Overview of JPEG2000 technology and error correcting codes for video streaming.

Report by: Oscar Medina Duarte
Erasmus Mundus MERIT programme student
Information Theory and Coding INGI2348
Advisor: Benoit Macq

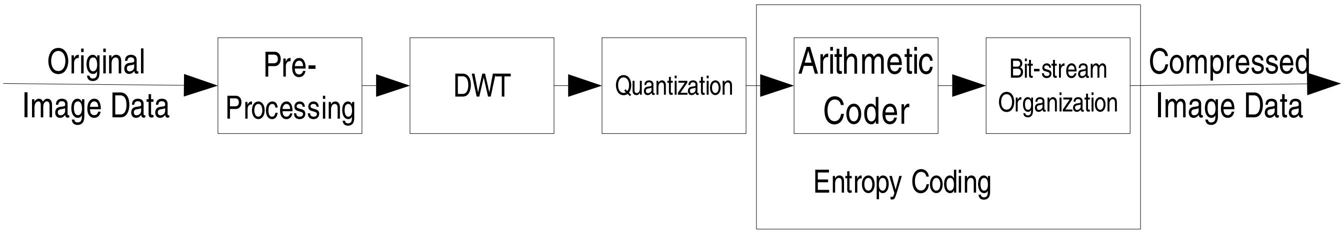
Introduction

JPEG2000 is a Joint Photographic Experts Group's standard proposed to meet some requirements to satisfy some areas where other standards have failed to produce the best quality of performance. This standard is scheduled to be issued in six parts. Part 1, is the core coding system, aimed to satisfy the same solved needs as JPEG. Part 2, this part's aim is to extend Part 1's characteristics to meet specific needs. Part 3, this part defines motion JPEG 2000 which is based in Part 1 with the addition of a file format. Part 4, defines conformance testing to ensure high-quality implementation of the standard. Part 5, defines reference software implementation for Part 1. Finally, Part 6 defines a compound image file format for document scanning and fax applications.

Thanks to JPEG 2000's characteristics on scalable quality and resolution, it is possible to create mechanisms using error correcting codes for unequal error protection depending on the relevance of the image stream's bytes. For this purpose, a Reed-Solomon code based strategy is examined.

Architecture of JPEG2000

For a general idea, we can take a look at the building blocks of JPEG 2000, located in the following image, where we can mention the basic modules of the process in a sequential manner, Pre-processing, Discrete Wavelet Transform (DWT), Quantization, and Entropy encoding, after applying these operations on the original image data, we obtain the compressed image data. The image data can be obtained by applying the same reverse operations in reverse order, Entropy Decoding, Dequantization, Inverse-DWT, Image reconstruction.



The input Image to JPEG 2000 may contain one or more components, typical images have three components as in the RGB system, but this could be extended to up to 16,384 (2^14) components to handle special kinds of images. The values for each component could be signed or unsigned values with a bit-depth in the range of 1-38 bits.

Pre-Processing

The first step to prepare the image for the DWT is to partition the image into rectangular and non-overlapping tiles of equal size (except the borders perhaps), the tile size is arbitrary and can be as large as the image itself of as small as a pixel. Each tile is compressed independently using its own set of parameters, this is useful for applications where the available memory is limited.

The next step, is to make all the values of the components symmetric around zero, this means that unsigned values will be subtracted the value 2^(B-1) where B is the bit-depth of the components. Signed values are already symmetric around zero so they remain untouched. This has no effect on the coding efficiency but it simplifies some implementation issues.

Finally, this values can be subject to a point-wise transformation to decorrelate the color data, for that purpose, all components must have the same bit-depth and dimension. There are two transformations, one is the irreversible color transform (ICT) which can only be used for lossy coding, and the other one is the reversible color transform (RCT) which is capable of recovering the original RGB data with no error.

At the decoder, the decomposed image is subject to the inverse color transformations if needed, and also to the removal of the shift abound zero if it was applied.

Discrete Wavelet Transform (DWT)

The DWT provides a multi-resolution image representation while also improving compression efficiency due to good energy compaction and the ability to decorrelate the image across a larger scale. Integer DWT filters can be used to provide both lossless and lossy compression within a single compressed bit-stream.

The DWT is understood as successive applications of a pair of low-pass and high-pass filtered, followed by down sampling by a factor of two after each filtering operation. The filtered samples that are the output from the forward DWT are referred to as wavelet coefficients. The filter pair is designed in such a way that after downsampling the output from each filter by a power of two, the original signal can still be completely recovered for the case where there are no quantization errors, this is called the perfect reconstruction (PR) property.

To recover the original signal from the wavelet coefficients, the decoder has another pair of low-pass and high-pass filters whose input is upsampled by a factor of two first, and their output is added together to produce the reconstructed signal.

To construct a two dimensional DWT, the filters are applied separately, first (it doesn't actually matter the order since the system is linear), the filter pair is applied to the rows of the image resulting in two sub-bands, and then, the same process is applied to each column, resulting in four sub-bands, this is called the first level wavelet decomposition, the following levels decompositions are found by applying the same filtering process to each lowest frequency sub-band, this process can be repeated until there's no tangible gain in compression. Most of the image's energy is found in the lower frequency sub-bands. From this point is not hard to see that the DWT provides a solution for the multiresolution requirement of the JPEG 2000 standard .

DWT implementation

The JPEG 2000 standard supports two filtering modes: a convolution-based and a lifting-based. For this modes to be implemented, the original signal has to be periodically extended at both ends to ensure that there's always a signal sample that corresponds to the filtering mask. The convolution-based filter consists in performing a series of dot products between the two filter masks and the extended signal, with the disadvantage of a very large amount of required memory, which in many cases is not available. The lifting-based approach solves this problem by operating on the same memory space and overwriting the input samples in the signal with the computed wavelet coefficients. This operation if formed by several steps, first, compute a trivial WT (lazy WT) in which the original signal is splitted in two sub-sequences, one with the odd and other with the even values and then modifying these values by using alternating (odd/even samples) prediction and update steps. A prediction step consists of predicting each odd sample as a linear combination of the even samples and subtracting it from the odd sample. The update step consists of updating the even samples by adding them to a linear combination of the already modified odd samples. The output of the final prediction step will be the high-pass coefficients, while the output of the final update step will be the low-pass coefficients.

Quantization

The transformed samples are then quantized, this operation is usually lossy unless a quantization step of 1 is selected and the coefficients are integers. Uniform quantization with deadzone is used to quantize all the wavelet coefficients. For each sub-band b, a basic quantizer step size $\triangle_b$ is selected bye the user an is used to quantize all the coefficients in that sub-band. The quantization rule is given by :

$$q = sign(y)floor(|y|/\triangle_b)$$

where y is the input to the quantizer, $\triangle_b$ is the quantizer step size, q is the resulting quantizer index, sign(y) denotes the sign of y, |y| denotes the absolute value of y, and floor(x) denotes the largest integer not larger than x.

At the dequantizer, the dequantization rule is given by:

$$z = [q + r\, sign(q)]\triangle_b \text{ , } q \mathrel{!}= 0$$

$$z = 0 \text{ , otherwise}$$

where q is the quantizer index, $\triangle_b$ is the quantizer step size, z is the reconstructed bias, where r = .5 results in midpoint reconstruction (no bias) and r < 0.5 biases the reconstruction towards zero.

Embedded Quantization

JPEG2000 is encoded one bit at a time, going from the MSB to the LSB, during this progressive encoding, the quantized wavelet coefficient is called insignificant if the quantizer index q is still zero. Once the first nonzero bit is encoded, the coefficient becomes significant and its sign encoded. Thus, giving the ability to create an encoded stream that its precision scalable, that is, scalable in quality.

Entropy Coding

The entropy coding is achieved in two stages, first, each bit-plane is divided into code-blocks of the same size, the size is variable, but must be between 4 and 4096 bits due to the entropy encoding scheme used. MQ-Coder is an encoder that uses arithmetic encoding to produce an encoded bitstream from the quantized coefficients.

An arithmetic based encoder was chosen over Huffman coding because Huffman coding poses several limitations for the kind of data to be encoded. If the probability estimations of the symbols have some error, then the Huffman code fails.

The second step is to multiplex the bitstreams coming from the code blocks in an order that optimizes the decoding of the compressed image.

JPEG2000 bit-stream organization

JPEG2000's flexibility is achieved thanks to the structure of the compressed data that is inherited from the methods commented above. These structures partition the image data into color channels (through components), spatial regions (though tiles), frequency regions (Trough sub-bands and resolution layers), and space frequency regions (trough codeblocks). Tiling provides access to the image data over large spatial regions, while independent coding of the codeblocks provides access to smaller units. Codeblocks can be viewed as as a tiling of the coefficients in the wavelet domain. A precinct is a collection of spatially contiguous codeblocks from all sub-bands at a particular resolution level. For each precinct, the compressed data for the codeblocks is first organized into one or more packets. A packet is simply a continuous segment in the compressed codestream that consists of a number of bit plane coding passes for each codeblock in the precinct. A layer is a collection of some consecutive bit-plane coding passes from all codeblocks in all sub-bands and components, each codeblock can contribute an arbitrary number of bit-plane coding passes to a layer.

Progression order

The order in which the packets appear in the codestream is called the progression order and is controlled by specific markers. For a given tile, four parameters are needed to uniquely identify a packet. These are component, resolution, layer and position (precinct). The packets for a particular component, resolution and layer are generated by scanning the precincts in a raster order. All the packets for a tile can be ordered by using nested for loop structures, where each "for" varies one parameter from the above list. By changing the nesting order of the "for loops", a number of different progression orders can be generated. There are five progression orders defined in Part 1 of the JPEG 2000 standard, layer(L)-resolution(R)-component(C)-position(P) [LRCP], RLCP, RPCL, PCRL and CPRL, each of them is useful to explode certain characteristics of the standard.

Video Streaming

The quality degradation in an Internet video transmission compared to a perfect transmission is mainly determined by the packet loss behavior observed at the video decoder. Congested routers are one typical source for packet losses. By posing a real-time constraint on a transmission system as meant by the term streaming, packets arriving too late at the receiver are another source of packet loss.

The robustness against packet losses, especially in cases where a feedback channel is not available, can be increased if the encoder protects the packets with an appropriate forward error correction (FEC) scheme. Since JPEG 2000 provides a scalable bit-stream, we can talk about a scalable video transmission, allowing for supporting client decoders to adapt to the available sending rate according to the amount of congestion in the network. As we have reviewed before, the layer structure of the JPEG 2000 is based on the significance of the bits, therefore, is possible to create a FEC with unequal protection depending on the relevance of the layer.

Layered packetization

Before applying any error correcting code, the data has to be organized in packets containing same amount of data belonging to different layers and then the error correcting code is applied this is done in order to send together error correcting information as well as layer information to avoid the case of being able to decode the enhancement information but not the base layer. Therefore, we have a block of packets in which the layer bits are distributed over a number of k different packets, while the rest of the packets are filled with redundancy information, like Reed-Solomon codes.

Reed-Solomon(RS) codes

The idea of FEC across packets is to transmit additional redundant packets which can be used at the receiver to reconstruct lost packets. RS correcting codes are widely used because they are the only known non-trivial maximum distance separable codes.

RS coding system is based on groups of bits, such as bytes, rather than individual 0's and 1's, making it particularly good at dealing with bursts of errors. We can say that a given RS code: RS(n,k) has length n and encodes k information symbols, this means that the encoder takes k data symbols of m bits each and adds parity symbols to make an n symbol codeword. There are $n-k$ parity symbols of s bits each. A Reed-Solomon decoder can correct up to t symbols that contain errors in a codeword, where $t = (n-k)/2$.

References

- Majid Rabbani, Rajan Joshi, **An overview of the JPEG 2000 still image compression standard**, Signal processing: Image Communication 17 (2002) 3-48.
- Charilaos Christopoulos, **THE JPEG2000 STILL IMAGE CODING SYSTEM: AN OVERVIEW**, ublished in IEEE Transactions on Consumer Electronics, Vol. 46, No. 4, pp. 1103-1127, November 2000.
- Majid Rabbani, Diego Santa Cruz, **The JPEG2000 Still-Image Compression Standard**, Thessaloniki, Greece, October 11, 2001.
- K. Stuhlmuller, M.Link, B. Girod, U.Horn, **Scalable Internet Video Streaming With Unequal Error Protection**, Packet Video Workshop 99, April 1999, NY.