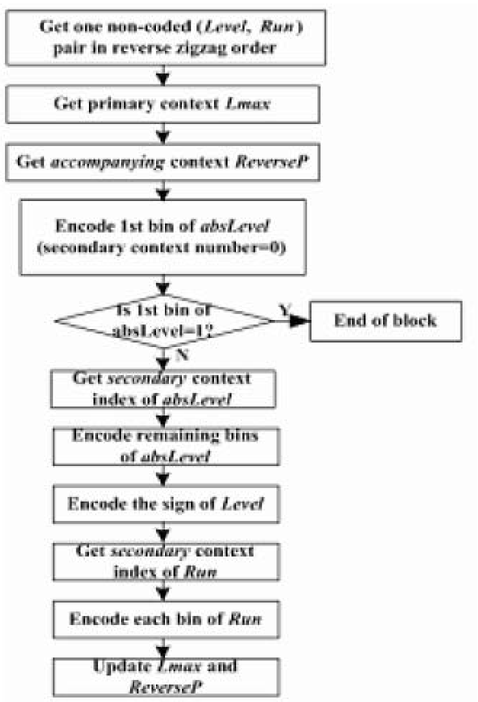


Figure 1. Coding process of a (*Level, Run*) pair

REFERENCES

[1]Wilhelm Burger, Mark J. Burge- “Principles of Digital Image Processing - Fundamental Techniques”.

[2]Yi-Huang Han, Jin-Jang Leou- “Detection and Correction of Transmission Errors in JPEG Images.”

[3] Fangjun Huang, Jiwu Huang, Yun Qing Shi- “Detecting Double JPEG Compression With the same Quantization Matrix.”

[4]Ralf Schafer and Thomas Sikora-“Digital Video Coding Standards and their Role in Video communications”, IEEE, 1995

[5] Jens- Rainer Ohm, Gary J.Sullivan, Heiko Schwarz, Thiow Keng Tan, Thomas Wiegand- “Comparison of the coding efficiency of video coding standards- Including high efficiency Video Coding (HEVC) ”, IEEE , 2012

[6] Hao Chieh Chang, Yi Chu Wang, Mei Yun Hsu and Liang Gee Chen-“Efficient Algrithms and Architecture for MPEG-4 Object-Based Video Coding”, IEEE, 2000

[7] Li Zhang, Xiaolin Wu, Ning Zhang, Wen Gao, Qiang Wang, and Debin Zhao, Context-based Arithmetic Coding Reexamined for DCT Video Compression, IEEE 2007

[8] Discrete Cosine Transform (N. AHMED, T. NATARAJAN, AND K. R. RAO) IEEE 1974

[9] Marc Servais, Gerhard De Jager, Video Compression using the Three Dimensional Discrete Cosine Transform (3D-DCT), IEEE 1997

[10] Graeme Nattress-“Chorma Sampling: an investigation”, Los Angeles Creative Prouser Group, July, 2005