FLEXIBLE, DYNAMIC AND COMPLIANT REGION OF INTEREST CODING IN JPEG2000

René Rosenbaum, Heidrun Schumann

University of Rostock
Computer Science Department
D–18051 Rostock
Germany
{rrosen, schumann}@informatik.uni-rostock.de

ABSTRACT

Image transmission with JPEG2000 can be speeded up by using regions of interest (RoIs). By such a mechanism, parts of an image not belonging to RoIs can be transmitted later or encoded at a lower bit rate. This paper describes a flexible dynamic RoI-scheme which supports the definition and handling of arbitrarily-shaped RoIs in JPEG2000. The scheme supports the dynamic definition and prioritization of new and existing RoIs at any time during image transmission. This can be done without having to decode and re-encode the JPEG2000 datastream. To reach this, we use the flexible Precinct/Layer-mechanism of the standard. The proposed scheme is fully compliant with the JPEG2000-Baseline-Codec.

1. INTRODUCTION

Wireless networks like GSM allow ubiquitous access to image content. Since the bandwidth offered by such networks is very limited, flexible image coding techniques, eg. JPEG 2000, have to be used which support the encoding of RoIs. By such a scheme, image parts not belonging to RoIs can be transmitted later or encoded using lower levels of detail (LoD).

The JPEG2000-standard describes two different techniques, which refine the RoI by quality. Part1 of the standard describes the Maxshift-method[3]. The encoding of the RoI is reached by a constant up-shifting of all RoI-coefficients in the wavelet domain. Thus, these coefficients are handled prior during the next steps of the codec. This method supports every RoI-shape, which doesn't need to be signaled to the decoder. However, the rising of bitplanes during the scaling procedure decreases the compression performance. Furthermore, the entire RoI has to be decoded before the background.

The second part of the standard introduces a more common form of the Maxshift-method. By using different scaling values, it is possible to prioritize RoIs, but in so doing the shape of the RoIs has to be signaled explicitly. Like the Maxshift-method, a possible loss of image information can appear.

A combination of the these techniques is presented in [4]. This technique reduces the encoded bitplanes by decreasing the quality of the image information.

Nevertheless, all of these techniques have the drawback of a static RoI-encoding. All RoIs must be defined during the encoding of the image. This is sufficient for a lot of applications, but in interactive environments a dynamic RoI-handling is needed. In [5] an approach for encoding dynamic RoIs is presented. Depending on the RoI-definition, this method fully transcodes the last steps of the codec. However, this approach needs a lot of computing power and creates a non-compliant datastream.

In this paper, we address these problems by proposing a new approach. The flexible Precinct/Layer-mechanism of JPEG2000 is used to realize a dynamic RoI-scheme which supports the definition and handling of arbitrarily-shaped RoIs, the dynamic definition and prioritization of new and existing RoIs and the creation of a compliant datastream. The paper is structured as follows: After discussing the basic idea of the method, we will describe the algorithm for the creation of a demand-driven RoI-enhanced JPEG2000 datastream and a framework which uses the new approach to transcode dynamic RoIs. Finally, we conclude with directions for further work.

2. THE MAIN IDEA OF THE METHOD

The basic principle of our approach is the prioritized handling of JPEG2000-packets belonging to RoIs. To reach the aspired RoI-functionality, our proposal makes use of:

- the flexible layer mechanism, to build a compliant datastream,
- the independence of JPEG2000-packets, to decode image information of RoI earlier and

This work was supported by SIEMENS AG CT IC 2 Munich.

- the pyramid-encoding property, to enable spatial access to image information belonging only to RoIs.

To allow an unique application of the approach, we use a flexible RoI/LoD-scheme. All used features are described in the manner to understand their embedding and functionality within JPEG2000.

2.1. Precincts, Layers and Packets

JPEG2000 uses a extensive mechanism to format the encoded coefficients to the final datastream. In the last coding step, all grouped and encoded coefficients of a subband covered by one cell of a defined regular grid are combined to a *precinct*. The cell size of the grid can be chosen independently for every resolution level. For every resolution level precinct-triples (HL, LH, HH) are created by concatenating precincts belonging to the same spatial region. Thus, the structure represents the complete encoded information of a spatial region at a particular resolution level. To enable the SNR-progressive refinement, it is necessary to spread this information using a number of *layers*. Every layer contains a certain amount of data from the considered precincts. This partial data from a precinct-triple is formatted and called a *packet*. If there is no information to include, an empty packet is created. The whole procedure is done on every resolution level, precinct and layer. The resulting packets are formatted to the final datastream. A single packet can be decoded independently from another, but it doesn't carry any tag to derive its position. This information is derived from a fixed packet order and the position of the packet in the final stream.

A JPEG2000-compliant decoder assumes this encoding scheme for every stream to decode. As long as we change only variables of the scheme we will be compliant.

2.2. Spatial access to RoI

The pyramid-encoding property of a common wavelet decomposition scheme (e.g. EZW, SPIHT) is well known. All coefficients of a pyramid belong to the same spatial area (e.g. a RoI) of an encoded image. To support RoIs, these coefficients can be handled and decoded first. In doing so, it is possible to exclusively enhance the RoI-area in resolution and quality as described in [1, 2].

Due to the quite different architecture of JPEG2000, to access every coefficient independently is very costly. To overcome this, we use a block-based approach, which uses precincts. Thereby a regular access grid is defined in the image domain. To be JPEG2000-compliant, the cell size in each direction has to be power a of 2 and influences the fineness of the spatial access to RoIs. This nominal cell size is propagated to the next lower resolution level by dividing the size by 2. This is done for every level, which ensures all levels have the same number of precincts and the pyramid-structures are build up suitably (Fig.1). Every pyramid and pyramid-level can be decoded independently and is accessible via its belonging packets. The resulting stream is stored once and is considered further on as the original JPEG2000-datastream.

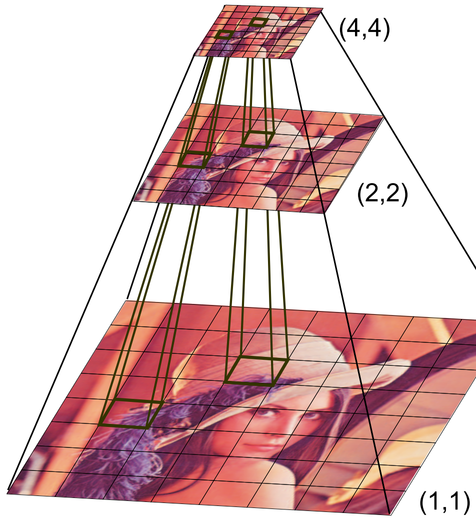


Fig. 1. Pyramid-encoding in JPEG2000 using precincts.

2.3. RoIs and LoDs

In our approach, each RoI is represented as a polygonal area in the spatial domain. To every RoI we assign a LoD, which will be refined depending on user demands. At any time the user can specify which layer (quality) of which resolution level (resolution) of which RoI shall be handled now. Thus, prioritizing a RoI can be easily done by exclusively considering image information belonging to this specific RoI.

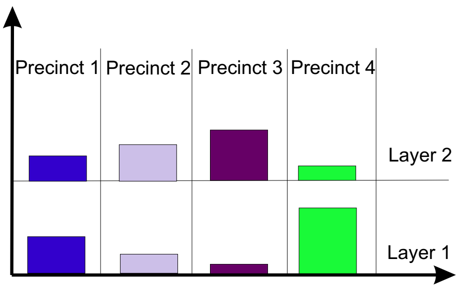
In case we have multiple and overlapping RoIs, the affected image information is assigned to the RoI which has the highest priority.

3. DESCRIPTION OF THE ALGORITHM

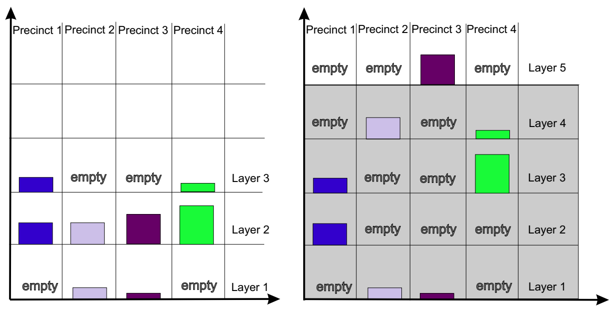
As a starting point of our approach, we need an original JPEG2000-datastream created in the described manner. Additionally, we have some user demands, e.g the definitions of RoIs, to be applied to the datastream. As a result we want a demand-enhanced JPEG2000-datastream, which considers these requirement.

The algorithm is embedded into a framework, which computes the precincts to be handled depending on actual demands and selects the belonging packets in the datastream.

Selected packets have to be arranged prior to currently remaining packets. Because of the missing position tag of a packet, it isn't possible to rearrange these packets without consideration of the whole datastream. To reach this, we use the property that a packet can be empty together with the flexible layer functionality. We call it *dynamic layer insertion*. The method is simple and handled using a transcoder. If only some packets of the actual layer have to be handled, a new layer is created, and all packets from non-important precincts are up-shifted by one. Thus, the actual layer contains only image information from the preferred precincts. Afterwards, the resulting non-staffed positions are filled with empty packets to retain the position information. Then the actual layer can be completely moved to the enhanced datastream.



(a) Original datastream.



(b) First and last transcoding step.

Fig. 2. Dynamic layer Insertion in JPEG2000 using transcoding (grey: layers of the enhanced datastream).

Fig. 2 shows an example of the procedure. The original JPEG2000 datastream contains 2 layers, 4 precincts and only one resolution level (Fig. 2a.). The packets are shown as boxes. We assume the creation of a new layer as one discrete step. In the example, image information from precinct 2 and 3 have to be placed first in the enhanced datastream. The transcoder creates a new last layer and moves all other packets to the next upper layer (Fig. 2b.). Free positions are filled with empty packets. Now the modification of the first layer is finished and the first part of the enhanced datastream created. This partial stream is JPEG2000-compliant and can be decoded without any knowledge of the following layers. Changes in demands are shown in the next image. After the finalization of the first layer, the example assumes a new demand, which requires information from precinct 1 only. The system creates a next layer and handles the affected packets as described. Depending on following demands, this procedure is done layer by layers and ends if all packets of the original datastream were handled. Every completed layer is appended at the layer former created. Because the main header of the enhanced datastream was created first and it contains the number of layers of the whole datastream, we have to adjust this value with every appended layer.

The proposed method offers a lot of advantages. During the creation of a layer all current demands, like new RoI-definitions and/or changes in prioritization of the RoIs, can be considered. This allows a dynamic and flexible RoI-handling. Because we only rearrange the datastream, this approach is very fast, needs only less computing power and produces a non-redundant JPEG2000-compliant datastream. Furthermore, the method doesn't change the encoded image information. If the original image was encoded losslessly, the enhanced datastream does not change this property.

The described approach uses precincts to access the image information of RoIs. Because a precinct acts as a box for lots of wavelet-coefficients, the spatial access can't be realized as granular as using coefficient-based approaches. Thus, we may transmit more data than necessary to display a RoI. This can be improved by downsizing the access grid cells during the encoding procedure of the original image. Hence, this leads to a worse compression performance. As several tests have shown, a cell size of 32x32 offers a good trade-off between compression and spatial access. By using empty packets to retain the right assignment of packets, the enhanced datastream becomes bigger than the original one. In our tests we measured a strong variance in the increment, because it depends on a number of values like image size, number of resolution levels and user demands. A typical interaction scenario with three successively defined RoIs and changes in the prioritization increases the original stream by $5 - 15\%$. To reduce the number of empty packets, we use the opportunity of JPEG2000 to split up the image in smaller independently handled regions, called tiles. Thus, empty packets need to be included in affected tiles only.

The proposed method can be suitably applied to every JPEG2000-datastream including more than one precinct per resolution level. Nevertheless, by using the described pyramid-encoding, the image information covered by the several pyramid-precincts is restricted to RoI-areas only.

4. A FRAMEWORK FOR TRANSCODING DYNAMIC ROIS

The proposed method was included in a system for demand-driven image transmission [1, 2]. To support JPEG2000,

Fig. 3. Dynamic Regions of Interest with JPEG2000 using the proposed method (image after 9 (left), 12 (center) and 15 (right) seconds using a GSM connection).

a Verification Model8.6-library of the JPEG2000-baseline-codec is used. The dynamic layer insertion is implemented as an additional transcoder at the server side. Every finalized layer is transmitted directly to the client. To provide RoI-functionality to the viewer, a web browser plugin has been developed. The client allows the easy specification of RoIs and the decoding of the received partial and compliant datastream during a running transmission. This is demonstrated in an example, where the initial definition of a central elliptic RoI is assumed (Fig.3/left). The image information of this region will be exclusively refined until all information was transmitted or the viewer changes his demands. In our example, the viewer defines a new rectangular RoI (Fig.3/center) to see the whole lighthouse and the surrounding areas. Hence, the priority of the first RoI is automatically decreased. The new RoI partially overlaps the former RoI. By means of a redundancy-free transmission, the overlap is handled automatically by the existing system. Lastly the viewer wants to see the whole image and defines a RoI containing all parts of the image (Fig.3/right). Since we have already transmitted a lot of image information and only differential data is transmitted, refinement of the whole image can be performed very fast.

5. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a new method for the incorporation of regions of interest into JPEG2000. Unlike many existing techniques, the proposed scheme allows the *dynamic* and *flexible* definition of regular *polygonal RoIs*. This is reached by rearranging an original datastream and using flexible layer insertion. This ensures a redundancy-free refinement of already transmitted image information and a fast and easy creation of a fully JPEG2000-compliant datastream.

In some cases it isn't suitable to transmit the information from 3 different subbands within one JPEG2000-packet. To overcome this limitation of JPEG2000, in further work we will enhance our approach to support the transmission of information from selected subbands only. This leads to a more adapted RoI-handling and saves bandwidth during the transmission of such a RoI.

6. REFERENCES

[1] U. Rauschenbach and H. Schumann, "Demand-driven image transmission with levels of detail and regions of interest," *Computers and Graphics*, vol. 23, no. 6, pp. 857–866, December 1999.

[2] Uwe Rauschenbach, René Rosenbaum, and Heidrun Schumann, "A flexible polygon representation of multiple overlapping regions of interest for wavelet-based image coding," in *Proc. IEEE ICIP2001*, Thessaloniki, Greece, October 7-10 2001.

[3] "JPEG 2000 image coding system, part1," Final Draft International Standard part 1 with corrigendum 1 and draft corrigendum 2, ISO/IEC JTC 1/SC 29/WG 1 (ITU-T SG8), 5 July 2001.

[4] Raphael Grosbois, Diego Santa Cruz, and Touradj Ebrahimi, "New approach to jpeg2000 compliant region of interest coding," *SPIE 46th anual meeting, Applications of Digital Image Processing*, July 2001.

[5] Diego Santa Cruz, Touradj Ebrahimi, Mathias Larson, Joel Askelöf, and Charilaos Christopoulos, "Region of interest coding in jpeg2000 for interactive client/server applications," *IEEE Third workshop on multimedia signal processing*, pp. 389–394, September 1999.