

JPEG XR an Image Coding Standard

Savita S. Jadhav and Sandeep K. Jadhav

Abstract—JPEG XR is an emerging image coding standard, based on HD Photo developed by Microsoft technology. It supports high compression performance twice as high as the de facto image coding system, namely JPEG, and also has an advantage over JPEG 2000 in terms of computational cost. JPEG XR is expected to be widespread for many devices including embedded systems in the near future. This review-based paper proposes a novel architecture for JPEG XR encoding. This paper gives discussion of image partition and windowing techniques. Further frequency transform and quantization is also addressed. A brief insight into Prediction, Adaptive Encode and Packetization has been provided in the paper.

Index Terms—JPEG XR, encoder, PCT, quantization, adaptive prediction, adaptive entropy coding.

I. INTRODUCTION

Many advanced multimedia applications require image compression technology with higher compression ratio and better visual quality. High quality, high compression rates of digital image and low computational cost are important factors in many areas of consumer electronics, ranging from digital photography to the consumer display equipment applications such as digital still camera and digital frame. These requirements usually involve computationally intensive algorithms imposing trade-offs between quality, computational resources and throughput. For high quality of digital image applications, the extended range of color range has becoming more important in recent emerging need of the consumer product. In the past, the digital cameras and the display equipments in the consumer market typically had 8 bits per channel. Today the condition is quite different. In the consumer market, digital cameras and the desktop display panels also have 12 bits of information per channel. If the information per channel of digital image is still compressed into 8 bits, 4 bits of information per channel is loss and the quality of the digital image is limited. Due to the improvement of the display equipments, the JPEG XR is designed for the high dynamic range (HDR) and the high definition (HD) photo size. JPEG XR which is already under organized by the ISO/IEC Joint Photographic Experts Group (JPEG) Standard Committee is a new still image coding standard and derived from the window media photo [1],[2]. The XR of JPEG XR means the extended range. It means that JPEG XR supports the extended range of information per channel. The goal of JPEG XR is to support the greatest possible level of image dynamic range and color precision, and keep the device

implementations of the encoder and decoder as simple as possible. For the compression of digital image, the Joint Photographic Experts Group, the first international image coding standard for continuous-tone natural images, was defined in 1992. JPEG is a well-known image compression format today because of the population of digital still camera and Internet. Another image coding standard, JPEG2000 [3], was finalized in 2001. Differed from JPEG standard, a Discrete Cosine Transform based coder, the JPEG2000 uses a Discrete Wavelet Transform based coder for better coding efficiency. The JPEG2000 not only enhances the compression, but also includes many new features, such as quality scalability, resolution scalability, region of interest, and lossy/lossless coding in a unified framework. However, the design of JPEG2000 is much complicated than the JPEG standard. The core techniques and computation complexity comparisons of these two image coding standard are shown in [4].

II. JPEG XR

The coding flow of JPEG XR is shown in Fig. 1. First of all, color space transformation is applied to an input image to transform RGB image to YUV image [5]. Then a transform called photo core transform (PCT) is applied to decompose an image into frequency components. The PCT is applied to a rectangular area called a macroblock. To reduce block noise, this occurs around macroblock boundaries, a transform called photo overlap transform (POT) is used with the PCT. These transforms are based on lapped biorthogonal transforms. Next, the transformed coefficients are quantized. In an inter-block coefficient prediction process, the quantized coefficients of a block are replaced by prediction errors to enhance compression rate. After the inter-block coefficient prediction, an adaptive scanning is processed and coefficients are rearranged from two-dimension form to one dimensional form. Finally, the scanned coefficients are entropy coded using adaptive Huffman tables. Hereafter, the details of each process are described.

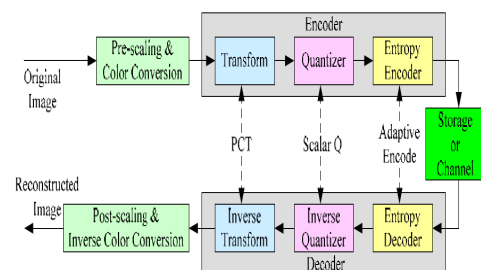


Fig. 1. The data flow of JPEG XR encoding

Manuscript received February 4, 2012; revised March 11, 2012

Savita S. Jadhav is with the Department of Electronics and Telecommunication Engineering, India.

Sandeep K. Jadhav is with Shivaji University, Maharashtra, India.

A. Color Conversion

1) Color Space

RGB color space is the most common way nowadays to represent color in computer display. One color is decomposed into red, green and blue components and stored separately. The drawback of RGB color space is that the three components have equal weights which are not so good for compression. The alternative triplet composed color space called YCbCr has advantage for compression.

Y represents Luminance which indicates the intensity of a picture while CbCr represents chrominance which describes how colorful a picture is. Cb gives the Blueness and Cr gives the redness. Research has proven that human eyes are more sensitive to Luma components which give an unequal importance for these three components. Then we could down-sample Cb and Cr components to reduce storage without visibly decreasing the image quality. JPEG use a linear transformation from RGB to YCbCr specified as follow:

$$Y = 0.299R + 0.587G + 0.114B. \quad (1)$$

$$Cb = -0.1687R - 0.3313G + 0.5B + 2^{Ps-1}. \quad (2)$$

$$Cr = 0.5R - 0.4187G - 0.0813B + 2^{Ps-1}. \quad (3)$$

In Equations "1", "2", "3" Ps denotes the precision of sample like Ps = 8 means that each color component has 8 bits. The transform is a little bit lossy due to roundoff error. JPEG XR specifies a lossless color space conversion, given by:

$$Y = G + \frac{[R-G + \frac{[B-R]}{2}]}{2}. \quad (4)$$

2

$$Cb = - \frac{[R-G + \frac{[B-R]}{2}]}{2}. \quad (5)$$

$$Cr = B - R. \quad (6)$$

JPEG XR is an integer-based standard. Equations "4", "5", "6" do not contain the rounding operation; therefore the inverse operation can be performed in decoder without loss.

2) Down Sampling

Down sampling is the synonym of sub sampling. The reason we use YCbCr color space is that Cb and Cr can be down-sampled without visible quality degrade. JPEG XR support YUV444, YUV422 and YUV420 down sampling ratio. It means, for example, YUV444, has a horizontal x vertical block number of Y: 4 x 4; U: 4 x 4, V: 4 x 4 in each macro block [6]. In this way, we can have Table I:

TABLE I: DOWN SAMPLING RATIO OF YUV COLOR SPACE

Down sampling Ratio	Y plane size	U plane size	V plane size	YUV ratio
YUV444	4x4	4x4	4x4	1:1:1
YUV422	4x4	2x4	2x4	2:1:1
YUV420	4x4	2x2	2x2	4:1:1

B. Image Partition and Windowing

The smallest element of an image is a pixel. Each 4x4 set of pixels is grouped as a block. Then each 4x4 set of blocks is

grouped as a macroblock. A set of Macroblocks can then be grouped as a tile, though the number of macroblocks included along the width and height may vary between tiles [6]. At the highest level of the hierarchy, the tiles come together to form the entire image. A diagram of this partitioning is shown in Fig. 2.

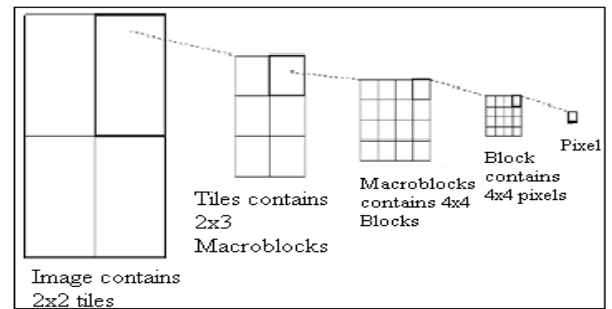


Fig. 2. An example of hierarchical image partitioning.

1) Tiles

Tile size has to be decided at the beginning of JPEG XR compression since it is largely related to the hardware architecture. Each tile in an image is processed independently like a small image. Dividing the image into small tiles has the following advantages:

- Flexibility: Tiles in the same image could have different process configurations.
- Memory saving: In adaptive prediction phase, one row of coefficients has to be saved. Smaller tile means less horizontal length which occupies less memory. It reduces either the memory accessing or silicon area.
- Robustness: Tile gives the system robustness. If one error occurs in one tile, the others will not get affected.

These features make JPEG XR more flexible and suitable for hardware implementation especially when memory size constraint is in high priority. Nevertheless, using too many tiles will decrease the compression efficiency since redundancy among tiles cannot be extracted. The compression ratio decreases when number of tile increases.

2) Macro Blocks

Macroblock is a basic data unit in JPEG XR. The size of a macroblock is predefined by the standard. In Luma plane, each macroblock consists of 16x16 pixels in which 1 is DC component, 15 are LP component, 240 are HP component. In Chroma plane of YUV422 and YUV420, each macroblock is 8 x16 and 8 x8 respectively. PCT, POT, quantization, prediction stages are all designed to process one macroblock each time.

3) Blocks

Block is the smallest data unit in JPEG XR. Each block consists of 4x4 pixels and one macroblock consists of 4 x 4 blocks.

C. Pre-Filtering

The pre-filtering scheme used in JPEG XR compression is called Picture Overlap Transform (POT). Frequency transform taking place in blocks cannot take advantage in solving redundancy across block boundary. Therefore some artifacts will come out in the image compressing. POT is a new feature in JPEG XR that could somehow "smooth" the artifacts on block boundaries. POT has three overlap levels:

one non-overlapping, one-level overlapping and two-level overlapping.

- Non-overlapping does not use any pre-filtering operation and is the fastest way but has low compression ratio.
- One-level overlapping has one stage POT. It has better compression ratio in the price of more complex computation and more processing time.
- Two-level overlapping mode has two-stage POT means highest computation complexity, memory usage and processing time. But at the same time, we get best PSNR under relatively low bit-rate.

Compared to PCT, which is only performed inside each 4x4 pixels block, POT is more complex and memory consuming since it performs not only inside each macroblock but also between the boundaries of adjacent macroblocks. The POT consists of two major operations: one is 4 x4 Pre-filtering module specified in Figure 3; the other is 4 point Pre-filtering module.

- 4x4 Pre-filtering: The 4x4 Pre-filtering is applied to all macroblock junction and fully interior areas. We can see the POT insists on 4 sub-operations: 2x2 Hadamard Transform TH, 2-point Forward Rotation TR, 2-point Forward Scaling TS, 2 x 2 Odd odd Transform Toddodd.
- 4-point Pre-filtering: Linear 4-point pre-filters are applied to the edge straddling 2 x 4 and 4 x 2 boundary areas of the image. It has 2-point Forward Rotation TR, 2-point Forward Scaling TS and a few micro-operations.

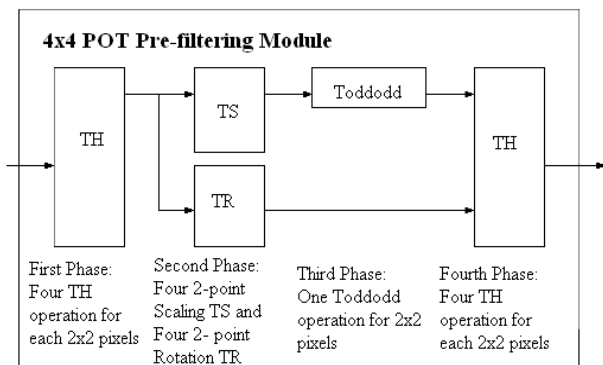


Fig. 3. Block diagram of 4 x 4 pre-filtering

D. Frequency Transform

The reason why we transform spatial data into frequency domain is: Human eyes are usually less sensitive to high frequency component so that it could be removed to reduce overhead. Frequency transform could extract frequency components which are uniformly distributed in spatial data and put same frequency components together. Once the high frequency component has been put together, it is easy to remove them by using quantization. That is why the quantization parameters in high frequency band are usually much larger than the low frequency and DC bands. The frequency transform function used in JPEG XR is called Photo Core Transform (PCT). After pre-filtering, it transforms an image from spatial domain into frequency domain, like the old Discrete Cosine Transform (DCT) from JPEG does. The differences are:

- PCT defines three clear frequency bands DC, Low pass, High pass which DCT does not. The advantage that

clearly defines three bands is they could be processed separately and is more flexible.

- JPEG XR usually uses a 4 x 4 pixels block as the minimum processing unit which is more fine-grained than the 8 x 8 block used in JPEG.

PCT has only one module which is specified in Figure 4[6]. It is simpler than POT since the input is always a 4x 4 block. Three sub-operations TH, Todd, Toddodd are employed in PCT. The 4x4 PCT module processes one block in 2 phases specified in Figure 5. In the first phase, 2 x2 Hadamard transform TH is applied to the corners, boundaries and interior areas of the block in parallel. In the second phase, TH is applied to the top-left corner while Odd transform Todd is applied to the top-right and bottom-left. The bottom-right corner is processed by odd odd Transform Toddodd.

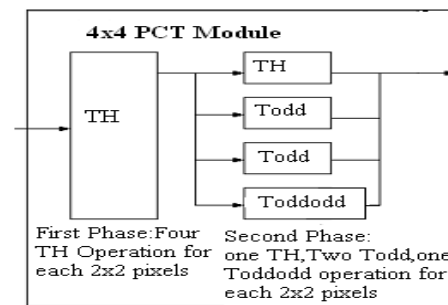


Fig. 4. Block diagram of 4 x 4 PCT

E. Quantization

The quantization is the process of rescaling the coefficients after the transform process is applied. The quantization uses the quantized value to divide and round the coefficients after PCT transformed to an integer value. For the lossless coding mode, the quantized value = 1. For the lossy coding mode, the quantized value > 1. The quantization of JPEG XR use integer operations. The advantage of integer operation keeps the precision after scaling operations and only uses shift operation to perform the division operations. The quantization parameter is allowed to differ across high pass band, low pass band and DC band. It varies to different value according to the sensitivity of human vision in different coefficient bands.

F. Prediction

Adaptive prediction is used in JPEG XR. The prediction performs only when the system makes sure two neighbor macroblocks have big enough similarity. It is called "adaptive" because it dynamically changes they predict direction to one of the neighbor macroblocks which has the biggest similarity with the current macroblock. The predict direction could be from left, top or top-left and the prediction scheme is variant across different frequency bands. The detailed DC, LP and HP prediction will be introduced in the following sections.

DC Prediction: The DC prediction is done in between macroblocks. It can happen from top, left and top-left. The DC level of the one deemed most similar is selected and the difference between that one and the present one is calculated. Figure 6 gives an example how DC prediction happens between macroblocks. Since there is only one DC coefficient in each macroblock, four values represent four macroblocks.

In the example, value 54 is the DC value in current macroblock. Comparing among the three current directions, value 53 in the left have the biggest similarity and then is chosen as the subtrahend. The result value 1 is finally passed on to the next stage, adaptive scan.

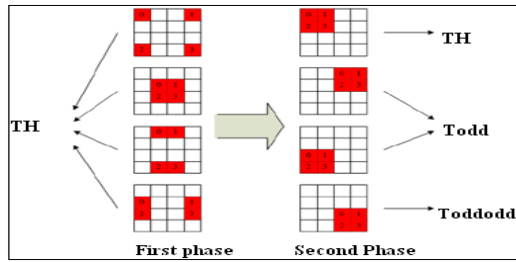


Fig. 5. Two phases sub-operations on each block

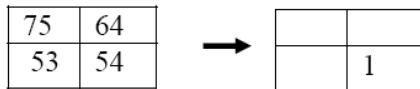


Fig. 6. Example of DC-prediction

LP Prediction: As DC prediction, LP prediction also happens between macroblocks. But it has only two predict directions which are from left and top. LP predicts direction is decided by the DC prediction mode if the quantization parameters are the same for the two macroblocks involved, otherwise it is skipped altogether. If DC predicts direction is from top-left, the prediction will not perform in LP.

HP Prediction: Different from the DC and LP prediction, the HP prediction performs within each macroblock. The intention is to extract and save the difference among blocks inside each macroblock. The HP predicts direction could be from left, top or not at all.

G. Adaptive Encode and Packetize

JPEG XR uses adaptive entropy coding which has several advanced features to reduce data overhead: First of all, not all of the bits need to be encoded. A process called normalization is used to decide which part of the bits should be kept as plain or be encoded. The plain bits in the DC and LP bands are called refinement while in the HP band they are called flexbits. After the bits are separated, the plain bits will be directly written into the bit stream while the rest of the bits need to be encoded by using a combination of both Run Level Encoding (RLE) and the Variable Length Coding (VLC). Since each macroblock has only one DC component, the DC band does not have Run Level encoding and some parts of the VLC coding. This makes the DC part simpler than the other two. Moreover, two data dependent loops are applied in this architecture. One is the normalization model bits, the other is the adaptive VLC code table selection. The final bit stream consists of three parts in each frequency band. They are Coded Bit Pattern (CBP), VLC coded bits and plain bits. CBP signals whether or not any non-zero coefficients are present in the corresponding band for this macroblock. It is a bit mask with a bit for each YUV color plane.

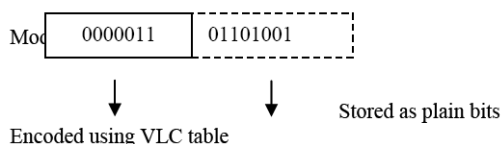


Fig. 7. Typical example of normalization

The actual coefficients are divided into VLC coded bits and plain bits by normalization. Figure 7 gives an example how VLC coded bits and plain bits are separated. Assume each coefficient has 16 bits. If the number of model bits equals 8, the 8 LSB of the coefficient will be kept as plain bits and directly written to the bit stream. The remaining 8 MSB will later be encoded using the VLC table.

III. CONCLUSION

Compared with the JPEG2000, the coding flow of the JPEG XR is simple and has lower complexity in the similar PSNR quality at the same bit rate. Hence, the JPEG XR is very suitable for implementation with the dedicated hardware used to manage HD photo size images for the HDR display requirement. Also with optimized configuration the PSNR performance of JPEG XR is on average 2 to 4 db better than JPEG. The advantage is more obvious especially when the compression ratio is increased. That means JPEG XR keeps better image quality when bit-rate is low. More image details are kept. JPEG XR uses 4 x 4 block as the minimum processing unit, and it is more fine-grained than the 8x8 block used in JPEG.

REFERENCES

- [1] D. D. Giusto and T. Onali, "Data Compression for Digital Photography: Performance comparison between proprietary solutions and standards," *IEEE Conference on Consumer Electronics*, pp. 1-2, January 2007.
- [2] B. Crow, "Windows Media Photo: A new format for end-to-end digital imaging," *Windows Hardware Engineering Conference*, May 2006.
- [3] JPEG 2000 Part I, ISO/IEC JTC1/SC29/WG1 Standard, Mar. 2000.
- [4] Y. W. Huang, B. Y. Hsieh, T. C. Chen, and L. G. Chen, "Analysis, Fast Algorithm, and VLSI Architecture Design for H.264/AVC Intra Frame Coder," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 378-401, Mar. 2005.
- [5] C. Ho Pan, C. Yen Chien, W. Min Chao, S. C. Huang, and L.G. Chen, "Architecture Design of Full HD JPEG XR Encoder for Digital Photography Applications" *IEEE Transaction on Consumer Electronics* Volume: 54, pp. 963 - 971, 2008.
- [6] L.Yu, "Evaluating and Implementing JPEG XR Optimized for Video Surveillance", Master thesis, Dept. Computer Eng. ,Linkoping Univ. ,Linkoping, Sweden, March 2010.



Savita S. Jadhav received the B.E and M.E. degrees in electronic engineering from Shivaji University, Maharashtra, India in 2002 and 2009, respectively. In 2002, she started her academic career as a Faculty member of Rajarambapu Institute of Technology, Maharashtra and currently with Pd. Dr. D.Y.Patil Institute of Engineering and Technology, Pimpri, Pune, Maharashtra, India, where she is a Lecturer with the Department of Electronics and Telecommunication Engineering. She is the author of seven research papers. Her research interests include signal compression, joint source-channel coding, multiple description coding and image processing. Mrs. Savita is Member of ISTE.

Sandeep K. Jadhav received the B.E degree in electronic engineering from Shivaji University, Maharashtra, India in 2003. In 2003, he started his academic career as an employee of Shapoorji & Paloonji Co. Ltd.Pune and currently with same organization, where he is a Sr. Manager. His research interests include signal compression and image processing.