EVOLVING ROI CODING IN H.264 SVC

Syeda Shamikha F. Shah and Eran A. Edirisinge
Loughborough University, Leicestershire LE11 3TU, UK

Keywords: SVC, ROI, Motion Estimation, Video Surveillance.

Abstract: Region-of-Interest (ROI) based coding is an integral feature of most image/video coding techniques/standards and has im-portant applications in content based video coding, storage and transmission. However, in the latest scalable extension of H.264 AVC video coding standard, i.e. H.264 SVC, motion estimation across the slice group boundaries does not preserve the coding quality and compression rate of the ROI. In this paper novel enhancements to the ROI based coding for H.264 SVC have been proposed to constrain the inter frame prediction across slice group boundaries. We show that the proposed algorithms do not negatively affect the rate-distortion performance of the coded video, but provide useful additional functionality that enables the extended use of the standard in many new application domains. Further, we pro-pose a method for supporting the coding of moving ROI in the scalable video coding domain, by adaptively changing the shape, size and position of the slice groups. We show that this additional functionality is particularly useful in video surveil-lance applications to effectively compress and transmit the ROI and reduce the storage and transmission requirements without any quality degradation of the ROI.

1 INTRODUCTION

The Scalable extension of H.264-AVC, i.e. H.264 SVC addresses the challenges of supporting heterogeneous users linked over heterogeneous networks. Each user might have different requirements and constraints. This includes different screen resolution or different QOS requirement of the application. Similarly, the condition of the network is not a constant factor owing to congestion and fluctuation of bandwidth. SVC provides the flexible encoding to cater to these changing requirements (Ziliani and Michelou, 2005). The application areas of Scalable video coding include digital video surveillance and network applications. The scalable standard should be able to discard parts of the video bit stream to meet channel requirements and provide better compression and performance efficiency (Mark et al., 2002).

ROI based coding is an important topic in video coding. A considerable amount of research has been carried out on enhancing the ROI coding as well as adapting it to the scalability domain. Some problems encountered in enabling ROI based coding, such as carrying out motion compensation and intra coding of macroblocks have been highlighted in Wang and Hannuksela, 2002. Bae et al., 2006 takes it on further and addresses the issues related to coding ROI in scalable mode. It shows how to overcome the problems posed by the dependency between slice groups (ROI) in intra-prediction, motion estimation, half-sample interpolation on the slice group boundary and upsampling in intra-base mode on the slice group boundary. It further suggests that the dependency between slice groups for motion estimation should be resolved by implementing constrained motion estimation.

The importance of limiting the inter prediction across slice group boundary has been realized by the H.264 SVC standard (Wiegand et al., 2006) by introducing the motion constrained SEI message. This message signals to the decoder that the samples from a given set of slice groups shall not refer to samples outside this set. The encoder shall provide the functionality to limit this reference and so should the decoder.

In this paper, novel techniques to restrict the motion estimation across slice group boundaries at the encoder have been proposed. These techniques do not require the transmission of the motion constrained SEI or any special handling at the decoder. Constrained inter-frame prediction across slice group boundaries is important for the

Shamikha F. Shah S. and A. Edirisinge E. (2008).
EVOLVING ROI CODING IN H.264 SVC.
In *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pages 13-19
DOI: 10.5220/0001085200130019

independent decoding of ROI and preserving its coding quality.

A further issue with ROI coding is the change in the shape, size and position of the ROI. (Wang and Hannuksela, 2002) proposes ways to code evolving ROI, that is, the shape of the isolated region grows/evolves with time. FMO map type 3, 4 and 5 provide the feature of growing and evolving slice groups. However, these map types do not cater to a moving slice group. The slice group can grow from its initial position but not change shape, or move horizontally or vertically across the frames. Therefore, these map types cannot be used for implementing moving ROI and special handling needs to be provided for changing/moving slice groups. In light of the above observations and practical significance, support for moving ROI in H.264 SVC has been proposed in this paper.

The rest of the paper is organized as follows. Section 2 describes the proposed algorithms for constrained inter frame prediction across slice group boundaries. Changing slice groups (moving ROI) feature is presented in Section 3. Section 4 provides the experimental results and their analysis. The conclusion to this research is drawn in Section 5 with suggestions for future work.

2 CONSTRAINED INTER FRAME PREDICTION

Constrained inter prediction across the slice group boundary is a useful functionality to allow for independent decoding of slice groups, and hence the ROI. The independent ROI decoding can increase the error resilience by limiting the motion search for the ROI to the same slice group in the reference pictures. It will restrict motion compensation from slice groups coded at lower quality. Restricted motion compensation, in turn, maintains the compression quality of different slice groups. A slice group that is coded at a lower compression rate would maintain its quality by not referring to the samples that are outside this slice group and are possibly coded with higher compression.

Three different techniques to constrain the inter prediction across slice groups boundaries are proposed as follows.

2.1 Boundary Padding of Non-ROI

The proposed method to restrict the inter-frame prediction across slice group boundaries is to eliminate the possibility of any sample in one slice group finding its best match from the other slice group. This can be done by padding the boundaries of the ‘non-current’ (current slice group being the one for whose samples, a best match is being found) slice group.

The size of padding should be equal in width to the minimum of search range specified in the encoder configuration and the width of ‘non-current’ slice group. The value with which this region is padded should be some value other than a permissible pixel sample value (both luma and chroma). This padding shall be applied to all reference pictures used for inter prediction, and not the current picture. The interpolation process for the reference frame, for creating half pixel accurate and quarter pixel accurate sample buffers, shall be carried out after the padding.

Figure 1 illustrates the padding process. In figure 1(b), the macroblocks from slice group B are padded with an undefined value. Although they fall inside the search range of the current macroblock, their undefined value cannot provide a match for this macroblock. Based on the implementation it is possible to restrict the motion vectors of just one slice group or multiple slice groups. The one draw back of this technique is that if the reference frames are padded only once and used for all slice groups, then the padded slice group can effect the motion estimation of its own samples.This is because the padded area would become ‘inaccessible’ to the padded slice group as well. A way to solve this can be to pad the reference frames for each slice group separately.

2.2 Limiting the Search Rectangle

Constrained inter prediction can be implemented by redefining the search range for each macroblock according to its position in the slice group. In this algorithm, the search range of the current macroblock is defined in a way that the rows and columns of macroblocks belonging to other slice groups are excluded from the search rectangle of the current macroblock. The technique is illustrated in figure 2.

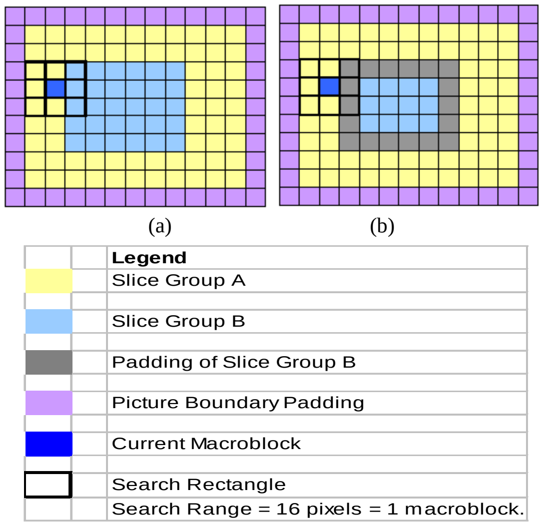


Figure 1: Padding slice group inner boundary (a) without padding (b) with padding.

The implementation requires the initialization of the macroblock for motion compensation to be changed. This technique is ideally suited to restrict inter-frame prediction for the foreground slice group in FMO map type 2.

2.3 Constrained Inter-frame Prediction at MB and Sub-MB Level

This technique involves restricting the motion vectors of each macroblock to point inside the slice group to which it belongs. This restriction has to be implemented in the form of checks at the macroblock and sub macroblock level, so that neither the 16x16 macroblock, nor any of its partitions have the motion vector pointing outside. Further, the restriction should be active for both full pel motion vectors and sub pel motion vectors.

The following algorithm is designed to check if the motion vector points to a macroblock or macroblock partition that belongs to the current slice group. If so, the motion vector is valid, otherwise this motion vector shall not be used in the motion estimation process.

- Let the Motion Vector (Mv) to be checked be (MvX, MvY).
- Let the partition size of the partition for which the motion is being estimated be Px (width) and Py (height).
- Let the coordinates of the current macroblock (MB) be (MBx, MBy).

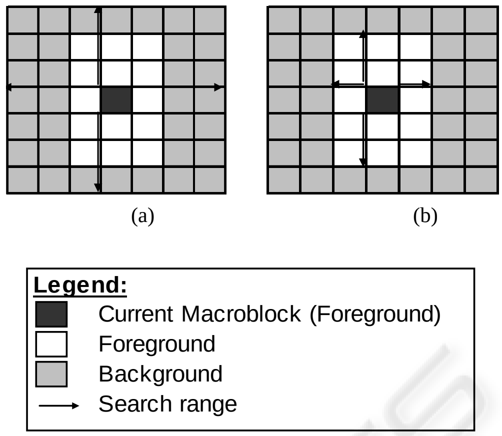


Figure 2: Limiting search rectangle (a) original search range (b) limited search range.

Step 1: The starting coordinates (X, Y) of the partition to be checked for best match, by calculating the SAD with the current macroblock partition, are derived as:

$$X = MBx + MvX$$
$$Y = MBy + MvY$$

Step 2: Determine the coordinates of the pixels that mark the four corners of the partition to be checked to give the best match.

- (X + Px, Y)
- (X, Y + Px)
- (X + Px, Y + Px)
- (X , Y)

The x and y coordinates of any of these pixels, that is calculated to be lying outside the picture boundary should be clipped to the nearest boundary coordinate.

- x = min (0, max (x, picture width in pixels))
- y = min (0, max (y, picture height in pixels))

Note that by doing so, the padded area outside the pixel boundary is mapped to a macroblock closest to the padded area but lying inside the boundary. Thus the slice group of this part of padded region would be inferred as the slice group of the closest boundary macroblock.

Step 3: The owner macroblocks of the four pixels, as given in step 2, shall be determined. If all of these owner macroblocks belong to the same slice group as the current macroblock, then the motion vector is valid, otherwise it is invalid.

Usage of the algorithm. The algorithm given above is used in the motion estimation process for restricting the motion vectors to the current slice group.

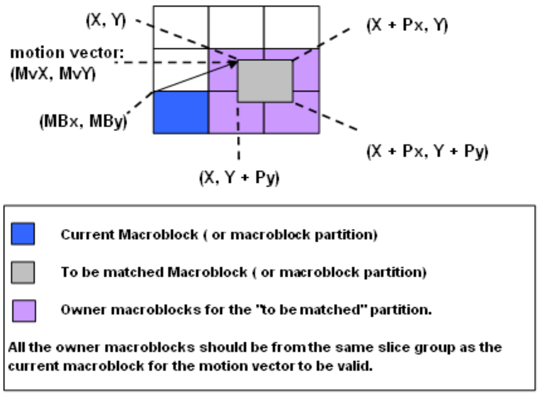


Figure 3: Determine valid motion vector.

Motion estimation process picks the predicted motion vector as the first best estimate for refining the motion vector. However, even when the predictor blocks belong to the current slice group, their motion vectors when translated to the current macroblock may point outside the slice group. Therefore the validity of the motion vector should be checked by using the proposed algorithm.

The motion estimation of the macroblock and its partitions, by zero vector and tree search (hierarchical search), shall also be restricted by applying the algorithm.

Moreover, this algorithm shall also be used in sub-pel motion estimation. The sub pixel motion estimation involves the half pel and quarter pel interpolation. As the first step towards determining if the motion vector points to a macroblock inside the current slice group, the macroblock to which the sub pel would belong should be identified. After the mapping from sub-pel to full pel, the validity of the motion vector shall be determined.

This technique is applicable to both the rectangular slice groups (FMO map type 2) and arbitrary shaped slice groups (FMO map type 6). Moreover, it inter predicts each slice group independently of the other, and is not restricted by the number of slice groups in all.

3 CHANGING SLICE GROUPS (MOVING ROI)

The FMO functionality in the SVC standard allows defining multiple slice groups in the frame. In the case of FMO type 2, the foreground slice group can be selected as the ROI. However this selection is fixed for the entire video sequence. In real life applications, the object constituting the ROI changes its position with time. This calls for updating the place and shape of the ROI from time to time.

The support for changing slice groups/ROI, as proposed in this paper, allows changing the ROI definition at the encoder. The relevant information is transmitted to the decoder in time to decode the changing slice groups. The changes are transparent to the decoder. Furthermore, it preserves the encoded quality of each slice group as ensured through constrained motion estimation techniques.

The following steps are involved in implementing changing slice groups.

3.1 Slice Group Map Redefinition per GOP

The slice group definitions for the video sequence are provided to the SVC encoder as the configuration parameters. The encoder subsequently generates the macroblock to slice group mapping for the entire video sequence, once it starts encoding the sequence. In order to redefine a slice group, the corresponding FMO parameters should be changed. The new parameters, such as the starting and ending MB for the ROI, should correspond to the updated size, position and shape of the ROI. These parameters can be obtained by repeated ROI identification per GOP, through some computer vision algorithm.

Following the parameter change, the FMO unit shall be reinitialized to construct the macroblock to slice group map according to the new definition of the slice groups. The frequency of changing the slice group can be as high as per frame. However, this increases the computation cost of the encoder. Thus it is advisable to change the slice group mapping once per group of pictures (GOP).

3.2 Reference Slice Group Map for Motion Estimation

In the reference SVC encoder (JSVM 8.13), one slice group map is used for all the frames of the video sequence. The constrained motion estimation process refers to the slice group map to find the slice group of a macroblock. This is to ensure that the best match macroblock is only picked up from the same slice group as the current macroblock.

When the moving ROI functionality is implemented, the slice group map changes every GOP. Since the key frame from the previous GOP is used in the motion estimation process for the frames of the current GOP, the need to have the current as well as the old slice group mapping is essential. For this reason, the slice group map is stored with each frame. A macroblock from the reference frame is selected as the best match only if it lies in the same slice group according to the slice group map of the reference frame.

3.3 Picture Parameter Set Update and Transmission

The FMO parameters, which include the first MB in a slice and the number of macroblocks in the slice are transmitted by the encoder in the slice header. However, the macroblock to slice group map is communicated to the decoder in the picture parameter set. Therefore, for moving ROI, the picture parameter set NAL unit is updated and transmitted by the encoder every time the slice group mapping is changed.

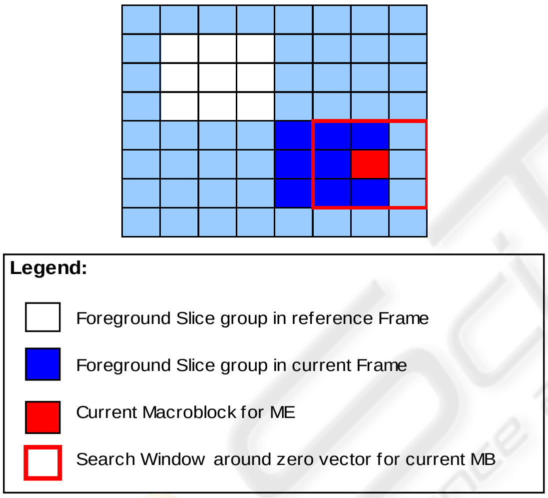


Figure 4: Non overlapping search window of current MB with foreground SG in reference frame.

3.4 Increase in Search Range

The constrained motion estimation process considers a macroblock in the reference frame as the best match if it belongs to the same slice as the current macroblock. Changing slice groups implies that the slice group definition in the reference picture may be different from that in the current picture. In the case when the motion between the GOPs is fast, it is possible that the search window for a macroblock may not overlap with the slice group mapping of its owner slice group in the reference frame. This is illustrated in figure 4.

In this case, no macroblock in the reference frame will fulfill the criteria that it belongs to the same slice group as the current macroblock. For this reason, even a macroblock with otherwise a very low SAD, will not be selected to estimate motion of the current macroblock, since this would be a compromise to the coding quality of the ROI and violate the principles of constrained motion estimation. To resolve this issue, the search range in the configuration parameters shall be increased. The increased search range would be effective for all the macroblocks in the entire video sequence.

4 EXPERIMENTAL RESULTS

The algorithms for constrained inter-frame prediction and the support for moving ROI has been implemented on JSVM reference software (version 8.13).

4.1 Constrained Inter-frame Prediction

In the experiments, Football and Foreman video sequences were coded with 2 slice groups defined using FMO map type 2. Constrained inter-frame prediction at MB and sub-MB level was tested for FMO map type 6 as well. Skip mode and Direct mode for motion estimation were not enabled for the constrained inter-frame prediction at MB and sub MB level. The loop filter was disabled for the experiments. Different QP values were set for the two slice groups. One of the slice groups is coded with QP value greater than 52. This is to make the distinction between the two slice groups visually apparent for testing purposes. The algorithms have been tested for two spatial layers (QCIF and CIF).

(a) (b)

Figure 5: Decoded frame (Foreman QCIF) (a) without constrained inter frame prediction (b) with boundary padding of non-ROI (ROI in grey).

(a) (b)

Figure 6: Decoded frame (Foreman QCIF) (a) without constrained inter-frame prediction (b) with limiting the search rectangle for ROI (ROI in grey).

(a) (b)

Figure 7: frame (Foreman QCIF) (a) without constrained inter-frame prediction (b) with constrained inter-frame prediction at MB and sub MB level.

Without constrained inter-frame prediction, the samples from one slice groups are compensating motion for the other slice group and hence no distinct boundary is seen for P or B frames. This distinction in quality is present with constrained inter-frame prediction.

4.1.1 Effect on Bitrate and PSNR

The PSNR and bitrate values were obtained on constrained inter-frame prediction at MB and sub MB level encoded with two spatial layers. It is observed that there is no significant difference in PSNR with or without the constrained ME. Therefore we conclude that the constrained ME technique does not effect the overall quality of the sequence.

Experiments were conducted for bitrate on foreman and foreground test sequences with the two slice groups coded with a base QP of 8 and 48 respectively. The experiments show a decrease in the bitrate. However, for some values of QP and QP difference between the two slice groups, the bitrate may increase. This is because the bitrate is a balance between the bits used to encode the error and the bits used to encode the motion vector. With constrained motion estimation, the magnitude of motion vector is reduced, since it is constrained to the same slice group. However, the error increases with the constrained ME, since the best match is forced to be selected from within the same slice group, which otherwise could have existed somewhere outside the slice group.

The magnitude of error as well as that of motion vector also depends on the size of the slice groups and the degree of motion between frames. Hence the effect on bitrate is controlled by all these factors.

4.1.2 Computational Complexity

The constrained inter-frame prediction techniques were implemented without any hardware accelerator. No special emphasis was given to optimized implementation of these techniques. The computation time of the Constrained Inter Frame Prediction techniques was computed using Intel® VTune™ Performance Analyzer 8.0 for windows. The computation time was calculated on both fast motion sequence (football) and slower motion sequence (foreman) with one and two spatial layers.

The technique with boundary padding of non-ROI shows an increase in computation time of roughly 23% for foreman and 7 to 10% for football. Constrained inter-frame prediction at MB and sub-MB level causes an increase of 32% for both test sequences. A decrease of about 6% in computation time is observed for constrained inter-frame prediction by limiting the search rectangle.

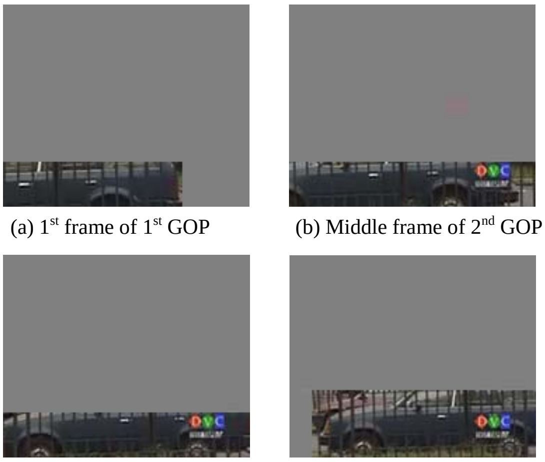
4.2 Changing Slice Groups (Moving ROI)

The support for moving ROI has been implemented on JSVM (version 8.13). Experiments were conducted on Foreman and Bus video sequence. Testing was done with and without spatial scalability. The loop filter was disabled for the experiment.

The ROI was selected using FMO map type 2, with the foreground slice group compressed with a lower QP then the background slice group. The basic QP for foreground slice group is set to 25 and as for the background slice group, it is set to a much higher value (out of range value to effectively nullify the background).

ROI identification per GOP was done by integrating the JSVM software with Intel OpenCV (version 1.0) Library.

The results show effective coding of ROI with change in position, size and rectangular shape across GOPs.



(a) 1st frame of 1st GOP (b) Middle frame of 2nd GOP (c) Middle frame of 3rd GOP (d) Middle frame of 4th GOP

Figure 8: Decoded frames of BUS test sequence (QCIF).

5 CONCLUSIONS

In this paper, three novel algorithms for constraint inter frame prediction have been proposed. The implementation of constrained inter-frame prediction algorithms on H.264 SVC reference encoder (version 8.13) gives encouraging results. There is no significant negative impact on the PSNR or bitrate of the coded video for carefully selected quantization parameter values. The computational complexity of the proposed techniques is high, and can be reduced in part by optimized implementation in software or more effectively; through hardware acceleration.

The paper also proposes the technique to support changing slice groups (moving ROI) in H.264 SVC. The technique, as implemented on JSVM (version 8.13), has been verified for both fast and slow moving video sequences. The results show effective encoding and decoding of the video sequence with ROI of changing shape, size and position.

REFERENCES

Ziliani F. and Michelou, J. (VisioWave), "Scalable video coding in digital video security", VisioWave, SA, 2005.

Mrak, M., et al., "Scalable video coding in network applications", *VIPromCom-2002, 4th EURASIP-IEEE region 8 international Symposium on Video/Image processing and multimedia communications*, Zadar, Croatia, June 2002.

Ye-Kui Wang and Hannuksela, M.H., "Isolated Regions: Motivation, Problems, and Solutions", *Input Document to JVT, JVT 3rd Meeting*, Fairfax, Virginia, USA, Document #JVT-C072, May 2002.

Bae, T.M., et al., "Multiple Region-of-Interest Support in Scalable Video Coding", *ETRI Journal, Vol. 28, Number 2*, April 2006.

Wiegand, T. et al., Joint Draft 6, *JVT 19th Meeting*, Geneva, Switzerland, Document #JVT-S201, April 2006.