ELSEVIER

Signal Processing: *Image Communication* 17 (2002) 105–111

SIGNAL PROCESSING:
IMAGE
COMMUNICATION

www.elsevier.com/locate/image

Region of interest coding in JPEG 2000

Joel Askelöf*, Mathias Larsson Carlander, Charilaos Christopoulos

Ericsson Research, Corporate Unit, Ericsson Radio Systems AB, S-164 80 Stockholm, Sweden

Abstract

This paper describes the functionality in the JPEG 2000 Part 1 standard, for encoding images with predefined regions of interest (ROI) of arbitrary shape. The method described is called the Maxshift method. This method is based on scaling of the wavelet coefficients after the wavelet transformation and quantization. By sufficiently scaling the wavelet coefficients used to reconstruct the ROI, all the information pertaining to the ROI is placed before the information pertaining to the rest of the image (background), in the codestream. By varying the quantization of the image and by truncation of the codestream, different quality for the ROI and for the background can be obtained. A description is also given of how the wavelet coefficients that are used to reconstruct the ROI (ROI mask) can be found. Since the decoder uses only the number of significant bitplanes for each wavelet coefficient to determine whether it should be scaled back, an arbitrary set of wavelet coefficients can be scaled on the encoder side. This means that there is no need to encode or send the shape of the ROI. This paper also describes how this can be used to further enhance the ROI functionality. The results in this paper show that the Maxshift method can be used to greatly increase the compression efficiency by lowering the quality of the background and that it also makes it possible to receive the ROI before the background, when transmitting the image.

Keywords: Wavelet; Image compression; JPEG 2000; Region of interest; ROI; Mask; Scaling; Maxshift

1. Introduction

In many applications, certain regions of an image to compress are of greater importance than other regions. In such applications it may be desirable to compress the image with higher quality for the ROI than for the background. It may also be desirable that when the compressed image is transmitted, the information pertaining to the ROI is transmitted earlier than the information pertaining to the background. The ROI functionality included in JPEG 2000 provides both of these features. This paper describes the method used in JPEG 2000 for encoding images with an ROI, called the Maxshift method. The principle of the Maxshift method is to scale up (shift up) the wavelet coefficients obtained from the wavelet transformation (for an overview of JPEG 2000 see [2]) so that the bits associated with the ROI are placed in higher bit-planes than the bits associated with the background. Then, during the embedded coding process, those bits are placed in the codestream before the background parts of the image. In Section 2 the Maxshift method is described for

*Corresponding author.

E-mail addresses: joel.askelof@era.ericsson.se (J. Askelöf), mathias.carlander@era.ericsson.se (M.L. Carlander), charilaos.christopoulos@era.ericsson.se (C. Christopoulos).

0923-5965/02/$ - see front matter
PII: S0923-5965(01)00026-1

encoding and decoding JPEG 2000 images with an ROI. Section 3 describes how to determine which wavelet coefficients need to be scaled in order to obtain the desired ROI. In Section 4 some experimental results are presented and in Section 5 the conclusions of this paper are presented.

2. The Maxshift method

When an image is encoded using JPEG 2000, a wavelet transformation of the image is performed followed by quantization of the obtained wavelet coefficients. The quantized wavelet coefficients are then divided into coding-blocks that are entropy encoded independently. The entropy encoder generates embedded bit-streams for each coding-block where the quantized wavelet coefficients have been encoded bitplane by bitplane (see [2]). A rate allocation algorithm can be used to interleave the bit-streams from the coding blocks forming a layer progressive code-stream, where the encoded bits are put in the codestream in order of significance [6,5].

2.1. Encoding an image with a region of interest

When an image is to be encoded with a ROI, the wavelet transformation and the quantization are first performed. Before the entropy encoding, the wavelet coefficients corresponding to the desired ROI are scaled (see [1,4]) so that all the bits pertaining to these coefficients are put in higher bitplanes than the significant bits of the rest of the coefficients (see Figs. 1 and 2).

When the wavelet coefficients are entropy encoded, all the bits of the ROI coefficients will be put earlier in the bit-streams of the coding blocks, than the bits of the background coefficients. When the bit-streams from the coding-blocks are interleaved into a layer-progressive code-stream, the information pertaining to the ROI will come before the information pertaining to the background. How to determine which coefficients to scale is described in Section 3. The process described above is performed independently for each tile and component in the image and the scaling values used for each tile and component are included in the JPEG 2000 code-stream. Note that no shape information for the ROI need to be included in the codestream.

2.1.1. Choosing the scaling value

The only restriction on the scaling value to use is that it must be chosen so that it is greater than the most significant bit in any background coefficient of that particular tile and component. The scaling value can either be found through a complete

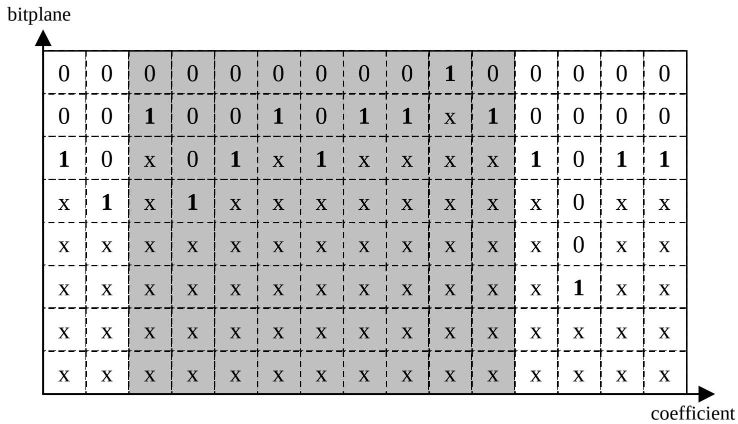

Fig. 1. One row of quantized wavelet coefficients. A '1' for each coefficient indicates the most significant bit and the subsequent bits are marked as 'x'. The gray color indicates the coefficients associated with the ROI and the white color indicates coefficients associated with the background.

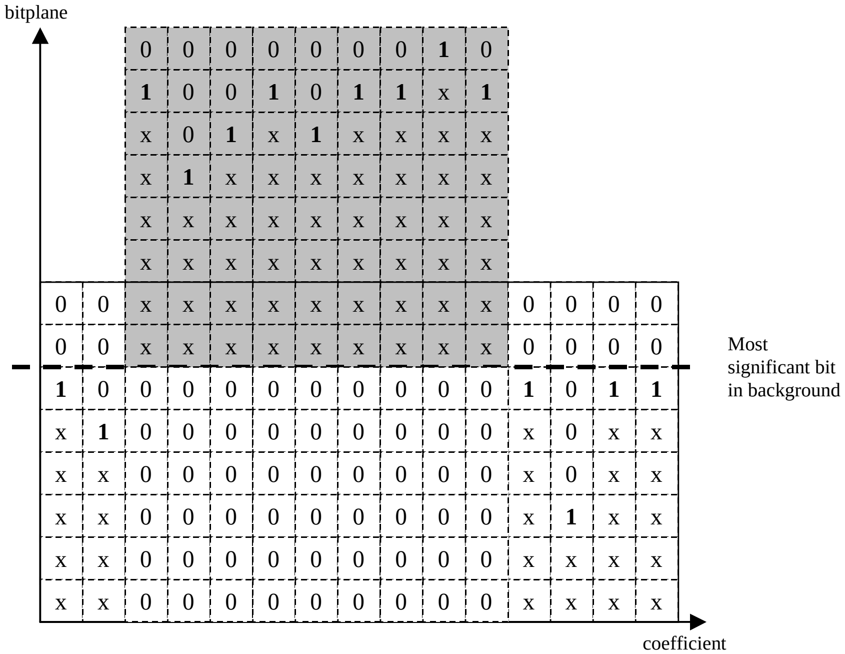

Fig. 2. One row of quantized wavelet coefficients after scaling. The coefficients associated with the ROI have been shifted up so that all the bits of those coefficients are in higher bitplanes than the most significant bit in the background. The new least significant bitplanes of the scaled coefficients are set to '0' and are encoded with the background coefficients.

search of the background coefficients as described above. Another way to choose the scaling value is to find an upper bound for the magnitude of the background coefficients. This can be done using the nominal bit-depth of the component, the analytical gain of the wavelet filters used and the number of wavelet decompositions (see [2]).

2.1.2. Using the rate allocator

Since the entropy encoder only generates embedded bit-streams for each coding-block, a rate allocation algorithm must be used to generate a layer progressive code-stream where the information pertaining to the ROI precedes that of the background. For a description of how to generate such a codestream see [6,5]. In some cases, entire coding blocks will belong either to the ROI or to the background. In these cases no scaling need to be performed to separate the ROI and the background coefficients within the coding block. Instead, the rate allocation algorithm can be used to put the entropy-encoded bitplanes in the codestream at the same level of significance as if all the coefficients in the coding-block had been scaled.

2.1.3. Encoding images with different quality for ROI and background

Using the Maxshift method, an image can be encoded with an arbitrarily shaped ROI so that the information pertaining to the ROI precedes that of the background. However, if all the bitplanes for the ROI and for the background are put in the code-stream, a complete decoding of the image will give the entire image at full quality. In order to encode an image with different quality for the ROI and for the background, the code-stream needs to be truncated at some point. By varying the quantization step, the quality of the ROI can be controlled and by truncating the code-stream (or by stopping the entropy encoding) the quality of the background can be sacrificed in order to increase compression efficiency.

2.2. Decoding an image with a region of interest

When an image that has been encoded with an ROI is to be decoded, the process described in Section 2.1 is reversed. However, whereas the encoder needs to determine which set of coefficients are to be scaled (see Section 3), the decoder only needs to consider the magnitude of the coefficients. Any coefficient that has a significant bit in a bitplane higher than the scaling value must have been scaled up in the encoder and is hence scaled down in the decoder.

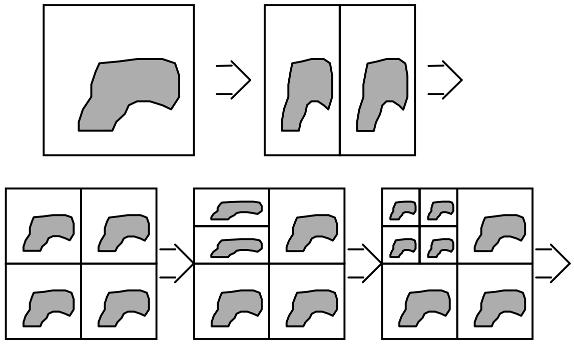

Fig. 3. The ROI mask is generated following the inverse wavelet transformation. In each step, the wavelet coefficients are found that are used to reconstruct ROI coefficients when two subbands are composed into one.

3. ROI mask generation

Since only the magnitudes of the quantized wavelet coefficients are used to scale down the ROI coefficients in the decoder, the set of coefficients that are scaled up in the encoder can be chosen completely arbitrarily. However, in some applications it is necessary that all coefficients that are used to reconstruct the ROI are scaled so that, for instance, the ROI can be encoded losslessly while the background is encoded lossy. The ROI mask is a bitplane indicating which wavelet belongs to the ROI and will be scaled up in the encoder. 0 describes how to generate an ROI mask of all coefficients that need to be encoded losslessly in order to obtain lossless results for the ROI. Section 0 describes some additional features that are possible due to the fact that the ROI Mask can be chosen arbitrarily. Some guidelines for efficient implementation of the generation of ROI masks are given in [3].

3.1. Mask generation for arbitrarily shaped ROI

In order to generate an ROI mask for lossless ROI, it is necessary to find all wavelet coefficients that are used to reconstruct the ROI. This can be done tracing the inverse wavelet transformation backwards. First a mask describing the ROI in the image domain is generated. Then the last inverse wavelet transformation, that was used to transform two subbands into the image, is traced to see which wavelet coefficients in those two subbands were used to reconstruct samples in the ROI. After that, the inverse wavelet transformations that were used to compose subbands into those two subbands are traced and so forth. Following the entire wavelet transformation this way, the complete set of wavelet coefficients, in all subbands, needed to reconstruct the ROI are found (see Fig. 3).

If, in one decomposition, one row or column of wavelet coefficients are considered as a 1D signal, the original samples are denoted $X(2n)$ and $X(2n+1)$ and the samples belonging to the low and high frequency subbands are denoted $L(n)$ and $H(n)$ respectively, then the ROI mask can be found by checking which $L(n)$ and $H(n)$ are required for the computation *of* $X(2n)$ and $X(2n+1)$. The ROI mask for the integer (5, 3) filter (which has 3 low-pass and 5 high-pass synthesis taps) is derived by looking at the inverse transform and checking what coefficients need to be in the mask. The inverse (5, 3)-wavelet transformation can be written as in Eqs. (1) and (2).

$$X(2n) = L(n) - \frac{H(n-1) + H(n)}{4}, \quad (1)$$

$$X(2n+1) = \frac{L(n) + L(n+1)}{2} + \frac{-H(n-1) + 6H(n) - H(n+1)}{8}. \quad (2)$$

The coefficients needed to reconstruct $X(2n)$ can be seen to be $L(n)$, $H(n-1)$ and $H(n)$. Hence if

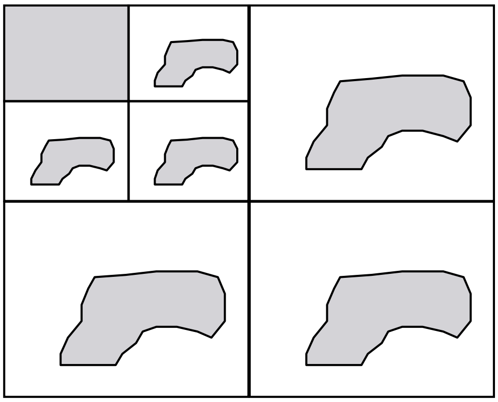

Fig. 4. ROI mask with all the wavelet coefficients in the lowest resolution level included in the mask.

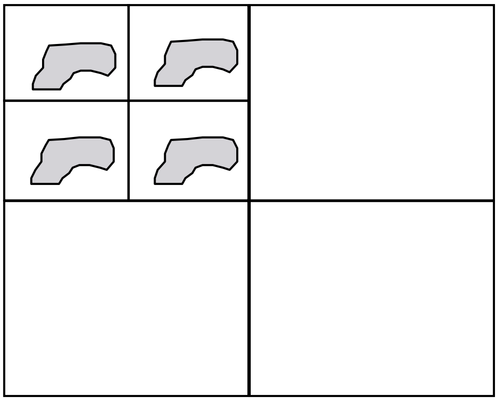

Fig. 5. ROI mask with only coefficients in the two lowest resolution levels included in the mask.

$X(2n)$ is in the mask, so should $L(n)$ and $H(n-1)$ and $H(n)$. Equivalently, if $X(2n+1)$ is in the mask, $L(n)$, $L(n+1)$, $H(n-1)$, $H(n)$ and $H(n+1)$ should be in the mask.

3.2. ROI in different resolution levels

Since it is possible to have an arbitrary set of wavelet coefficients, it is possible to have different ROI in different resolution levels. This can be used for some additional features. If, for instance, the ROI chosen is much smaller than the image, there will be no information received for the background at early stages of transmission of the image. If it is desirable to get some background information at an early stage, it is possible to include all the coefficients of the lowest resolution level in the ROI mask (see Fig. 4).

This means that a low-resolution version of the image would be transmitted together with all the information pertaining to the ROI followed by the rest of the information pertaining to the background. Any number of full resolution levels may be included in the ROI mask this way.

Another possibility is to have a ROI present only in some resolution levels (see Fig. 5). This means that the ROI would be transmitted up to a certain resolution level followed by the remaining information pertaining to the whole image. It is also possible to have different ROI that are present in different resolution levels.

4. Results

The following results are an evaluation of the Maxshift method used with seven images from the JPEG 2000 test set, covering various types of imagery. The images "bike" (2048×2560) and "cafe" (2048×2560) are natural, "cmpnd1" (512×768) and "chart" (1688×2347) are compound documents consisting of text, photographs and computer graphics, "aerial2" (2048×2048) is an aerial photograph, "target" (512×512) is a computer generated image and "us" (512×448) an ultra scan image. All these images have a bit-depth of 8 bits per pixel.

Fig. 6 shows the PSNR calculated using the average RMSE values for the test. The PSNR values were calculated for both the ROI and the entire image and a comparison was made with results for progressive decoding of non-reversible bit-streams without ROI. The region of interest in each case was a rectangular area with its upper left corner at (0.5, 0.5) times the image dimensions and the width and height of the ROI were 0.25 and 0.25 of the image dimension, respectively. A single non-reversible SNR progressive bit-stream was generated for each image and it was then decoded at 0.125, 025, 0.5, 1.0 and 2.0 bits/pixel.

The graph clearly shows that the quality of the ROI is significantly better than the quality of the

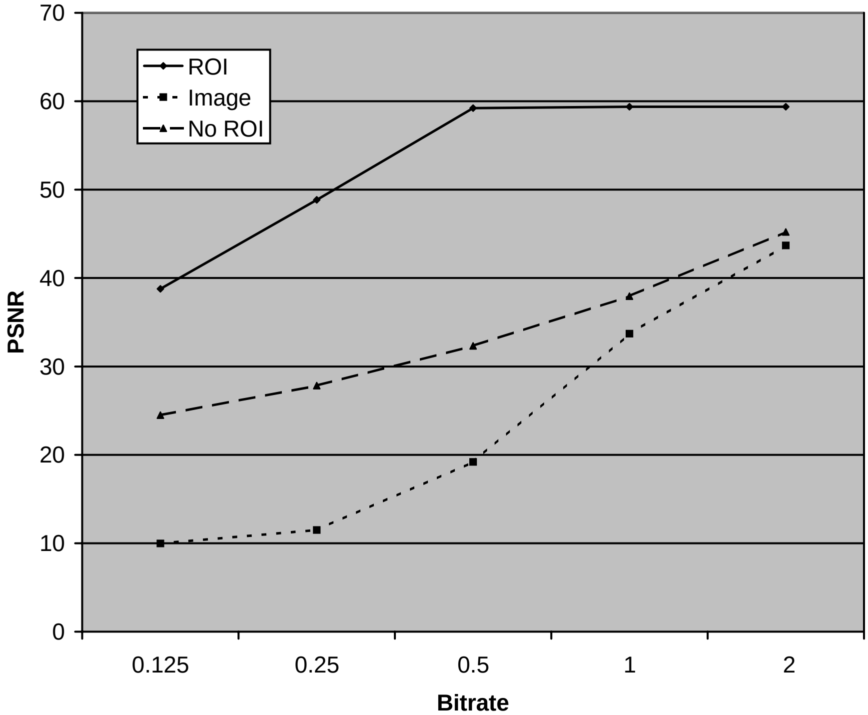

Fig. 6. Comparison of PSNR calculated from mean RMSE for the ROI and for the background for images encoded with an ROI and the PSNR calculated from mean RMSE for the same images encoded without ROI.

image encoded without an ROI, particularly at low bit-rates. The graph also shows that the quality of the ROI and of the rest of the image converges and they will both eventually reach the same quality when the full bit-stream has been decoded. For the images tested here, this occurs at bitrates higher than 2 bpp. The average increase in file-size for these images when encoding with the ROI was 5%.

5. Conclusions

The JPEG 2000 standard provides the functionality to encode images with arbitrarily shaped ROI. When a JPEG 2000 image encoded with an ROI is transmitted, the information pertaining to the ROI is sent before the information pertaining to the rest of the image. It is also possible to encode an image with different quality for the ROI and for the background. The method used is called the Maxshift method. Due to the fact that the decoder does not need to be sent in the shape of the ROI, an arbitrary set of wavelet coefficients can be included in the ROI. This makes it possible to have different ROI in different resolution levels. It is also possible to include entire resolution levels in the ROI. The results presented in this paper show that when transmitting a JPEG 2000 image encoded with an ROI, progressively, the quality of the ROI improves much faster than if the image was encoded without an ROI. The increase in file-size for the image due to the ROI coding is approximately a few percent.

References

[1] E. Atsumi, N. Farvardin, Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees, in: Proceedings of the IEEE International Conference on Image Processing (ICIP-98), 4–7 October 1998, Chicago, Illinois, USA, pp. 87–91.

[2] C. Christopoulos, A.N. Skodras, T. Ebrahimi, The JPEG 2000 still image coding system: an overview, IEEE Trans. Consumer Electron. 2000 46 (4) (November 2000) 1103–1127.

[3] C. Christopoulos, J. Askelof, M. Larsson, Efficient methods for encoding regions of interest in the upcoming JPEG 2000 still image coding standard, IEEE Signal Process. Lett. 7 (9) (September 2000) 247–249.

[4] D. Nister, C. Christopoulos, Lossless region of interest with a naturally progressive still image coding algorithm, in: Proceedings of the IEEE International Conference on Image Processing (ICIP 98), 4–7 October 1998, Chicago, Illinois, pp. 856–860.

[5] D. Taubman, High performance scalable image compression with EBCOT, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), 24–28 October 1999, Kobe, Japan.

[6] D. Taubman, A. Zalkor, Multirate 3-D subband coding of video, IEEE Trans. Image Process. 3 (September 1994). 572–578.