

# H.264/SVC ROI ENCODING WITH SPATIAL SCALABILITY

Lino Ferreira

*Instituto de Telecomunicações, Instituto Politécnico de Leiria/ESTG, Leiria, Portugal*  
*lino@estg.ipleiria.pt*

Luís Cruz

*Instituto de Telecomunicações, Universidade de Coimbra / DEEC, Coimbra, Portugal*  
*lcruz@deec.uc.pt*

Pedro Assunção

*Instituto de Telecomunicações, Instituto Politécnico de Leiria / ESTG, Leiria, Portugal*  
*assuncao@estg.ipleiria.pt*

Keywords: ROI, Spatial scalability.

Abstract: This paper proposes two H.264/AVC compliant methods for encoding Regions-of-Interest (ROI) with spatial scalability and evaluates their respective rate-distortion-complexity performance. The base layer is kept unchanged and provides lower resolution images with roughly constant quality, without identification of the ROI. In the proposed methods there is no need to encode contour information because the ROI is implicitly defined in the upper layer of the spatial resolution in a transparent way by using different encoding parameters for the ROI and its complementary region. It is shown, that spatial scalability in ROI can be efficiently used to enhance specific regions of an image sequence in both spatial resolution and quality with low coding complexity. The proposed encoding scheme is suitable for remote surveillance, medical applications and entertainment, where higher resolution and higher quality ROI is a useful functionality for object/face recognition, selective encryption, detail analysis, etc.

## 1 INTRODUCTION

The region-of-interest (ROI) functionality in visual information representation and transmission systems defines a set of methods and tools which allow selection, extraction and specific processing of important regions within the image acquisition/display area. Support for region specific differentiated coding has long been sought as a desirable feature for both image and video compression algorithms, as evidenced by any overview of the existing literature on this topic. In image coding several solutions have been proposed, taking different forms depending on the underlying coding principles, e.g. whether the coding algorithm is based either on block transforms or wavelets. In video transmission and storage ROI-based coding has also been studied in the recent video coding standards (e.g., H.264/AVC) which already offer some limited support for this functionality through the use of Flexible Macroblock Ordering (FMO) into different slice groups (Thang, 2005), (Bae, 2006), (Lambert, 2006) and (Van Leuven, 2006). In recent years, along with the

development efforts of the scalable extension of H.264, H.264/SVC, availability of scalable ROI coding functionalities was identified as an important requirement to be fulfilled in future standards (ISO/IEC TC1/SC29 WG11, 2005).

The H.264/SVC standard supports scalability in terms of spatial and temporal resolution as well as the variation of reconstruction quality (SNR). This type of encoding is more flexible and adjustable for different communication technologies and user requirements (bandwidth, resolution, etc.). In scalable ROI a single frame can be split into several independent regions which in turn may be encoded at different SNR, spatial and temporal scalabilities. In general, spatial or temporal qualities can be assigned to ROI in order to guarantee a predefined quality level while the background region can be encoded at lower quality.

This paper deals with spatially scalable ROI coding where the aim is to achieve efficient encoding of ROI with both better quality than the background area and higher spatial resolution than the base layer. The base layer is intended to provide a low resolution signal with

an acceptable spatial quality in the whole image while in the enhancement layer the ROI is the only useful image area. Therefore spatial and quality scalability is only achieved for the ROI, which should contain the image area of interest for target applications. In the following sections, the Rate-Distortion and Complexity performance of two methods, compliant with H.264/SVC, is evaluated and compared with straightforward encoding without ROI.

## 2 H.264/SVC ROI WITH SPATIAL SCALABILITY

The underlying idea to achieve efficient encoding of the ROI in the higher resolution layer is to minimise the number of bits spent in the background region of the higher resolution images. In the base layer there is no distinction between ROI and background. One of the methods proposed in this work is based on coarse quantisation of the background region and finer quantisation of the ROI in the high resolution layer. In this method, the macroblocks (MBs) of the background region, i.e., outside the ROI, are encoded with the maximum quantisation scale allowed by H.264/SVC ( $Q_p=51$ ) in order to maximise the number of null coefficients. The other method is based on setting to zero the transform coefficients of the MBs outside the ROI regardless their value. Note that in this case quantisation is avoided for these MBs. In both methods, the ROI is defined by a mask, providing a ROI map (ROImap) which is used by the encoder to identify the ROI MBs though it is not encoded into the video stream.

### 2.1 $Q_p_{51}$ outside ROI

The functional implementation of this method is depicted in Fig. 1. In each MB of the high resolution layer, the QP value is switched between 51 and the QP value selected for the current MB, either for MBs located outside the ROI or within the ROI, respectively. The ROI is not defined in the base layer, thus the whole image is normally encoded at a lower resolution.

Therefore, the quality of ROI MBs is much higher than that of the MBs outside the ROI and consequently most of the bits used in the high resolution layer are assigned to the ROI. Note that in the high resolution layer the only useful information that needs to be encoded is the ROI itself, because the lower quality and resolution of the background region provided by the base layer should be enough for the envisaged application.

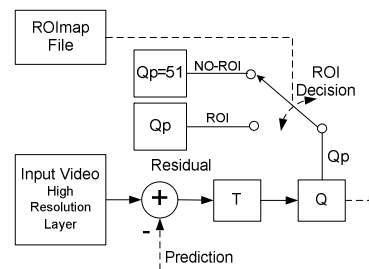


Fig. 1 -  $Q_p_{51}$  functional diagram.

### 2.2 Set-to-Zero

The objective of this method is the same as the previous one: to spend no bits in the MBs outside the ROI and to increase the subjective quality of ROI in the higher resolution layer. In the Set-to-Zero method, the transform coefficients of residual blocks are set to zero for those MB outside the ROI. Thus, the encoder sets the syntax element *coded block pattern* (CBP) to 0. The Fig. 2 shows Set-to-Zero functional diagram.

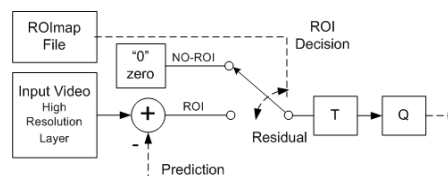


Fig. 2 - Set-to-Zero diagram.

## 3 SIMULATION RESULTS

The performance of the two methods described in the previous section was evaluated in regard to rate-distortion and encoding complexity. Separate experiments were carried out for *Intra* and *Inter* coding modes. The proposed methods were implemented using the JVT reference software, version 8.9, as a basis framework. The test sequence “Mobile” was used in the experiments with two layers QCIF@30fps (base layer), CIF@30fps (enhancement layer) and two ROIs (ROI1, ROI2) with different sizes were used. ROI1 is a 192x144 pel image region covering the area of the calendar numbers and ROI2 is the whole calendar, as shown in Fig. 3.

In the experiments the following settings were used for the *Intra* test: two spatial layers (QCIF and CIF) at 30fps; *NumberReferenceFrames* 1; *FastSearch*; *Loop Filter on*. The coding parameters were as follow: for the base layer: *CABAC*; *Basic QP* 35; *FRExt no*; for layer 1: *CABAC*; *InterLayerPred on*; *FRExt on*. The *Inter* tests the were used: two spatial layers (QCIF and CIF); 30 frames; *NumberReferenceFrames* 1; *FastSearch*; *Loop Filter on*; *MaxDelay* 1200; *GOPsize*

16; *IntraPeriod* 16. The configurations of base layer and the layer 1 are equal to *Intra* test.



Fig. 3 - ROI1 and ROI2 definition.

The simulations were performed on PC with a 2.4GHz processor and 1.0 GB of RAM memory. The rate-distortion performance of both methods was evaluated as well as the computational complexity measured as the processing time per frame.

The bitrate shown in Fig. 4 and Fig. 5 is the sum of both bitrates of base layer and layer1. The various bitrates were obtained by using different QP in layer1 while the QP of base layer is constant (QP=35). The ROI PSNR (i.e. the PSNR computed for the pixels within the ROI) is shown in the figures for *Intra* and *Inter* coding, respectively. For reference, the two proposed methods' results are compared with results from an experiment where the higher layer is totally encoded using the same QP without distinguishing the ROI and the background. These "ground-truth" results are labelled *SVC-without\_ROI*.

### A. Intra coding

The rate-distortion performance of the *Intra* case is shown in Fig. 4. The Set-to-Zero method is compared with  $Qp_{51}$  and with *SVC-without\_ROI*. The encoding complexity is shown in TABLE 1 for both ROIs. From the figures it is clear that the efficiency of the Set-to-Zero method is consistently better for both ROIs in the *Intra* case. In ROI1 this method produces a PSNR about 2dB higher than the  $Qp_{51}$  method. As one can see in the figures, the overall quality gain of the proposed methods is much higher when compared to *SVC-without\_ROI*.

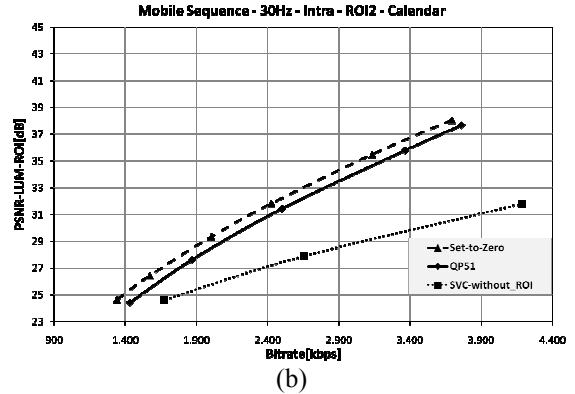
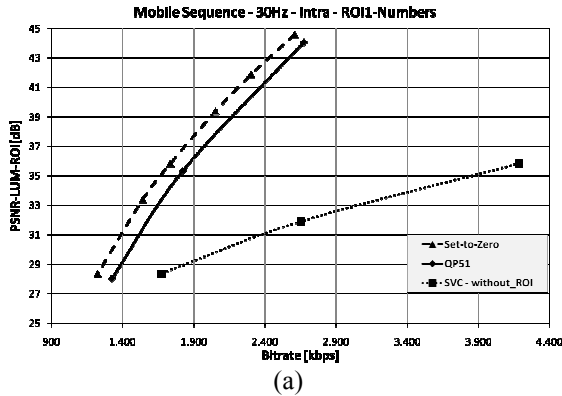


Fig. 4 - *Intra* case: Rate-Distortion (a) ROI1-Numbers (b) ROI2-Calendar.

For the lower bitrates in ROI1, the Set-to-Zero method produces a PSNR about 6.5dB higher than *SVC-without\_ROI* and at higher bitrates the gain is about 13dB. For the ROI2 the gains of Set-to-Zero are smaller than in the case of ROI1. About 0.4dB-0.5dB higher than  $Qp_{51}$  and 2.5dB-7.5dB higher than *SVC-without\_ROI* for low and high bitrates, respectively. For the same PSNR, both the  $Qp_{51}$  method and *SVC-without\_ROI* produce more bits than Set-to-Zero for encoding ROI1 and ROI2.

TABLE 1  
Processing time of encoding: (a) ROI1 (b) ROI2

QP	Set-to-Zero [ms/frame]	$Qp_{51}$ [ms/frame]	SVC-without_ROI [ms/frame]
25	182,53	195,83	262,26
35	174,37	187,35	225,24
45	167,54	179,58	192,12

(a)

QP	Set-to-Zero [ms/frame]	$Qp_{51}$ [ms/frame]	SVC-without_ROI [ms/frame]
25	204,97	217,07	262,26
35	189,10	198,85	225,24
45	174,37	182,66	192,12

(b)

Table I shows the processing time of the two proposed methods as well as *SVC-without\_ROI*. From this table one can conclude that the coding complexity of the Set-to-Zero method is smaller than that of the other two ( $Qp_{51}$  and *SVC-without\_ROI*) for both ROIs. For ROI1, the processing time is reduced by 12% to 30% with Set-to-Zero compared to *SVC-without\_ROI* and by 7% compared to the  $Qp_{51}$  method. In the case of ROI2, the processing time of Set-to-Zero is reduced 9% to 22% compared to *SVC-without\_ROI* and 5% compared to  $Qp_{51}$ . The lower complexity achieved by the Set-to-Zero is mainly due to the fact that quantisation is not

computed for the Mbs outside the ROI which significantly reduces the number of computations.

### B. Inter coding

The performance of *Inter* coding is shown in Fig. 5. In this case, the efficiency of Set-to-Zero is closer to  $Qp_{51}$ . In ROI1 the gains of both proposed methods are practically the same for low bitrates, while for higher bitrates the  $Qp_{51}$  method produces gains of about 0.8dB and 1.2dB compared with Set-to-Zero and *SVC-without ROI*, respectively. In ROI2, the Set-to-Zero yields better results relatively to the other methods. It is about 0.4dB better than  $Qp_{51}$  and nearly 2.6dB better than *SVC-without ROI*. TABLE 2, shows that the processing time depends on either the ROI dimension, the QP and the coding methods used. In this case, the processing time is greater than in the *Intra* case. However, as in the *Intra* case, the Set-to-Zero method is better than the other methods.

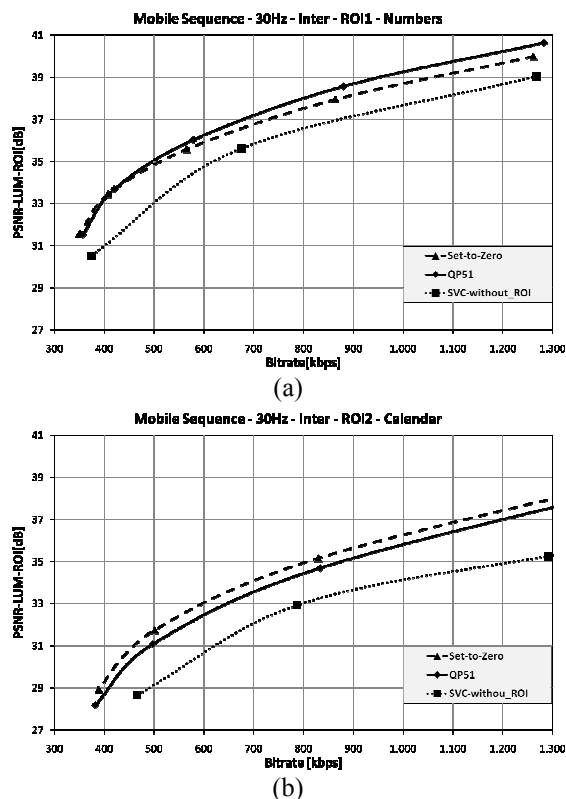


Fig. 5 - *Inter* case: Rate- Distortion (a) ROI1-Numbers (b) ROI2-Calendar.

TABLE 2  
Processing time of encoding: (a) ROI1 (b) ROI2

QP	Set-to-Zero [ms/frame]	$Qp_{51}$ [ms/frame]	SVC- without ROI [ms/frame]
25	168000,72	169000,05	170000,31
35	168000,72	169000,03	169000,41
45	168000,71	168000,87	169000,40

(a)

QP	Set-to-Zero [ms/frame]	$Qp_{51}$ [ms/frame]	SVC- without ROI [ms/frame]
25	168000,74	169000,05	170000,31
35	168000,73	169000,04	169000,41
45	168000,71	169000,03	169000,40

(b)

## 4 CONCLUSION

The performance of the ROI coding methods proposed in this paper shows that spatially scalable ROIs can be obtained at very good quality by using selective encoding for each region in the higher resolution layer. The results obtained also show that the Set-to-Zero method is less computationally complex than  $Qp_{51}$ , which makes it a good candidate for software-based implementations. By keeping the coded stream fully compatible with the H.264/SVC standard, the proposed methods are suitable for a wide range of applications where only specific regions of a video sequence are needed at higher spatial resolution (e.g., remote surveillance, medical apps, etc).

## REFERENCES

- T.C. Thang, T.M. Bae, Y.J. Jung, Y.M. Ro, J.-G. Kim, H. Choi, 04/2005 "Spatial Scalability of Multiple ROIs for Surveillance Video". In ISO/IEC MPEG & ITU-T VCEG JVT-O037.
- T. M. Bae, T. C. Thang, D. Y. Kim, Y. M. Ro, 01/2006 "Multiple ROI support in scalable video coding". In Proc. SPIE Electronic Imaging, Vol. 6074.
- P. Lambert, W. De Neve, Y. Dhondt, R. Van de Walle, 2006, "Flexible Macroblock Ordering in H.264/AVC", in Journal of Visual Communication and Image Representation, vol. 17.
- Van Leuven Sebastiaan, Van Schevensteen Kris, Dams Tim and Schelkens Peter, 07/2006 "An Implementation of Multiple Region-Of-Interest Models in H.264 /AVC", Masters Thesis, Department of Industrial Sciences and Technology, University College of Antwerpen, Belgium.
- ISO/IEC JTC1/SC29/WG11, 2005, "Applications and Requirements for Scalable Video Coding", N6880.