

Multiple Region-of-Interest Support in Scalable Video Coding

Tae Meon Bae, Truong Cong Thang, Duck Yeon Kim, Yong Man Ro, Jung Won Kang, and Jae Gon Kim

ABSTRACT—In this letter, we propose a new functionality to scalable video coding (SVC), that is, the support of multiple region of interests (ROIs) for heterogeneous display resolution. The main objective of SVC is to provide temporal, spatial, and quality scalability of an encoded bitstream. The ROI is an area that is semantically important to a particular user, especially users with heterogeneous display resolutions. Less transmission bandwidth is needed compared to when the entire region is transmitted/decoded and then sub-sampled or cropped. To support multiple ROIs in SVC, we adopt flexible macroblock ordering (FMO), a tool defined in H.264, and based on it, we propose a way to encode and, independently, decode ROIs. The proposed method is implemented on the joint scalable video model (JSVM) and its functionality verified.

Keywords—ROI, scalable video coding, MPEG.

I. Introduction

Currently, ISO/IEC MPEG and ITU-T VCEG are jointly making a scalable video coding (SVC) standard that is based on the hierarchical B frame structure (UMCTF) and the scalable extension of H.264/AVC [1]. The joint scalable video model (JSVM3.0) has been released, which describes the specific decoding process and bitstream syntax of the proposed SVC [2]. The objective of this codec is generating a temporal, spatial, and quality scalable coded stream that provides users with quality-of-service-guaranteed streaming service independent of video consuming devices in a heterogeneous network environment.

Manuscript received Dec. 02, 2005; revised Jan. 02, 2006.

Tae Meon Bae (phone: +82 42 866 6289, email: heartles@icu.ac.kr), Truong Cong Thang (email: tcthang@icu.ac.kr), Duck Yeon Kim (email: moonst55@icu.ac.kr), and Yong Man Ro (email: yro@icu.ac.kr) are with the School of Engineering, Information and Communications University, Daejeon, Korea.

Jung Won Kang (email: jungwon@etri.re.kr) and Jae Gon Kim (email: jgkim@etri.re.kr) are with Digital Broadcasting Research Division, ETRI, Daejeon, Korea

Reducing the picture resolution may not be the best solution for devices that have restrictions in size and display resolution such as handsets or PDAs. Instead, defining a semantically meaningful region such as an ROI, and displaying it, could be better because it provides important information while not reducing resolution. For this reason, the support of ROI is one of the SVC requirements [3].

The MPEG-4 object-based codec and H.263 can also support ROI functionality [4], [5]. The basic concepts of MPEG-4 object-based coding and H.263 independent segment decoding (ISD) mode for ROI decoding are the same in the point of treating an ROI as a whole picture, but due to differences in the detailed encoding scheme, specific considerations differ.

In this letter, we consider the support of a scalable ROI in SVC. Currently, SVC provides an extraction scheme that produces a spatial, temporal, and quality reduced bitstream from the originally encoded one without transcoding. However, ROI-related functionality is not yet supported, thus the proposed functionality enables the extraction of an ROI from the SVC bitstream. The extracted bitstream may have more than one ROI in the picture, and each ROI can be decoded independently with spatial, temporal, and quality scalabilities.

To accomplish the objectives, we apply flexible macroblock ordering (FMO) to the JSVM in order to describe ROIs. Based on utilizing the FMO to describe an ROI, we analyze the requirements to enable the independent decoding of the ROI.

II. Problems of Multiple ROI Support in SVC

Supporting an ROI by a video codec means that it provides a way to describe and encode/decode ROIs independently from

the whole picture. In addition to this support, SVC should provide scalabilities for the ROI, which means additional functionalities for the ROI should not conflict and should be well-harmonized with already existing functionalities for scalabilities.

1. Multiple ROI Representation in SVC

In this letter, we adopt FMO to describe ROIs. FMO is a tool of H.264 that enables the grouping of macroblocks into a slice group and the decoding of the slice group, independently, in order to make it possible to decode the remaining parts of a picture when there is a loss of the slice group that composes the picture [6]. FMO provides six types of macroblock-to-slice-group maps. Among them, map type 2, named ‘foreground and leftover’, groups macroblocks located in rectangular regions into slice groups, and the macroblocks not belonging to a rectangular region are grouped into one slice group. We use map type 2 to describe ROIs in the picture.

If more than one ROI are defined in the picture, we should consider the overlapped region between ROIs. If each ROI is described as one slice group, as in Fig. 1, the overlapped region between ROI 1 and ROI 2 has to belong to the slice group that has the lower slice group id. If the slice group id of ROI 1 is 0 and that of ROI 2 is 1, the overlapped region will belong to ROI 1. Therefore, if the slice group that represents ROI 2 is decoded, the result will show ROI 2 excluding the overlapped region. To overcome the problem, a new slice group is assigned to the overlapped region, enabling it to be decoded independently. If ROI 2 is to be decoded, slice groups for ROI 2 and the overlapped region should be decoded as well. To keep the FMO rule, the slice group id of the overlapped region should be lower than those of the related ROIs.

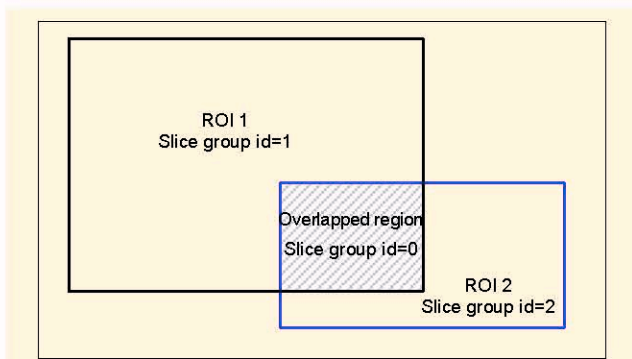


Fig. 1. Description of multiple ROIs with overlapped region by FMO.

2. Problems of Independent ROI Decoding in SVC

FMO is not enough for independent decoding of an ROI due to the characteristics of predictive coding and inter-pixel

dependent processing. To prevent decoding dependency between slice groups, FMO disables intra-prediction from the macroblocks outside of a slice group. However, it only avoids the decoding dependency that resides in the current picture; there still exists decoding dependency in the temporal direction by motion compensation. In addition, in the boundary of an ROI, half-sample interpolation for motion estimation (ME)/ motion compensation (MC) and upsampling for Intra_Base mode also cause problems due to interdependency between slice groups.

A. Constrained Motion Estimation

As mentioned before, constraining the motion search range into the ROI region is required to prevent inter-frame dependency between different slice groups. The ISD mode of H.263 also performs constrained ME, but MPEG-4 visual part 2 allows referencing samples outside of the video object plane (VOP) for an unrestricted motion vector [5].

An overlapped region may be allowed for the ME/MC of a non-overlapped region. However, the slice group for the overlapped region should be decoded before decoding the slice group for the non-overlapped region.

B. Handling Half-Sample Interpolation on the Slice Group Boundary

SVC and H.264 perform ME/MC using motion vector accuracy of one-quarter of a luminance sample grid spacing displacement. A 6-tap finite impulse response filter is used to construct the half sample, and bilinear interpolation is then applied for quarter sample construction [6]. Figure 2 shows the interpolation for the half sample.

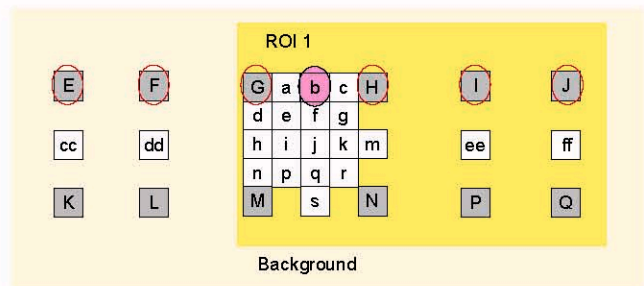


Fig. 2. Half-pel interpolation in the ROI boundary.

Equation (1) represents the luminance value of the half-sample position labeled ‘b’ by applying the 6-tap filter to the nearest integer position samples in the horizontal direction.

$$b = \text{round}((E - 5F + 20G + 20H - 5I + J) / 32) \quad (1)$$

If the interpolation for the half sample is performed near the slice group boundary, it requires integer samples outside of the

slice group. As shown in Fig. 2, the half sample labeled ‘b’ requires integer samples labeled ‘E’ and ‘F’, which are samples located outside of the slice group. Therefore, if only the ROI is decoded without background, there will be a mismatch between encoding and decoding in the half-sample interpolation, which would lead to a decoding error. To avoid this mismatch problem, there should be an agreement in referencing the integer sample outside of the slice group in half-sample interpolation. The same problem occurs in the picture boundary, and the current SVC and H.264 solve the problem by extending the picture boundary by using zeroth-order extrapolation in the horizontal and vertical directions. Therefore, these values can be replaced by the nearest integer sample in the slice group. Another method for solving this problem is to restrict the motion search range within the ROI region inside two samples [7]. However, this method decreases the coding efficiency.

Since the base layer of SVC should be compatible with H.264, a more restricted motion search should be used instead of extending the boundary of the slice group even though it decreases the coding efficiency.

Because H.263 uses a bilinear interpolator for the half sample, it does not suffer the same problem as that of SVC. Also, MPEG-4 visual part 2 allows referencing samples outside of the VOP by padding the VOP boundary using mirroring boundary samples [5].

C. Handling Upsampling of Intra_Base Mode on the Slice Group Boundary

In Intra_Base mode, inter-layer intra texture prediction is performed. By using the texture of the base layer, the encoder predicts that of the enhancement layer. When the spatial resolution of the base layer is half of that of the enhancement layer, the texture of the base layer should be upsampled. The interpolator for half-sample construction is used for the upsampling; therefore, the referencing sample outside the slice group occurs in the slice group boundary.

Because the cause of the problem is the same as that of the half-sample interpolation, the approach to handle the problem is also similar. However, the number of maximum referencing samples is three for upsampling, while being two for half-sample interpolation. Figure 3 shows a way of implementing the proposed handling by padding the slice group boundary.

For the macroblocks in inter-layer residual texture prediction mode, residual textures are reconstructed using bilinear interpolation, and there is no referencing of samples outside of the macroblock. Therefore, the error in the case of Intra_Base does not occur in inter-layer residual texture prediction.

In the case of H.263, it uses a different bilinear interpolation

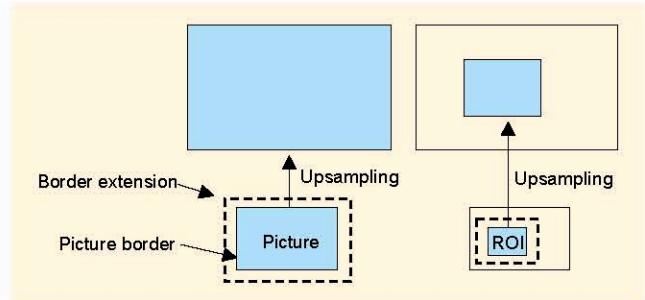


Fig. 3. Border extension of a picture and ROI for upsampling.

filter in the region (picture) boundary when ISD is used with spatial scalability.

D. Disabling the Deblocking Filtering

The deblocking filter in H.264 aims at smoothing the blocking effect. Because the deblocking filtering is inter-macroblock processing, it also causes a problem in the boundary when the ROI is decoded alone. H.264 and SVC are able to control the deblocking filter by setting the variable ‘disable_deblocking_filter_idc’. By setting this variable to ‘2’, we can disable the deblocking filter in the slice boundary.

III. Simulation and Analysis

We implemented the proposed method in JSVM, performed functional verification of ROI-independent decoding, and experimented on the effect of boundary handlings in the slice group. SVC test sequences, ‘BUS’ and ‘ICE’, were used for the experiment. For the ‘BUS’ sequence, a two layer configuration—{QCIF, 15 fps}, {CIF, 30 fps}—was used for encoding the sequence with two ROIs. For the ‘ICE’ sequence, a three layer configuration—{QCIF, 15 fps}, {CIF, 30 fps}, {4CIF, 30 fps}—was used for encoding the sequence with one ROI. Figure 4 shows the original test sequence of ‘BUS’ and ‘ICE’ sequences as well as the defined ROI regions. The sizes of the ROIs are 128×128 in CIF resolution.

Figure 5 shows the decoded result when the boundary handling for half-sample interpolation is not applied. As shown in Fig. 5, there are noticeable errors in the picture, which look like lines. These are due to referencing the half- or quarter-sampled boundary macroblock. In addition, these errors are drift with motion. Figure 6 shows the decoded result when the boundary handling for upsampling is not applied. The errors in the slice group boundaries are obvious. Figure 6(a) shows an ROI in CIF, and 6(b) shows an ROI in 4CIF. Figure 6(b) shows a more severe error than Fig. 6(a), which is due to propagation of the upsampling to the upper layer. Figure 7(a) represents the decoded result when the boundary handlings for both



Fig. 4. Original sequences and ROI region: (a) 'BUS' and (b) 'ICE'.

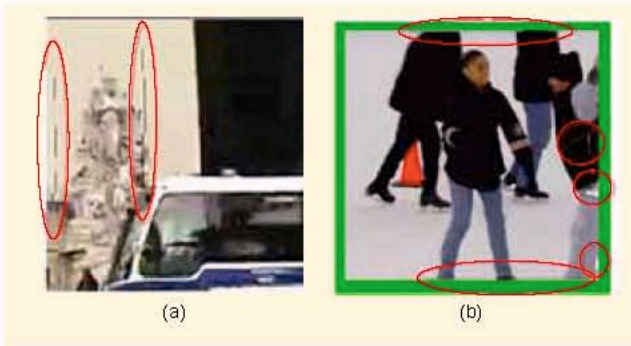


Fig. 5. Decoded results when boundary handling for half-pel interpolation is not applied: (a) 'BUS' and (b) 'ICE'.



Fig. 6. Decoded results when boundary handling for upsampling is not applied: (a) 'BUS' and (b) 'ICE'.



Fig. 7. Decoded results when boundary handlings for both half-pel interpolation and upsampling are (a) not handled and (b) handled (3 spatial layer).

upsampling and half-sample interpolation are not applied, which shows combined errors. And Fig. 7(b) shows an error-

free result due to the proposed boundary handlings of the ROI.

Due to the FMO and constrained motion estimation, coding efficiency decreases when an ROI is present. The increases of bitrate due to the presence of an ROI are 10.9% and 3.5%, respectively, for the test sequences 'BUS' and 'ICE' with the same configuration of boundary error test.

IV. Conclusion

In this letter, we suggest a way to support independently decodable multiple ROIs in SVC. The proposed method is implemented in JSVM, and the simulation result verifies ROI support of SVC. We proposed an exceptional handling of half-sample interpolation and upsampling in the slice group boundary in the Joint Video Team meeting [8].

References

- [1] ISO/IEC JTC 1/SC 29/WG 11, *Working Draft 4 of ISO/IEC 14496-10:2005/AMD3 Scalable Video Coding*, N7555, Nice, Oct. 2005.
- [2] ISO/IEC JTC 1/SC 29/WG 11, *Joint Scalable Video Model (JSVM) 4.0 Reference Encoding Algorithm Description*, N7556, Nice, Oct. 2005.
- [3] ISO/IEC JTC 1/SC 29/WG 11, *Scalable Video Coding Applications and Requirements*, N6880, Hong Kong, Jan. 2005.
- [4] ITU-T, "Video Coding for Low Bitrate Communication," *ITU-T Recommendation H.263, Ver. 2*, Jan. 1998.
- [5] ISO/IEC JTC1/SC 29/WG 11, *Information Technology—Coding of Audio-Visual Objects—Part 2: Visual*, ISO/IEC 14496-2 (MPEG-4), 1998.
- [6] ISO/IEC JTC 1/SC 29/WG 11, *Text of ISO/IEC FDIS 14496-10: Advanced Video Coding*, 3rd ed, N6540, Redmond, July 2004.
- [7] ISO/IEC JTC 1/SC 29/WG 11, *Isolated Regions: Motivation, Problems, and Solutions*, JVT-C072, Fairfax, May 2002.
- [8] ISO/IEC JTC 1/SC 29/WG 11, *Boundary Handling for ROI Scalability*, JVT-Q076, Nice, Oct. 2005.