LETTER

Fast Mode Decision on the Enhancement Layer in H.264 Scalable Extension*

Tae-Kyoung KIM†, Jeong-Hwan BOO†, Nonmembers, and Sang Ju PARK†(a), Member

SUMMARY_ Scalable video coding (SVC) was standardized as an extension of H.264/AVC by the JVT (Joint Video Team) in Nov. 2007. The biggest feature of SVC is multi-layered coding where two or more video sequences are compressed into a single bit-stream. This letter proposes a fast block mode decision algorithm in spatial enhancement layer of SVC. The proposed algorithm achieves early decision by limiting the number of candidate modes for block with certain characteristic called same motion vector block (SMVB). Our proposed method reduces the complexity, in terms of encoding time by up to 66.17%. Nevertheless, it shows negligible PSNR degradation by only up to 0.16 dB and increases the bit-rate by only up to 0.64%, respectively.

key words:_MPEG-4 AVC, H.264, scalable video coding, inter-layer prediction, fast mode decision

1. Introduction

These days, a wide variety of video services are delivered through various channels having different bandwidths such as wireless and cable networks. And same contents are viewed on wide variety of display devices from high-definition television (HDTV) to mobile phone. Hence the video contents should be adapted to usage environments such as network characteristics, terminal capabilities, and user preferences. In order to meet these various requirements, scalable video coding is one of very attractive solutions [1]. Recently, a new scalable video coding was standardized in the Joint Video Team (JVT) as MPEG-4 AVC/H.264 Amd. 3 Scalable Video Coding (SVC) [2], [3].

SVC has spatial, temporal, and quality scalabilities. Each scalability can be used either individually or combined with others. The base-layer is compatible with the non-scalable H.264/MPEG-4. On the other hand enhancement-layer is designed to be able to wisely use the encoded information from the base-layer or lower enhancement-layer.

In the encoding stage of spatially scalable enhancement-layer, SVC make reference to motion vectors and decoded macro-block (MB) information from lower layer for coding efficiency improvement. However, this process suffers from severe computational complexity increase because it requires iterated rate distortion (RD) computation for finding block mode having minimum RD cost. More specifically, in order to determine the macro block modes at the spatial enhancement layer, the JSVM (Joint Scalable Video Model) reference encoder [4] computes RD costs of all the candidate modes which consist of the mode predicted from the base layer and all modes in non-scalable MPEG-4 AVC/H.264 standard. Consequently, while SVC can offer a gain in coding efficiency, it can only be obtained at the cost of significant encoding complexity increase.

A few fast mode decision algorithms have been proposed to lower the encoding complexity of SVC. Li et al. proposed an algorithm that reduces candidate modes in enhancement-layer by considering the modes determined for the base-layer [5], [6]. They reduced the candidate modes when the corresponding base layer mode was 8 × 8, 16 × 8, or 8 × 16. Then, the algorithm was improved by reducing candidate modes in 16×16 modes by Lee et al. [7].

In this letter, we propose a new fast mode decision algorithm employing candidate mode reduction for spatial enhancement-layer encoder of SVC.

2. Analysis of MB Mode Predicted from Base layer

It is observed that there is correlation between the MB modes of base-layer and those of enhancement-layer. We compared MB modes of base-layer and its enhancement-layer. The condition of experiments is listed in Table 1.

Among the blocks for which mode predicted from the base layer (ModeBL) is 16×16, we define SMVB (Same Motion Vector Block) as follows. For each block with ModeBL = 16×16, let and respectively be the forward and backward motion vectors predicted from the corresponding base layer block. And also let and be the forward and backward motion vectors respectively found by conducting 16 × 16 mode RD optimization in the enhancement layer under consideration. We mark a block as SMVB when and .

| Table 1 | Condition of experiments. |
|---------|--------------------------|
| Reference Codec | JSVM 9.14 |
| Size of GOP | 8 |
| Frames | 100 |
| Motion Search Range | 32 pixel |
| Motion Search accuracy | 1/4 pixel |
| Motion Search Function | Full pixel: SAD |
| | Sub pixel: SATD |
| FGS Scalability | Do not use |
| Input Sequence | Base-layer: QCIF 15 fps |
| | Enhancement-layer: CIF 15 fps |

*Manuscript revised August 21, 2009.
†The author are with School of Electronics, Information & Communication Engineering, Hongik University, Seoul, Korea.
This work was supported by the Seoul R&BD Program No.10555.
a) E-mail: sjpark@hongik.ac.kr
DOI: 10.1587/transinf.E92.D.2545

Copyright © 2009 The Institute of Electronics, Information and Communication Engineers
Figure 1 shows how SMVB is defined. Experimental results are listed in Table 2. It should be noted that for most of the blocks, ModeBL_pred are 16 × 16 for all the test sequences and different choices of QP (Quantization Parameter). Therefore reducing the computational complexity of 16 × 16 mode blocks via fast mode decision must have great impact on reducing the overall SVC encoder complexity. Furthermore most of the 16 × 16 mode blocks are found to be SMVBs. A specific 16 × 16 mode block is marked as SMVB basically means that its characteristic in both the base layer and the enhancement layer are quite similar. That, in turn implies that it is highly unlikely that best block mode in optimal rate distortion sense found from exhaustive search in the enhancement layer has smaller block size than 16 × 16. So main idea is, for blocks marked as SMVB, to limit prediction mode search candidates as either 16 × 16 which is the optimal prediction mode already found in the base layer or SKIP.

3. Proposed Algorithm

We propose a new fast decision method for blocks with ModeBL_pred = 16 × 16 which actually take up most of the blocks as can be seen in the previous section. For blocks decided as not SMVB, find the mode that has the best RD cost, in the same manner as the conventional scalable encoder such as [4]. For blocks found to be SMVB, first, check the modes for already encoded blocks on the upper and left side of current block†. Based on the assumption that prediction modes of adjacent blocks are to be correlated, if either of the modes checked is 16 × 16 or SKIP, then limit the candidate modes of current block also to 16 × 16 and SKIP. Otherwise, do not limit the candidate modes. For blocks with mode predicted from base layer is not 16 × 16, method given in [6] is applied.

Table 3 summarizes the reduced candidates modes con-

†The first block at the top left corner (which does not have neighboring blocks to the left or above) is also encoded in conventional manner.
Table 4 Encoding time (sec).

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Mother & Daughter | JSVM    | 121   | 110   | 105   | 104   | 103   |
| Li's       | 115     | 103   | 99    | 98    | 97    |       |
| Proposed   | 81      | 62    | 52    | 45    | 40    |       |

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Foreman    | JSVM    | 156   | 136   | 126   | 116   | 112   |
| Li's       | 145     | 128   | 118   | 111   | 107   |       |
| Proposed   | 86      | 83    | 81    | 79    | 79    |       |

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Harbour    | JSVM    | 205   | 168   | 137   | 118   | 108   |
| Li's       | 192     | 155   | 128   | 110   | 102   |       |
| Proposed   | 186     | 143   | 108   | 84    | 67    |       |

Table 5 PSNR (dB) values of reproduced video.

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Mother & Daughter | JSVM    | 44.50 | 42.14 | 39.45 | 36.69 | 33.67 |
| Li's       | 44.50   | 42.14 | 39.45 | 36.69 | 33.67 |       |
| Proposed   | 44.46   | 42.07 | 39.38 | 36.62 | 33.64 |       |

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Foreman    | JSVM    | 42.61 | 39.81 | 37.07 | 34.31 | 31.38 |
| Li's       | 42.61   | 39.81 | 37.08 | 34.31 | 31.38 |       |
| Proposed   | 42.56   | 39.69 | 36.92 | 34.15 | 31.24 |       |

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Harbour    | JSVM    | 40.85 | 36.82 | 33.80 | 30.33 | 27.65 |
| Li's       | 40.85   | 36.82 | 33.80 | 30.33 | 27.64 |       |
| Proposed   | 40.83   | 36.79 | 33.25 | 30.26 | 27.57 |       |

Table 6 Bit-rates (kbps) of compressed bit-stream.

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Mother & Daughter | JSVM    | 531.9 | 281.5 | 156.3 | 89.2  | 47.3  |
| Li's       | 536.0   | 285.3 | 158.3 | 90.1  | 47.5  |       |
| Proposed   | 531.7   | 281.5 | 155.9 | 88.8  | 47.0  |       |

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Foreman    | JSVM    | 1122  | 609.1 | 337.8 | 191.9 | 108.1 |
| Li's       | 1130    | 615.8 | 343.1 | 195.7 | 110.2 |       |
| Proposed   | 1130    | 612.2 | 338.4 | 195.1 | 107.4 |       |

| Sequence   | Encoder | QP 20 | QP 25 | QP 30 | QP 35 | QP 40 |
|------------|---------|-------|-------|-------|-------|-------|
| Harbour    | JSVM    | 2669  | 1488  | 771.9 | 395.8 | 194.3 |
| Li's       | 2663    | 1493  | 775.1 | 397.5 | 194.6 |       |
| Proposed   | 2662    | 1490  | 772.3 | 396.3 | 194.4 |       |

4. Simulation Results

To check the validity of proposed method, we tested our algorithm on three test sequences (Mother and Daughter, Foreman, and Harbour) with different characteristics. Specifically, “Mother and Daughter” has little or slow movements, “Foreman” has many and fast motions, and “Harbour” has complex textures.

Simulation conditions given in Table 1 are tested on Intel Core2 Quad F3MPC’s having 2.83 GHZ clock speed and 4 GBytes of main memories.

We compared proposed method with both reference encoder which is provided from JVT [4] and Li’s method of “Selective Reduction of Candidate Modes” [6]. Performance improvements of Li’s method over reference encoder were not very significant. Therefore although the results of Li’s method are listed on the tables, we will discuss the relative performances of our method with respect only to the reference encoder. Table 4 shows the actual encoding times for 3 encoders and 3 test sequences with varying QPs. Compared to the reference encoder, computation time is reduced by from 9.27% to as much as 66.17%. Encoding time decrease was most notable for sequences with slow movements and larger value of QP. This can be expected, since for these cases, relatively more blocks are to be decided as SMVBs. For example, for Mother & Daughter sequence with $QP = 40$, speed up by a factor of 2.5 could be achieved with the proposed method compared to the reference encoder.

Table 5 shows the quality degradation due to limiting the candidates modes for SMVBs. PSNR values of proposed method are lower than those of the reference coder only by from 0.01 dB to 0.16 dB which are almost negligible, especially considering the encoding time decreases mentioned above.

Finally, Table 6 shows the bit-rates of encoded bit-streams. Even at worst case, bit-rate is increased only by 0.64% compared to the reference encoder. For some cases, bit-rates are even decreased by small amount.

5. Conclusion

We proposed a new fast mode decision method for spatial enhancement layer of SVC. Experimental results verified that proposed method improves the performance by significantly reducing the encoding time while only slightly lowering the reproduced video quality and compression ratio.

References

[1] J.-R. Ohm, “Advances in scalable video coding,” Proc. IEEE, vol.93, no.1, pp.42–56, Jan. 2005.
[2] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), “Joint draft 10 of SVC amendment,” JVT-W201, San Jose, April 2007.
[3] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), “ITU-T Rec. H.264 ISO/IEC 14496-10 version 8,” Nov. 2007.
[4] Joint Video Team (JVT of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/Sc29/WG11 and ITU-T SG16 Q.6), “JSVM 9 software,” JVT-A203, Hannover, July 2008.
[5] H. Li, Z.G. Li, and C. Wen, “Fast mode decision for spatial scalable video coding,” Proc. ISCAS, pp.3005–3008, May 2006.
[6] H. Li, Z.G. Li, and C. Wen, “Fast mode decision algorithm for inter-frame coding in fully scalable video coding,” IEEE Trans. Circuits Syst. Video Technol., vol.16, no.7, pp.889–895, July 2006.
[7] B. Lee, M. Kim, S. Harm, C. Park, and K. Park, “A fast inter-layer mode decision method in scalable video coding,” The Korean Society of Broadcast Engineers, Journal of Broadcast Engineering, vol.12, no.4, pp.360–372, July 2007.