

A method to reduce the complexity of conditional replenishment algorithm for hybrid 3DTV with mixed resolution

Kyeong-Hoon Jung¹ · Sung-Hoon Kim² ·
Dong-Wook Kang¹

Received: 4 November 2015 / Revised: 8 January 2016 / Accepted: 29 January 2016 /
Published online: 4 February 2016
© Springer Science+Business Media New York 2016

Abstract SC-MMH (Service Compatible 3DTV using Main and Mobile Hybrid) is a kind of hybrid 3DTV system where the resolutions of stereoscopic views are not identical and has become one of ATSC standards for terrestrial 3DTV. And CRA (Conditional Replenishment Algorithm) has been proposed to enhance the visual quality of this asymmetrical hybrid 3DTV with mixed resolution. CRA uses a variable-sized square block as a PU (Processing Unit) and the quad-tree is employed to represent the overall hierarchical structure. The largest and smallest sizes of PU of quad-tree affect the RD (Rate-Distortion) performance as well as the computational complexity of CRA. In this paper, we examine the effect of quad-tree levels on the BD-PSNR and the processing time. And we suggest the proper set of top and bottom levels which reduces the complexity of CRA with a negligible degradation of quality. Simulation for the stereoscopic sequences of HD resolution showed that the processing time of CRA can be considerably reduced to 70 % but the decrease of BD-PSNR is only 0.02 dB by employing of the combination of (7, 2) levels instead of conventional combination of (8, 0) levels. And the processing time could be more reduced to 60 % for UHD sequences when we used the combination of (7, 3) levels.

Keywords Hybrid 3DTV · CRA (Conditional Replenishment Algorithm) · VEI (Video Enhancement Information) · Quad-tree · Inter-view vector

✉ Kyeong-Hoon Jung
khjung@kookmin.ac.kr

¹ School of Electrical Engineering, Kookmin University, 77 Jeongneung-roSeongbuk-guSeoul 136-702, South Korea

² Electronics and Telecommunications Research Institute, 218 Gajeong-roYuseong-guDaejeon 305-700, South Korea

1 Introduction

There have been various kinds of approaches to initiate the stereoscopic 3D broadcasting service [3]. And SC-MMH (Service Compatible 3DTV using Main and Mobile Hybrid) system has become one of prominent standards for terrestrial 3DTV by ATSC [4, 11]. The stereoscopic 3D video can be transmitted via two independent video elementary streams in SC-MMH system, where one video (e.g. left view) is compatible with the legacy 2D TV service based on ATSC A/53 and the other (e.g. right view) is compatible with ATSC Mobile DTV service based on ATSC A/153 [1, 2]. Therefore the backward-compatibility with the existing fixed and mobile 2D services is completely guaranteed and it is the main advantage of SC-MMH 3DTV. Figure 1 illustrates the concept of SC-MMH system.

The environments for fixed and mobile services in SC-MMH system are quite different, that is, one view is encoded by the MPEG-2 codec with HD resolution of higher quality and the other view is encoded by the AVC codec with mobile resolution of lower quality. This problem of resolution or quality mismatch can be resolved by the binocular suppression effect which says the perceived quality of the stereoscopic views by human visual system is strongly weighted towards that of the higher quality view when the qualities are different from each other [12]. However, it is required to improve the quality of mobile view especially when the quality gap between two views is quite large or there exist severe coding artifacts. The CRA (Conditional Replenishment Algorithm) has been proposed to deal with this mismatch problem [10].

In order to compose 3D video at the receiver, the small-sized view from mobile channel needs to be enlarged to the same size of the view from main channel. The straightforward method is to simply interpolate it. But we know there is strong correlation between two stereoscopic views and the basic concept of CRA is to utilize this correlation. CRA can be considered as a switching process between two candidate as shown in Fig. 2; one is the view which is compensated from the main view by disparity or inter-view vector and the other is simply interpolated from the mobile view. The inter-view vector means 2D vector which represents the shifted amount between the left and right views of 3D video. And the better substitute is selected and assigned as mode to generate the enhanced view. The optimum inter-view vector and mode information need to be estimated at the encoder and transmitted as overhead data to be utilized at 3DTV receiver. This supplemental information is generated at the encoder and called as VEI (Video Enhancement Information) [5]. When VEI is transmitted, A SC-MMH receiver decodes VEI and can generate the 3D view of enhanced quality. It was showed that just a small amount of overhead data could greatly improve the visual quality of 3D video by using CRA [10].

Since VEI requires some additional bandwidth, it is desirable to keep the amount of it as small as possible. Thus the pixels which have the same inter-view vector and mode are merged

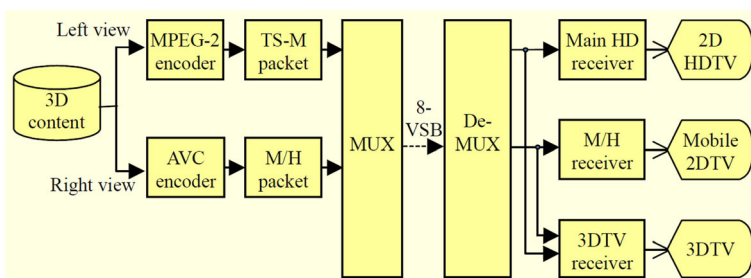


Fig. 1 Block diagram of SC-MMH 3DTV system [11]

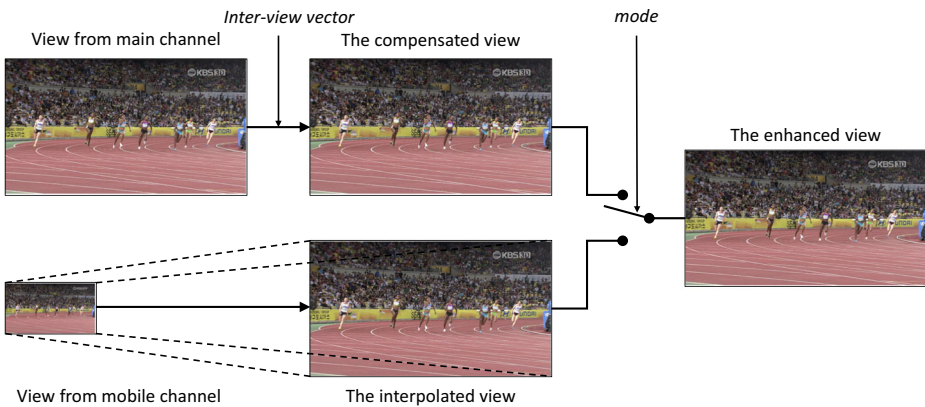


Fig. 2 The concept of CRA (Conditional Replenishment Algorithm)

into a single PU (Processing Unit) and the hierarchical quad-tree structure is employed to represent the overall distribution of PUs in CRA. Another issue of CRA is the computational complexity to obtain VEI. In order to find the optimum mode and inter-view vector of variable-sized PU, we need to calculate every inter-view vectors at each level of CRA and compare their rate-distortion performances. Thus it is undesirable to excessively increase the depth of quad-tree structure from the viewpoint of complexity.

In this paper, we investigate the effect of hierarchical structure of quad-tree on the performance and complexity of CRA. In section 2, the quad-tree structure of VEI is discussed. And the top and bottom level of quad-tree are determined in section 3 by examining the trade-off relationship between quality degradation and processing time. Finally, the conclusion is given in section 4.

2 The hierarchical quad-tree structure of CRA

CRA adopt a hierarchical quad-tree structure which is composed of PUs, to represent the switching mode for 3D composition. PU plays a role of basic unit for CRA and every pixel in a single PU has the same mode or inter-view vector. The optimum size and mode of PU depends on the spatial characteristics and inter-view correlation of two stereoscopic views.

In order to estimate the optimum mode, the RD (Rate-Distortion) cost function is employed as Eq. (1) where D and R denote MAD (Mean Absolute Difference) and the amount of required bits, respectively, and λ corresponds to the Lagrange multiplier.

$$J = D + \lambda R \quad (1)$$

At a given level, the cost functions of every possible modes are calculated for each PU of which size is fixed as $2^{level} \times 2^{level}$. And the mode which has the smallest cost is selected as the candidate mode of that PU. As described in introduction, there are basically two modes; one is 'M mode' to select the inter-view compensated view and the other is 'H mode' to select the simply interpolated view. In addition, the third candidate is considered because there is much temporal correlation in inter-view vectors and modes of consecutive frames. This third 'S mode' indicates the mode and inter-view vector are the same as those of the collocated PU in previous frame. But the third mode is not always allowed in order to create a random access point. Therefore there are two frame types in mode decision of CRA; one is INTRA type in which 'M mode' and 'H mode'

are available and the other is INTER type in which the third ‘S mode’ is also included [5]. The resulting mode and frame type of CRA are summarized in Table 1.

After every PU at every level is examined, the next step is to determine whether four child-PUs need to be merged or not. The sum of costs of four child-PUs at the current level is compared with the cost of their parent PU at the upper level. And they are merged into a single PU when the cost at the upper level is smaller. Consequently, CRA generates a hierarchical distribution of variable-sized PUs with quad-tree structure.

The Lagrange multiplier λ , which used for RD optimization process, plays an important role to control the partition of quad-tree and to regulate the bitrate of additional data. When we use smaller λ , the small-sized PU is preferred and vice versa. When we use fixed λ , the amount of overhead bitrate for VEI is variable but it can be controlled as constant as possible by adopting variable λ [6]. Typical examples of applying CRA are shown in Fig. 3, where black empty PU denotes ‘H mode’ and the other PU denotes ‘M mode’ and the PUs are successfully partitioned into hierarchical quad-tree structure. As expected, the higher PSNR gain was obtained when more bitrate is used for VEI.

Meanwhile, the hierarchical depth of quad-tree which is specified by the largest PU (LPU) and the smallest PU (SPU) is another important parameter of CRA. Generally speaking, as the size of PU becomes smaller, the tiny details are likely to be well recovered and the PSNR gain becomes greater. On the contrary, we can expect more coding gain when the size of PU becomes larger because only a single mode is assigned to each PU. As a result, it is better to keep the levels of quad-tree structure as large as possible from the viewpoint of PSNR gain.

However, the complexity can be a critical problem when the number of levels is too large since the inter-view vectors need to be estimated at every level and the computational load for calculation of inter-view vectors is considerable. Thus it is required to limit the depth of quad-tree in order to reduce the complexity. That is the reason why it is important to determine the smallest and largest size of PU, that is, the bottom and top level to balance the PSNR performance and complexity.

3 Determination of quad-tree levels of CRA

3.1 Simulation environment for HD sequences

Firstly, total 10 HD sequences with vertical resolution of 1080p are used for simulation. Figure 4 shows typical scenes of test sequences which are obtained from real video and animation. As you can see, the spatial and temporal characteristics of test sequences are quite diverse.

According to the service scenario of SC-MMH system, the left view is encoded by MPEG-2 with 12Mbps. And the right view is firstly resized to the resolution of 240p and encoded by

Table 1 The mode and frame type of CRA

Mode	Description	Available frame type
M mode	To compensate the main view by inter-view vector	INTRA, INTER
H mode	To interpolate the mobile view	INTRA, INTER
S mode	To use the mode and inter-view vector of previous frame	INTER

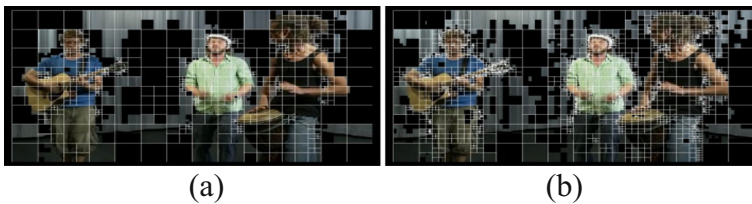


Fig. 3 Examples of quad-tree structure (PSNR at additional bitrate). **a** 36.6 dB at 120 kbps **b** 37.5 dB at 480 kbps

AVC with 480 kbps. The reference case is when the SPU is at level 0 and LPU is at level 8, that is, the smallest and largest sized of PU are 1×1 and 256×256 , respectively.

In order to measure the quality gain of CRA, we use the BD-PSNR [7]. The PSNRs of the enhanced right view of test sequences are calculated and averaged. And the differences of BD-PSNR between the reference and comparison cases are compared. It can be thought that the quality degradation is less when the difference is smaller.

And the TR (Time Ratio) is used as relative measure of computational complexity [9]. The TR can be calculated as Eq. (2), where T_c and T_r are the processing times of comparison and reference cases, respectively. The processing times of CRA are measured and averaged for every test sequences. And the ratio between the processing time of comparison case and that of reference case is calculated. As the complexity reduces, we can get the smaller TR.

$$TR = T_c/T_r \times 100 (\%) \quad (2)$$

3.2 Decision of the top and bottom levels of quad-tree

The first experiment is to decide the top level, that is, the size of LPU (Largest PU). We fix the smallest size of PU to 1×1 (level 0). And the sizes of LPU are changed from 32×32 (level 5) to 256×256 (level 8). As said before, the reference is taken as the case when the top level is 8, and the remaining three cases are compared. For this experiment, the Lagrange multiplier λ is fixed to 100, 200, 400, and 800.

It can be found from Fig. 5 that the difference of BD-PSNR decays exponentially as the top level decreases. When LPU is 128×128 (level 7), the difference of BD-PSNR is quite small. But it decreases to -0.07 dB when LPU is 64×64 (level 6) and there is a considerable gap of -0.33 dB when LPU is further reduced to 32×32 (level 5). Therefore, it is desirable to select the size of LPU not as small as 64×64 . Meanwhile, Fig. 5 says that TR decreases almost linearly as the top level decreases. If the level of LPU is reduced by one, then the decrease rate of TR is about 13 %.

It is worthy to notice the unique characteristic especially when the sequence is obtained from computer graphics. The Fig. 6 shows the difference of BD-PSNR according to the top

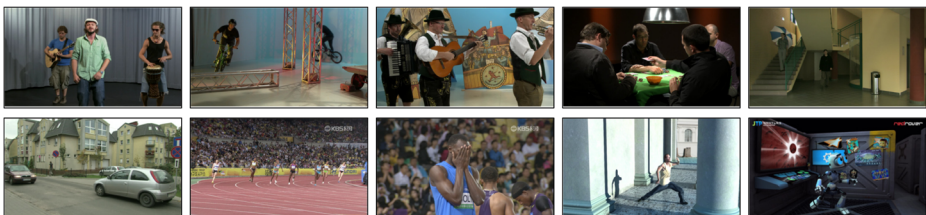


Fig. 4 Test HD sequences

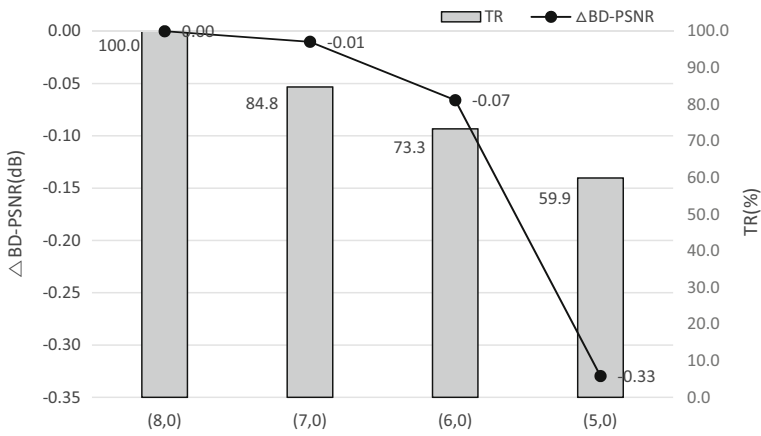


Fig. 5 TR and ΔBD-PSNR according to the top level for HD sequences

level when we divide the test sequences into two groups; one is from real video and the other is from animation. As expected, the BD-PSNR shows a remarkable drop for animation sequence as the top level decreases since the PSNR gain from using large-sized PU disappears. This fact means that it is desirable to determine the sizes of PUs according to the genre of sequence.

Next experiment is to decide the bottom level, that is, the size of SPU (Smallest PU). For this purpose, the largest size of PU is fixed to 256×256 (level 8) and the sizes of SPU are changed from 1×1 (level 0) to 8×8 (level 3). The bottom level of reference case is level 0, that is, the size of SPU is 1×1 , and the remaining three cases are compared.

Figure 7 shows the experimental result. It can be noticed that the difference of BD-PSNR is negligible until the size of SPU becomes 4×4 (level 2), but there is a considerable drop when the bottom level is 3. Therefore, it is desirable to limit the size of SPU not as large as 8×8 in order to avoid severe degradation in BD-PSNR. Meanwhile, TR decreases as the bottom level increases as before and the average rate of decrease is about 10 %.

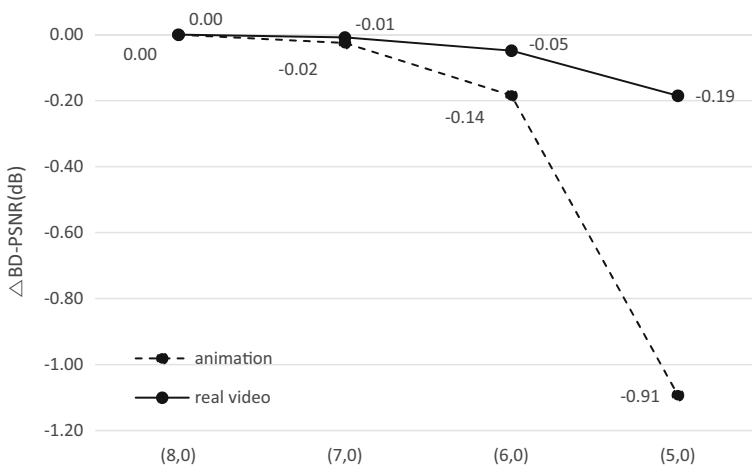


Fig. 6 ΔBD-PSNR comparison of real and animation sequences according to the top level

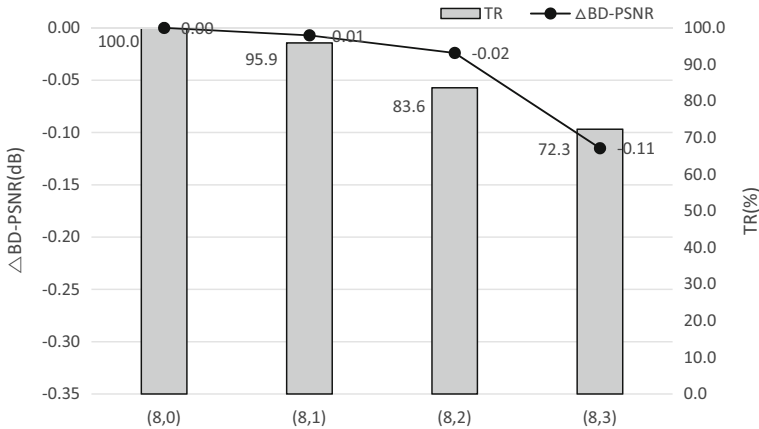


Fig. 7 TR and ΔBD-PSNR according to the bottom level for HD sequences

3.3 The result of combining the top and bottom levels

From the above experimental results, it is reasonable to choose the level 7 as a top level and level 2 as a bottom level, that is, to employ the combination of (7, 2) levels. In this case, the size of LPU is 128×128 and that of SPU is 4×4 and the hierarchical depth of quad-tree is six. Next experiment is to examine the performance and complexity when we employ this combination. The target of reference is when the sizes of PU is from 1×1 to 256×256 , that is, the combination of (8, 0) levels is used. For this experiment, the Lagrange multiplier λ is variable and the target bitrate for VEI are changed from about 120 to 480 kbps with a step size of 120 kbps.

Figure 8 shows that the RD curves of reference and comparison cases are very close. The quality gap between them is only -0.02 dB in terms of BD-PSNR and almost indistinguishable from the perceptual point of view. Meanwhile, it can be noticed from Fig. 9 that the TR remains nearly constant and slightly over than 70 % regardless of the amount of additional bitrate for VEI.

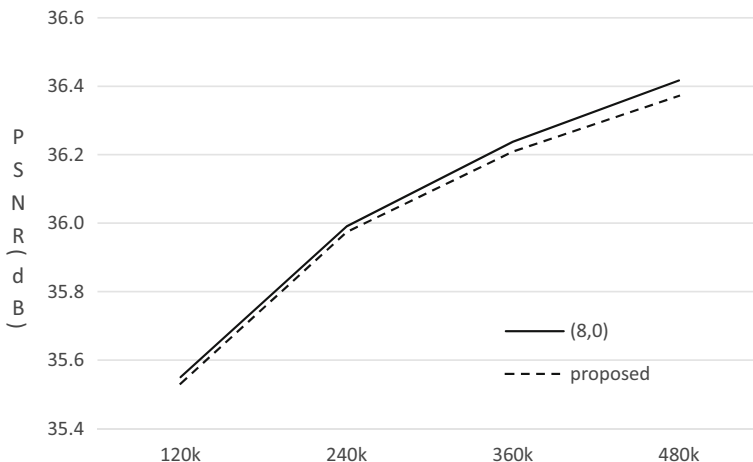


Fig. 8 RD curves of combination of (7, 2) level according to overhead bitrate for HD sequences

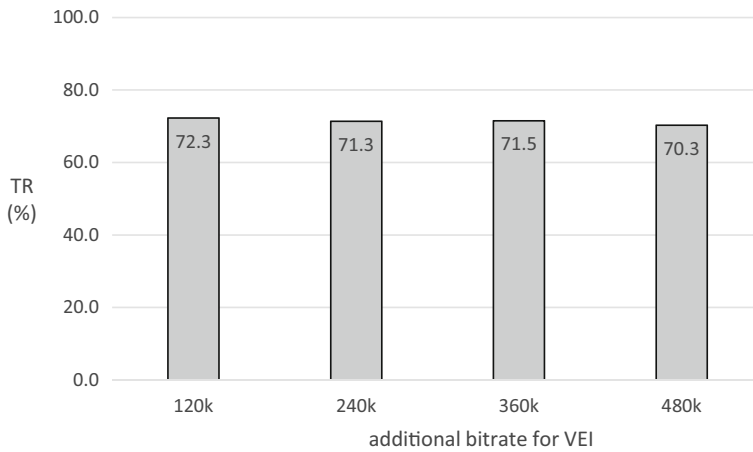


Fig. 9 TR of combination of (7, 2) level according to overhead bitrate for HD sequences

Consequently, we can get about 30 % reduction of computational complexity but the penalty of PSNR degradation is almost negligible by using the combination of (7, 2) levels.

3.4 The quad-tree levels of CRA for UHD sequences

Similar experiment is performed for several UHD sequences as shown in Fig. 10. In this case, the vertical resolution of left view is 2160p and we used 240p as a vertical resolution of right view. Both views are encoded by HEVC and the bitrates are 19 and 500 kbps, respectively [8]. And the reference case is when the SPU is at level 0 and LPU is at level 8, as before.

Figure 11 shows that the results of CRA according to the top level when SPU is fixed to level 0. The difference of BD-PSNR is only -0.010 dB when we use (7, 0) combinations but there is considerable decrease when the LPU is at level 5. The complexity decreases almost linearly and the slope is about 13 % as the level of LPU is reduced by one.

Meanwhile Fig. 12 shows that the results of CRA according to the bottom level when LPU is fixed to level 8. As you can see, the difference of BD-PSNR is only -0.006 dB when we use

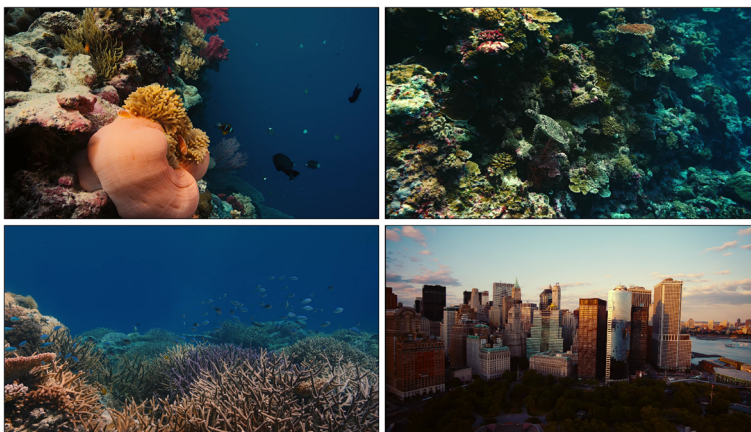


Fig. 10 Test UHD sequences

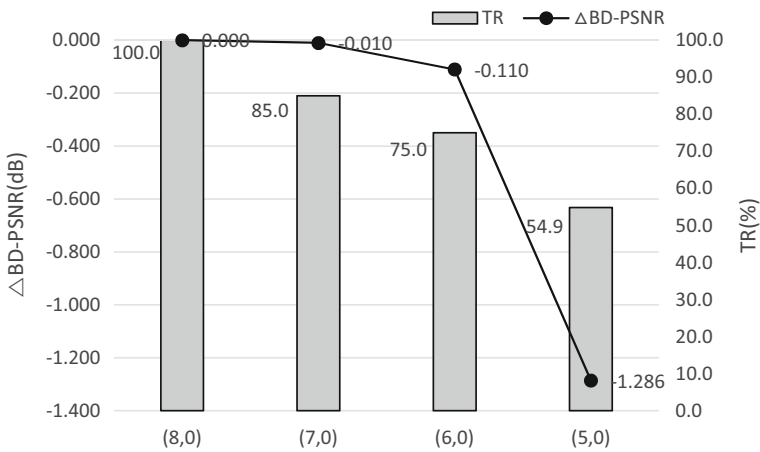


Fig. 11 TR and Δ BD-PSNR according to the top level for UHD sequences

(8, 2) or (8, 3) combinations. The average PSNR degradation is lower than that of HD sequences, since there is much spatial correlation in UHD sequences. The complexity decreases almost linearly and the slope is about 9 % as the level of SPU increments by one.

Based on the above experiments, the combination of (7, 3) levels can be a good choice for UHD sequences in terms of complexity. The Figs. 13 and 14 shows the RD curves and TR when we apply CRA and use (7, 3) combination for UHD sequences. The computational complexity can be greatly reduced by about 40 % without PSNR penalty.

4 Conclusion

SC-MMH is one of the ATSC standards for terrestrial 3DTV system and CRA is a very efficient solution to deal with the quality gap between two stereoscopic views of SC-MMH. The 3D visual quality can be greatly enhanced with additional of VEI which is generated by

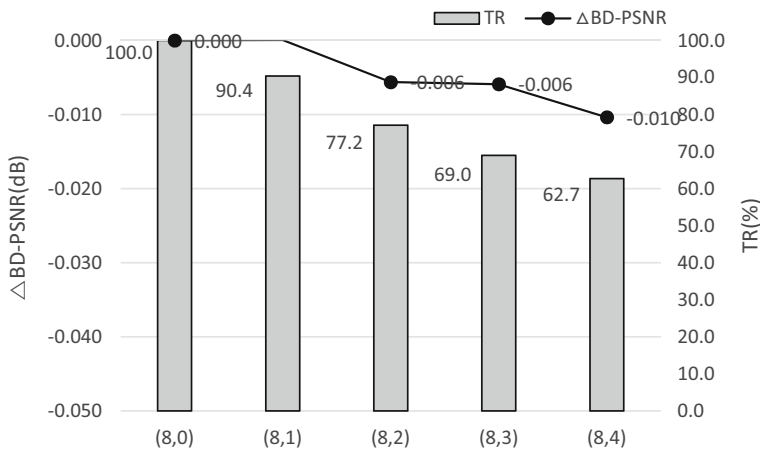


Fig. 12 TR and Δ BD-PSNR according to the bottom level for HD sequences

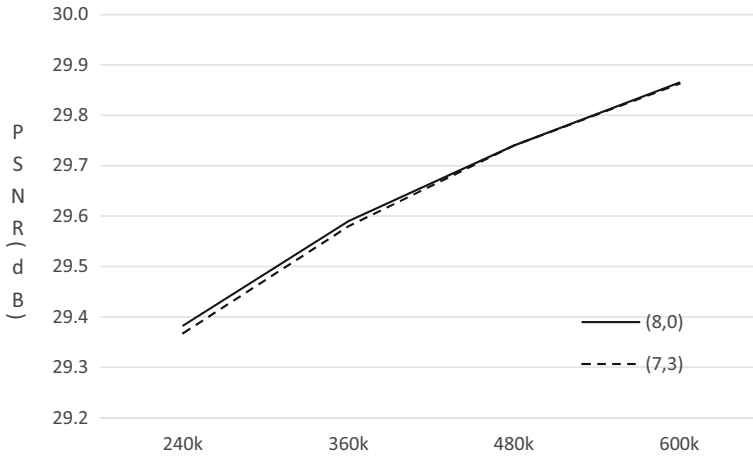


Fig. 13 RD curves according to overhead bitrate for HD sequences

applying CRA. We showed that the top or bottom levels of quad-tree are closely related to not only the RD performances but also the computational complexity. From the viewpoint of PSNR gain, the large depth of quad-tree is advantageous. But when the depth is too large, the processing time of CRA can be one of the critical factors to be considered. Therefore the hierarchical structure of quad-tree needs to be carefully determined.

In this paper, we examined the effect of the sizes of largest and smallest PU on the quality and the complexity of CRA. Total 10 HD and 4 UHD video sequences were tested for simulation. To find the effect of top level, SPU was fixed to 1×1 (level 0) and TR and BD-PSNR were compared as the LPU varied from 32×32 (level 5) to 256×256 (level 8). On the contrary, LPU was fixed to 256×256 (level 8) and TR and BD-PSNR were compared as the SPU varied from 1×1 (level 0) to 8×8 (level 3) or 16×16 (level 4) to decide the bottom level. Simulation results showed that the complexity and RD performances are in the trade-off relation. In case of HD sequences, the processing time can be reduced by 30 % when we employed the combination of (7, 2) levels. And we could get 40 % reduction of computational complexity by using the combination of (7, 3)

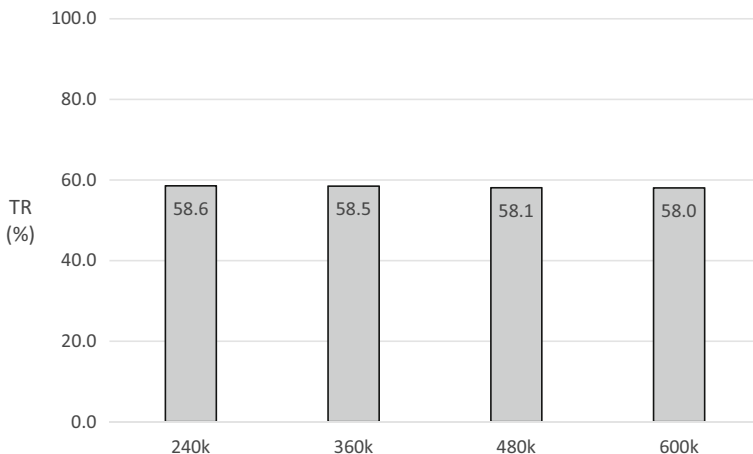


Fig. 14 TR of combination of (7, 2) level according to overhead bitrate for HD sequences

levels for UHD sequences. However the decrease of BD-PSNR is nearly negligible as compared to the conventional combination of (8, 0) levels. Therefore it is expected that the suggested set of top and bottom levels is used for fast implementation of CRA.

Acknowledgments This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program.

References

1. ATSC (2013) ATSC digital television standard, part 3 – Service multiplex and transport subsystem characteristics. Doc. A/53, Part 3:2013
2. ATSC (2013) ATSC mobile/handheld digital television standard, part 3 – Service multiplex and transport subsystem characteristics. Doc. A/153 Part 3:2013
3. ATSC (2014) ATSC standard: 3D-TV terrestrial broadcasting, part 1. Doc. A/104, Part 1:2014
4. ATSC (2014) ATSC standard: 3D-TV terrestrial broadcasting, part 5 - Service compatible 3D-TV using main and mobile hybrid delivery. Doc. A/104, Part5:2014
5. ATSC (2015) Candidate standard: revision of 3D-TV terrestrial broadcasting, part 5 – service compatible 3D-TV using main and mobile hybrid delivery. Doc. S12-237r6
6. Bang M, Kang D, Kim S, Jeong S, Jung K (2014) Quality control of conditional replenishment algorithm for hybrid 3DTV with mixed resolution. Proc. of ICCE, 410–411. doi: [10.1109/ICCE.2014.6776062](https://doi.org/10.1109/ICCE.2014.6776062)
7. Bjøntegaard G (2001) Calculation of average PSNR differences between RD curves. Doc. VCEG-M33
8. Dissanayake M, Abeyrathna D (2015) Performance comparison of HEVC and H.264/AVC standards in broadcasting environments. J Inf Process Syst 11(3):483–494. doi:[10.3745/JIPS.03.0036](https://doi.org/10.3745/JIPS.03.0036)
9. Goswami K, Kim B, Jun D, Jung S, Choi J (2014) Early coding unit-splitting termination algorithm for high efficiency video coding (HEVC). ETRI J 36(3):407–417. doi:[10.4218/etrij.14.0113.0458](https://doi.org/10.4218/etrij.14.0113.0458)
10. Jung K, Bang M, Kim S, Choo H, Kang D (2014) Enhancement for hybrid 3DTV with mixed resolution using conditional replenishment algorithm. ETRI J 36(5):752–760. doi:[10.4218/etrij.14.0113.1092](https://doi.org/10.4218/etrij.14.0113.1092)
11. Kim B, Bang M, Kim S, Choi J, Kim J, Kang D, Jung K (2012) A study on feasibility of dual-channel 3DTV service via ATSC-M/H. ETRI J 34(1):17–23. doi:[10.4218/etrij.12.0111.0270](https://doi.org/10.4218/etrij.12.0111.0270)
12. Stelmach L, Wa JT, Meegan D, Vicent A (2000) Stereo image quality: effects of mixed spatio-temporal resolution. IEEE Trans Circuits Syst Video Technol 10(2):188–193. doi:[10.1109/76.825717](https://doi.org/10.1109/76.825717)



Kyeong-Hoon Jung received the B.S., M.S., and Ph.D. degrees from the Dept. of Electronics Eng., Seoul National University, Seoul, Korea, in 1987, 1989, and 1996, respectively. After working at the Korea Broadcasting Institute from 1991 to 1997, he joined Handong University as a faculty member. In 2005, he transferred to Kookmin University, where he is now a professor in the School of Electrical Engineering. His research focused on broadcasting system, video coding, and multimedia signal processing.



Sung-Hoon Kim received B.S., M.E., and Ph.D. degrees in electronics engineering from Kookmin University, Seoul, KOREA in 1994, 1996 and 2008, respectively. Since 2000, he has been a Principal Researcher in Electronics and Telecommunications Research Institute (ETRI), Korea. His research interests are digital signal processing in field of Terrestrial & Mobile DTV, 3DTV and UHD TV broadcasting system.



Dong-Wook Kang received the B.S., M.S., and Ph.D. degrees from the Dept. of Electronics Eng., Seoul National University, Seoul, Korea, in 1986, 1988, and 1995, respectively. He joined as a faculty member for the Dept. of Electrical Eng., Kookmin University, Seoul, Korea, in 1995 and now is a Professor of School of Electrical Eng., Kookmin University. He is a trustee of the Korean Society of Broadcasting Engineers. His research area includes video coding, multimedia signal processing, 3DTV and digital culture technology.