

Module 6

STILL IMAGE COMPRESSION STANDARDS

Lesson 18

JPEG-2000 – Region of Interests Coding

Instructional Objectives

At the end of this lesson, the students should be able to:

1. State the objective of Region of interests (ROI) coding.
2. Explain the basic principle of scaling based method for ROI encoding.
3. Outline the steps involved in implementing the scaling based method.
4. Generate the ROI masks for each stage of subband decomposition.
5. Explain the limitations of scaling based method in arbitrarily shaped ROI.
6. Outline the MAXSHIFT method for ROI coding.
7. State the advantages of MAXSHIFT method, as compared to general scaling based method.

18.0 Introduction

In lesson-17, we have studied the basic architecture of the advanced still image coding standard JPEG-2000, along with its basic filtering scheme and embedded bit-stream structure. JPEG-2000 offers several new features, were not addressed in the prior coding standards - one of these is the region-of-interest (ROI) encoding and the other is the error resiliency. The former is the topic of study in the current lesson.

While viewing a still image, the viewer hardly concentrates on the entire image, especially if the image is of fairly large size. The attention is usually focused on a specific region of interest, which should be encoded and reconstructed with good fidelity, even by sacrificing quality at the other regions of the image. Though the earlier standards did not explicitly address the ROI quality, some preferential treatment of the ROI could always be given by performing finer quantization of transform coefficients at the ROI, as compared to coarser quantization for the rest. This requires explicit transmission of shape information of ROI by the encoder, requiring more bits. The JPEG-2000 standard, based on EBCOT algorithm, which performs multi-pass encoding of bit-planes of subband coefficients starting with the most significant bit-plane for the code-block, the non-ROI coefficients can be down-scaled with respect to the ROI coefficients, so that the bit-planes of the latter are transmitted first in the embedded bit-stream that is composed, without explicitly transmitting any shape or binary map of the ROI. We are first going to explain this approach, known as the *scaling based method*. The success of this method depends upon the computation of ROI masks at the encoder for different subbands. This requires an insight into the analysis and synthesis filtering so as to identify the wavelet coefficients which are used to compose the ROI pixel. This aspect will be covered. We shall thereafter present the MAXSHIFT algorithm, which is more effective in encoding the ROI as compared to the generalized scaling based method.

18.1 Scaling Based Method for ROI encoding

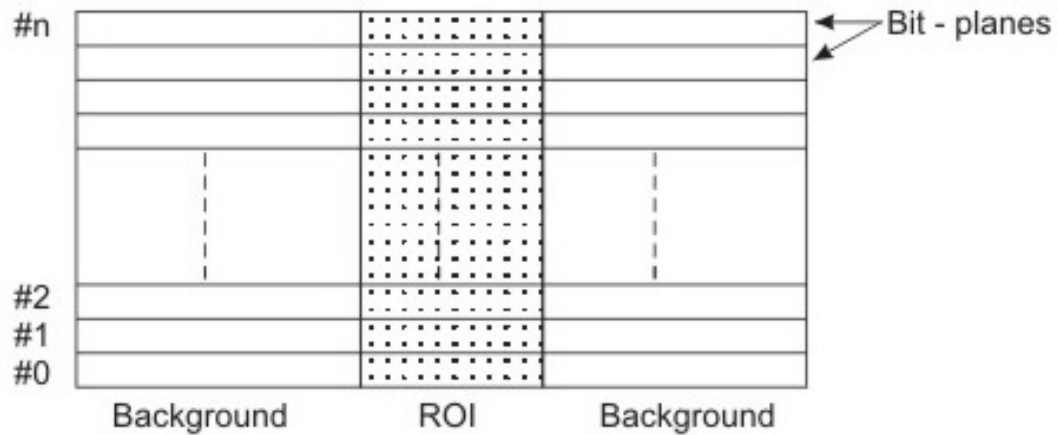


Fig18.1: Bit-planes of subband coefficients without special consideration to ROI

Fig.18.1 indicates the bit-planes of the subband coefficients, where no special consideration is given to the region of interest and fig.18.2 illustrates the scaling based method for ROI encoding.

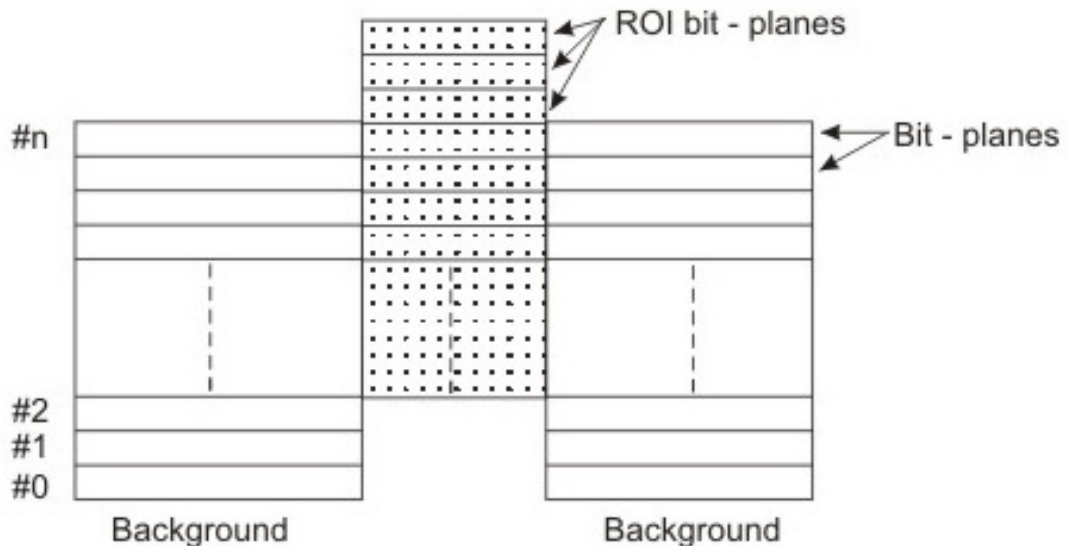


Fig18.2: Scaling based ROI Encoding.

As shown in fig.18.2, the coefficients in the ROI are scaled up with respect to the non-ROI coefficients. This increases the total number of bit-planes for encoding, but the most significant bit planes in this situation are occupied by the ROI coefficients, so that in the EBCOT algorithm for JPEG-2000, where the encoding starts with the most significant bit planes, the initial part of the

embedded bit-stream will contain the ROI only. Only after a certain number of bit-planes, the encoding of non-ROI coefficients and the lower significant bit-planes of the ROI begins. If the encoding is allowed to continue till the least significant bit-plane, the ROI, as well as non-ROI regions will be having good reconstruction quality, but if the bit-stream is truncated in between, the reconstruction fidelity of ROI will be better than the non-ROI region.

18.2 Implementation of Scaling Based Method

The scaling based approach stated above has been implemented as per the following steps in JPEG-2000 standard:

1. Calculate the wavelet transform of the image and decompose into subbands.
2. Define the region of interest through a binary mask for the original, that is, the highest resolution image, such that the mask bits having a value of “1” corresponds to the ROI and those having a value of “0” correspond to the non-ROI.
3. Propagate this mask to the subbands at different resolutions. The algorithm for generation of ROI masks for different subbands will be presented in the next section.
4. Quantize the wavelet coefficients.
5. Downscale the coefficients in the non-ROI by a specific scaling value.
6. Perform arithmetic coding of the resulting coefficients, starting with the most significant bit plane.

In addition to the subband coefficients, the bit-stream has to carry (i) the scaling value assigned to the ROI and (ii) the coordinates of the ROI. If instead of specific scaling value, an arbitrary scaling value is used, along with an arbitrarily shaped ROI, the shape information of the ROI has to be encoded and this increases the number of bits considerably.

18.3 Generation of ROI mask for different subbands

The first step involved with the generation of ROI masks for different subbands is to specify an ROI for an image at its highest spatial resolution, that is, the original. An example of this is shown in fig.18.3. The ROI may be either specified interactively by a user or automatically through some ROI features, such as skin color if the image is a color head and shoulder with face as the foreground or using motion as a clue if the ROI includes a moving object against static background. The ROI will be arbitrarily shaped in general and may contain holes

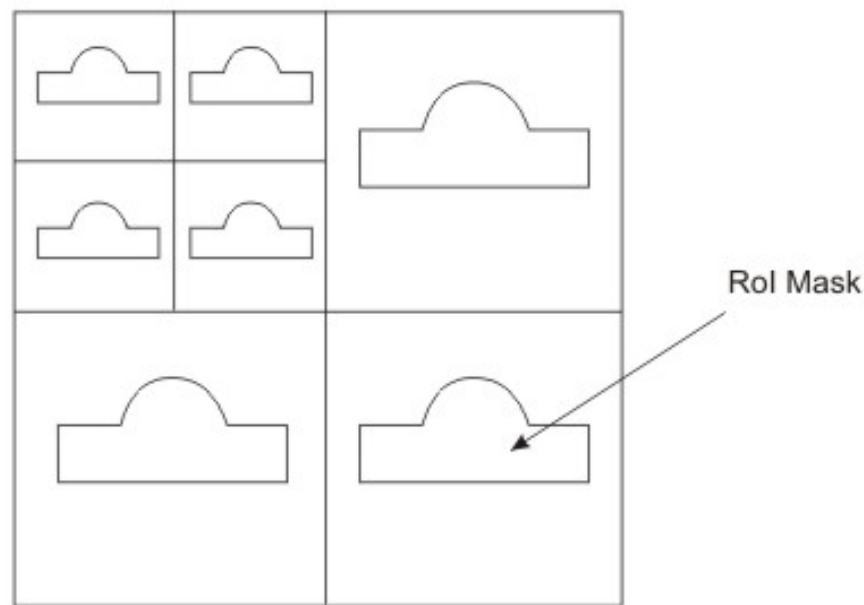


Fig. 18.3 An example ROI mask generation for different subbands

and disjointed combination of segments. The ROI input is fed in the form of a binary mask that has a value of “1” for the ROI and “0” for the rest. Once this is specified for the highest resolution original image before it is wavelet decomposed, the generations of masks for subbands after the wavelet decompositions proceed as follows.

Each level of 2-D multi-resolution decomposition is achieved by iterative application of the one-dimensional decomposition – along the rows and then along the columns of pixels. In each level of wavelet decomposition, the inverse wavelet transform is analyzed to see which of the subband coefficients actually contribute to the mask, so that the mask that exists before the decomposition can be obtained back after the inverse wavelet transform. The corresponding subband coefficients that contribute to the mask therefore serve as the updated mask after decomposition and this is done at each level of decomposition- along the rows and also along the columns. This is illustrated in fig.18.3. Click on this diagram repeatedly to see how the ROI mask is generated for each level of decomposition.

To illustrate how the corresponding subband coefficients for ROI masks are decided, let us consider the case of a 5/3 tap filter, which we had discussed in lesson-17 (Section-17.5). We show the synthesis filter structure in fig.18.4 to illustrate the synthesis of two subband coefficients $X(2n)$ and $X(2n+1)$ losslessly.

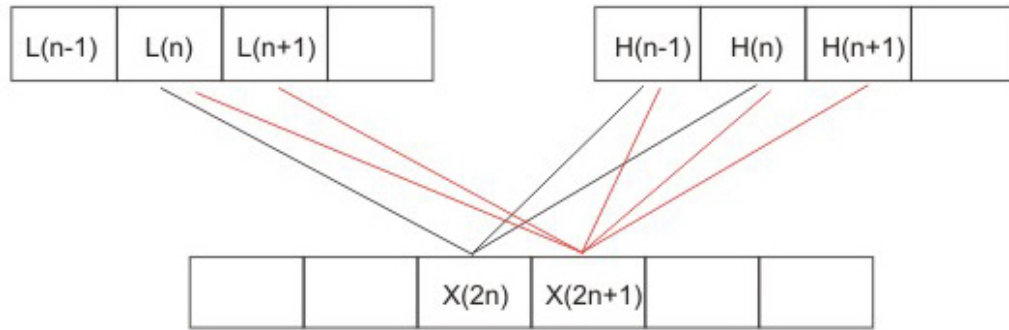


Fig.18.4: Lossless synthesis filter structure for $X(2n)$ and $X(2n+1)$

It is seen that following coefficients: $L(n), L(n+1), H(n-1), H(n), H(n+1)$ are needed. Hence, all these should belong to the ROI for the decomposed subbands L and H .

Since all the contributing coefficients from the lower resolution subbands are included in the ROI, the ROI increases in size with each level of decomposition.

18.4 Limitations of Scaling Based Methods

The scaling based method, presented in Section-18.2 has the following limitations:

1. For arbitrarily shaped ROIs having arbitrary scaling values, the shape information has to be encoded by the encoder and a shape decoder has to be used at the decoder to recover the shape information. This makes the design of encoders and decoders more complex and increases the bit rate.
2. The decoder has to generate the ROI mask and this increases the computational complexity and memory requirements at the decoder.
3. There are no independent controls over the quality of the ROI and non-ROI.

18.5 MAXSHIFT Method

The limitations posed by scaling based methods are overcome in another ROI encoding approach, known as the MAXSHIFT, which was incorporated in Part-I of JPEG-2000 standard. The basic philosophy of this algorithm is illustrated in fig.18.5.

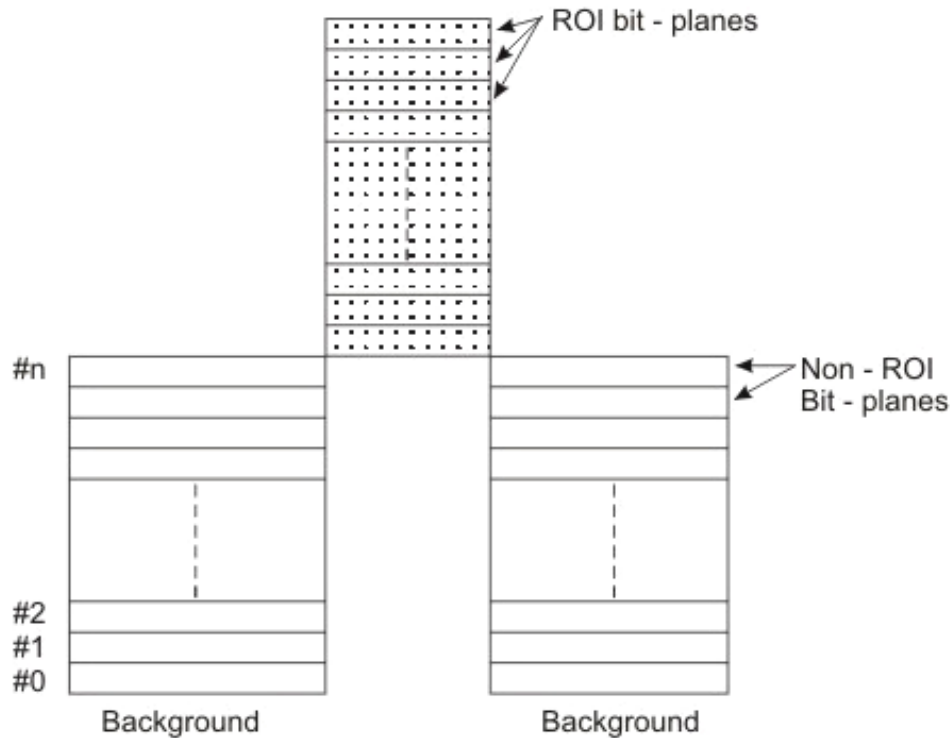


Fig 18.5: MAXSHIFT Method for ROI encoding

The MAXSHIFT method is actually an extension of the general scaling based method in which no desired scaling value is specified. The scaling value is such that the minimum value of the magnitude of coefficient in ROI is larger than the maximum coefficient magnitude (M) in the non-ROI area and hence all the bit-planes of the ROI are embedded first in the bit-stream before the bit-planes pertaining to the non-ROI area. This is achieved by having a scaling value of $n = \lceil \log_2 M \rceil$, such that $2^n \geq M$. Hence, while encoding, any coefficient having a magnitude greater than or equal to 2^n belong to the ROI. Otherwise, it belongs to the non-ROI. Thus, no shape coding for arbitrarily shaped ROI are needed. No arbitrary scaling value is possible, as the scaling value will always depend upon the maximum coefficient magnitude. As the non-ROI coefficients are down-scaled, the decoder should take care to upscale the received non-ROI coefficients.

18.6 Advantages of MAXSHIFT method

The MAXSHIFT method offers several advantages over the generalized scaling based method. These are:

1. Encoding of arbitrarily shaped ROI is possible without the need for explicit shape encoding and decoding and without the need for calculating the ROI mask at the decoder.
2. Since the bit-planes with ROI are complete separate from the bit-planes of non-ROI, the number of bit-planes for ROI and non-ROI may be decided independently. This allows independent controls over the bit rate and quality of ROI and non-ROI.
3. These added features are achieved only by a marginal (1% or so) increase in the bit rate.

18.7 Encoding and Decoding Aspects of MAXSHIFT

It is possible to have a misconception that in MAXSHIFT method, the decoder would first receive the entire ROI before receiving any background information. This is not correct, since the EBCOT encoded bit-stream has resolution scalability and at each resolution level, both ROI and non-ROI are encoded, before the next higher resolutions are considered. At each resolution and at each quality layer, the entire image will be received with the ROI first. It may also be noted from the discussions we presented in Section-18.3 that at each level of decomposition, the ROI mask expands and covers most of the areas in the lower resolution subbands. In MAXSHIFT method, the decoder does not need to generate the mask and hence any mask can be used at the encoder. For example, the mask can encompass the entire LL subband, so that at an early stage of progressive transmission, the encoder sends a low-resolution version of the entire background. This is done by the scaling of all the quantized transform coefficients of the entire subband. The user can decide at which subband he starts receiving the ROI and it is not necessary to wait for the entire ROI before any background information is received.