A FLEXIBLE REFERENCE PICTURE MANAGEMENT SCHEME

Xiaolong Jiang[1], Qin Yu[2], Kai Zhang[3], Siwei Ma[2], Honggang Qi[1], Shaw-Min Lei[3]

[1]University of Chinese Academy of Sciences, Beijing, China
xljiang@jdl.ac.cn, hgqi@ucas.ac.cn
[2]Institute of Digital Media, Peking University, Beijing, China
{qyu, swma}@pku.edu.cn,
[3]Mediatek (Beijing) Inc., Beijing, China
{kai.zhang, shawmin.lei}@mediatek.com

ABSTRACT

Reference picture management is an important part in video encoding and decoding process, and has great influence on the coding performance. In this paper, a flexible reference picture management scheme is proposed for the second generation of Audio Video Coding Standard (AVS2). In the proposed scheme, reference configuration set (RCS), which consists of reference picture information, is used to manage the reference picture. Based on RCS, the reference picture set for current coding picture can be arbitrarily configured. Experimental results show that the proposed flexible reference picture management scheme achieves significant bitrate reduction in AVS2 encoder. For low delay P and random access common test condition, the average coding gain can be up to 4.3% and 5.1% respectively.

Index Terms**—**AVS2, reference configuration set (RCS), reference picture management

1. INTRODUCTION

The first generation of Audio Video Standard (AVS1) was developed by the AVS Working Group of China [1] and has been officially approved as the China national Standard in 2006 [2]. AVS1 was extended in many aspects and brought huge revolution for video coding in China. Now AVS1 has been widely used in many applications, such as internet video, set top box, chip designing and TV broadcasting. However, with the growing needs for HD video, the coding performance of AVS1 no longer satisfies the increasing compression requirement. Hence, the AVS Working Group started the standardization work of the second generation of AVS in 2012, which is also called as AVS2 [3]. Compared with AVS1, AVS2 has 49.5% bit rate saving for random access common test condition and 44.0% bit rate saving for low delay P common test condition [4].

The encoder framework of AVS2 is shown in Fig. 1. For inter prediction there are nine inter prediction modes in AVS2, including direct/skip mode, 2Nx2N, 2NxN, Nx2N partition prediction modes and four asymmetric partition modes [3]. Moreover, weighted-skip mode and bi-prediction technique are used for inter prediction. For intra prediction there are 33 luminance prediction directions and 5 chrominance prediction directions [3]. The transform block size of AVS2 is available within the set of {8x8, 16x16, 32x32}. And the entropy coding used in AVS2 is arithmetic entropy coding (AEC).

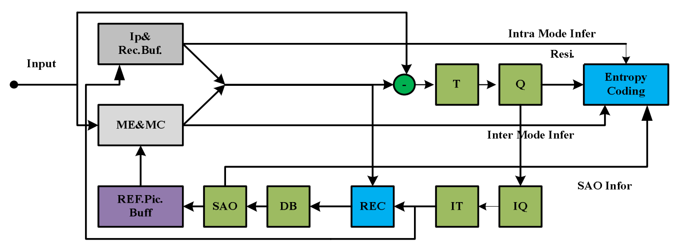


Fig.1. Framework of AVS2

For inter prediction, the coding performance can be improved with more reference pictures. However, buffering too many reference pictures may require more memory resources and increase the computational complexity. In AVS1, the maximum number of reference pictures is fixed as 2. One forward picture and one backward picture are selected as reference pictures for B picture and two forward reference pictures for P picture [5]. And same with AVS1, AVS2 also adopts multiple reference pictures for inter prediction. There are two reference pictures for hierarchical B coding structure and four reference pictures for low delay P coding structure [3]. The optimal reference pictures are selected according to the following equation

$$ref_{opt} = \arg\min_{n}(SAD_n(B_c, B_p) + \lambda R_n(MVD)) \quad (1)$$

where B_c and B_p denote the current coding block and the reference block, R_n *(MVD)* denotes the number of coding bits of motion vector difference and reference index.

To design a better reference management strategy and improve the coding efficiency of AVS2, this paper proposes a flexible reference management scheme and optimization algorithm for low delay P coding structure. The rest of the paper is organized as follows. Section 2 reviews the reference picture management scheme that AVS2 adopted before and introduces the related proposals. Section 3 describes the proposed reference picture management scheme. And Section 4 gives the experimental results. Finally, Section 5 draws the conclusion.

2. RELATED WORKS

Currently there are three kinds of coding structure in AVS2, low delay P structure, all intra coding structure and hierarchical B coding structure. Before the 46th AVS2 meeting, the reference picture management schemes that AVS2 adopted has some disadvantages. In low delay P coding structure, the four nearest reconstructed pictures are chosen as reference pictures [6], as is shown in Fig. 2. And AVS2 used a sliding window to remove unused picture from the decoded picture buffer (DPB). In hierarchical B coding structure one forward reconstructed picture and one backward reconstructed picture are selected as reference pictures, as is shown in Fig. 3. And AVS2 defined six layers for hierarchical B coding structure maximally and each layer has a different reference adjustment strategy [6]. The encoder allocates each coding picture to a fixed layer and adjusts the reference picture according to the layer. The reference structure and DPB management mechanisms of hierarchical B coding structure was fixed according to each layer of the coding picture.

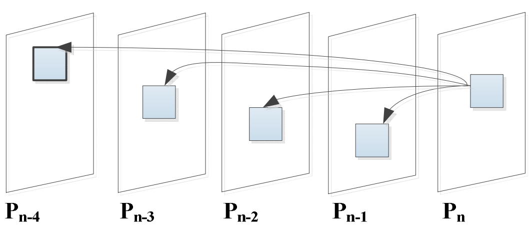
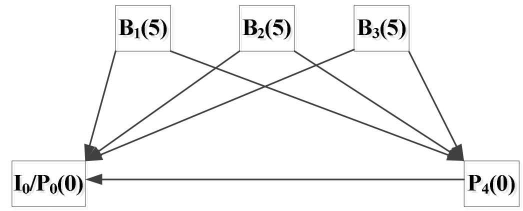
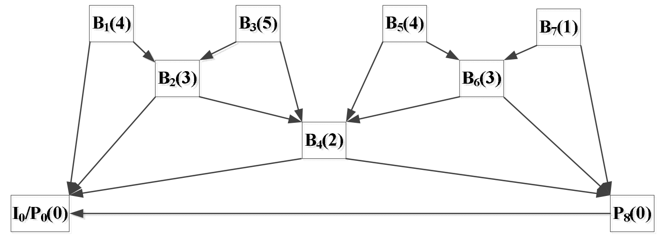


Fig. 2. Low delay P coding structure

The layer based reference picture management scheme is not very flexible and may not be able to get adapted to all the sceneries. For instance, the nearest four reconstructed pictures may not be the best reference pictures since the quality of these pictures may vary. In hierarchical B frame coding structure, six layers structure cannot adapt to different size of Group of Picture (GOP) well. The encoder will not work if the GOP size is larger, such as GOP size 16 and GOP size 24. In addition, the coding order is not flexible. Taking GOP size 8 of hierarchical B coding structure as an example, the coding order must be 0, 8, 4, 2, 1, 3, 6, 5 and 1 and cannot be 0, 8, 4, 2, 6, 1, 3, 5 and 7.



(a) GOP size = 4



(b) GOP size = 8

Fig.3. Hierarchical B coding structure

To resolve the inflexible characteristic of reference picture management scheme, many reference picture management schemes are proposed [7] [8] [9] [10] [11]. In [7], a reference layer based method was proposed. The total number of layer was extended from 6 to 8 and the maximum layer ID was set 7. The reference layer based strategy solved the problem of maximum GOP size and could support large GOP size well, e.g. 16. However, this strategy still has some limitations. Firstly, its reference structure is fixed according to the layer of the coding picture and it is unable to set reference picture for each coding picture flexibly. And lower layer picture cannot be set as reference picture for higher layer picture. Secondly, the coding order is fixed according to layer and the output order will be wrong once the coding order is changed.

3. PROPOSED REFERENCE MANAGEMENT SCHEME

In this section a flexible reference picture reference scheme based on RCS is proposed. Through the information in RCS, the encoder can arbitrarily configure the reference picture set for current coding picture.

3.1. Reference configuration set

RCS consists of reference picture information of current coding picture including the COI, playing order index (POI), QP offset, the number of reference pictures, delta COI between current picture and reference picture, number of pictures that need to remove from buffer and delta COI between pictures from buffer and current picture.

The coding order index (COI) is used to represents the coding order of current picture and reference pictures. And playing order index (POI) is used to represent the playing order of current picture and reference pictures. QP offset is used to set the QP of current coding picture. And the number of reference pictures and the delta COI between current picture and reference picture are used to set up reference picture set for current picture. Meanwhile, the number of pictures that need to remove from buffer and delta COI between pictures from buffer and current picture are used to manage DPB.

At the encoder side the syntax element *coi* is used to transmit the COI of current picture, which is defined as follows

$$coi = COI \% 256 \quad (2)$$

By marking and transmitting the COI information, the decoder side could figure out the actual coding order of current picture even there are some pictures lost during transmission. Meanwhile, the *delta_poi* is defined to derive the POI in decoder side, which is defined as the equation bellow.

$$delta\_poi = COI - POI \quad (3)$$

The syntax element *delta_poi* is used to denote the only display order of current picture. And the decoder side derives the *POI* of current coding picture by subtract *delta_poi* from *COI*.

In the reference picture list and removed picture list, the delta COI of reference pictures is recorded. And by deriving the COI and POI of reference pictures the reference pictures for current coding picture can be set.

At the encoder side, it is necessary to write the reference manage information with additional bit. However, it is redundant to transmit RCS for each picture since the RCS is of great similarity for the same position in the different GOP. The bits for reference information will account for a high percentage especially for low resolution video sequences. In our proposed scheme, several RCS sets are used and signaled in the sequence header to save bit rate instead of writing reference configuration set of each coding picture in picture header. And only the index of RCS is transmitted in picture header, signaling the reference scheme that the current coding or decoding picture takes. Meanwhile, a special RCS set can be also written into picture header if none of the RCS scheme from the sequence header is taken for the current decoding picture. And a flag is used to signal whether an index of sequence header is used or a special RCS is transmitted in picture header.

3.2. Encoding process

After the determination of RCS, the next step is to encode the information of RCS and manage the DPB. In the proposed scheme, the encoder reads reference management information from the configuration file and then initializes the reference information of current picture. The COI and POI of current coding picture is configured in encoder side. The reference list and total number of reference pictures are initialized manually. The whole process includes the following four aspects.

Firstly, locate the position of current coding picture in the RCS set.

$$rcs\_index = (COI - 1) \% GOP\_size \quad (4)$$

where *rcs_index* represents the actual position of in the RCS from configuration files. And COI denotes the actual coding order of current picture and *GOP_size* denotes the size of each coding group of picture. And the reference list will not be set up if the COI equals 0, namely the type of current picture is Intra picture.

Secondly, configure the QP for current coding picture. The QP of current coding picture is configured according to the QP offset in the RCS as bellow.

$$QP = QP_I + QPOffset \quad (5)$$

where the *QPOffset* is derived from the RCS of current picture.

Thirdly, set up the reference list. The total number of reference pictures and delta COI of reference pictures are derived from the RCS of current picture. The COI and POI of reference pictures is derived as bellows.

$$COI\_ref[i] = COI - delta\_coi\_ref[i] \quad (6)$$

where *COI_ref* denotes the actual coding order of reference pictures and *delta_coi_ref* is derived from the RCS of current picture. The COI that equals *COI_ref* in the buffer is selected as reference picture and the reference picture is considered as non-exist when the *COI_ref* is lower than 0 or the POI of reference picture is smaller than the playing order of random access point (RAP), and then set the COI and POI of RAP as the COI and POI of reference picture.

Fourthly, remove the pictures that will not be referenced from buffer. The total number of useless pictures and delta COI of useless picture list is set in the RCS of current picture.

$$COI\_useless[i] = COI - delta\_coi\_useless[i] \quad (7)$$

where *COI_useless* represents the actual COI of reference picture and *delta_coi_useless* denotes the delta COI of current picture. The picture will be removed from buffer if the COI of current picture equals *COI_uselss* and the COI of corresponding position in buffer is set -1, indicating that position in buffer is empty and can be set for other reference pictures.

3.3. Decoding process

There are six steps to set up reference pictures list and manage reference pictures as follows.

Firstly, initialize the variable of reference related. The current random access point (*currRAP*) and the last reconstructed picture are set 0. And decode sequence header and then set next random access point (*nextRAP*).

Secondly, derive the actual COI. Decode the syntax element *coi* in picture header. And if the *coi* is smaller than the *coi* of last decoded picture, add 256 to current *coi*. And the actual COI of current picture equals.

$$COI = coi + COI\_Offset \quad (8)$$

Thirdly, derive the RCS of current picture. Decode the syntax element *use_rcs_flag* from the header of current picture. And if *use_rcs_flag* is 1 then decode the *rcs_index* and locate the RCS that the sequence header points at. Otherwise, decode the RCS from the picture header.

Fourthly, derive the POI of current decoding picture. The *delta_poi* is derived from the RCS. And the actual POI is derived as follows.

$$POI = COI - delta\_poi \quad (9)$$

The *currRAP* will be set as the value of *nextRAP* if POI is larger than the *currRAP*.

Fifthly, set up reference picture list. The pictures in the buffer that satisfy (6) are selected as reference pictures of current coding picture.

Finally, derive useless picture information from buffer and delete from buffer. A picture in buffer is considered useless if it is not referenced by other pictures and needed to be removed from DPB. Firstly set the number of useless pictures *num_of_useless*, which is derived from RCS of current picture. Secondly derive the COI of useless reference picture.

$$COI\_useless = COI - delta\_coi\_useless \quad (10)$$

where *COI_useless* represents the coding order of useless picture in buffer. The variable *delta_coi_useless* could be derived from RCS.

3.4. Writing reconstructed file order

After the encoder or decoder finishes coding or decoding a picture, the motion vector and reconstructed current picture will be stored in buffer if the current picture will be referenced by other pictures, which is signaled by the flag *refered_by_others* in RCS. Meanwhile, the encoder and decoder will check whether to write the reconstructed picture that stored in buffer into file and clean buffer. An output flag *last_output* is used to signal the POI of previous output reconstructed picture. And search buffer to derive the picture that the POI of which is right after current output picture and then write into reconstruction file.

4. EXPERIMENTAL RESULTS

To validate the flexibility of the proposed reference management strategy, we incorporate the proposed scheme into AVS reference software RD4.0 [12].

Firstly, the experiments are conducted for hierarchical B coding structure with GOP size set 16. The results are as shown as in Table 1. Compared with the RD4.0 anchor, the performance gain of the proposed reference management scheme can be up to 10.8%, 26.7% and 27.7% on Y, U and V respectively. And the average performance gain is 5.6%, 14.4% and 14.0% respectively.

Secondly, a hierarchical P coding structure is designed and configured to improve coding efficiency. Traditionally, the QP of P pictures are set the same and four nearest previous constructed pictures are selected as reference pictures. In the proposed method, four pictures are configured as a GOP and pictures of different positions of a GOP are coded with different QP. As can be seen from Fig. 4, the picture that signaled with the black color uses the smallest QP and is referred the most times. The results are as listed in Table 2. It can be seen that, compared with the RD4.0 anchor, it can be seen that the performance gain of the proposed hierarchical P can be up to 9.6%, 9.5% and 9.7% on Y, U and V respectively. And the average performance gain is 4.3%, 7.0% and 7.3% respectively.

Table 1. Performance comparisons between the proposed scheme of hierarchical B with GOP size set 16 and RD4.0 anchor

Sequences	Y	U	V
pku_girls	-9.3%	-16.7%	-16.4%
Pku_parkwalk	-9.2%	-11.2%	-12.4%
Traffic	-5.2%	-16.2%	-16.9%
Cactus	-4.8%	-13.6%	-14.3%
BasketballDrive	-0.8%	-11.2%	-7.0%
BasketballDrill	-9.8%	-15.0%	-13.5%
BQMall	-4.3%	-16.0%	-14.1%
PartyScene	-10.8%	-20.1%	-20.9%
RaceHorsesC	1.4%	-7.8%	-5.8%
BasketBallPass	-1.5%	-9.9%	-2.8%
BQSquare	-9.0%	-26.7%	-27.7%
BlowingBubbles	-9.4%	-19.3%	-21.2%
RaceHorses	0.4%	-6.8%	-6.5%
City	-8.8%	-17.9%	-19.4%
Crew	0.4%	-4.9%	-6.4%
Vidyo1	-7.0%	-14.9%	-15.1%
Vidyo3	-8.2%	-16.7%	-18.3%
Average	-5.6%	-14.4%	-14.0%

Table 2. Performance comparisons between the proposed scheme of hierarchical P and RD4.0 anchor for low delay P

Sequences	Y	U	V
pku_girls	-5.0%	-8.0%	-8.3%
Pku_parkwalk	-7.1%	-11.3%	-13.9%
Traffic	-5.3%	-6.8%	-6.8%
Sunflower	-4.7%	-8.3%	-8.9%
Pedestrian_area	-5.1%	-6.9%	-5.3%
Kimono1	-2.0%	-3.5%	-3.8%
Cactus	-7.6%	-9.3%	-11.0%
BasketballDrive	0.0%	-2.6%	-1.9%
BasketballDrill	-9.6%	-9.5%	-9.7%
BQMall	-2.2%	-5.0%	-4.5%
PartyScene	-7.9%	-11.4%	-12.4%
RaceHorsesC	-0.8%	-1.8%	-1.9%
BasketBallPass	-2.5%	-3.6%	-3.5%
BQSquare	-7.3%	-21.0%	-19.6%
BlowingBubbles	-7.3%	-10.1%	-10.9%
RaceHorses	-1.5%	-2.2%	-2.7%
City	-2.2%	-6.5%	-5.2%
Crew	-0.2%	-1.4%	-2.0%
Vidyo1	-4.2%	-7.2%	-5.9%
Vidyo3	-3.3%	-4.7%	-6.9%
Average	-4.3%	-7.0%	-7.3%

Finally, to validate the flexibility of the proposed reference picture management scheme, a GOP size 24 of hierarchical B coding structure is configured. The results are given in Table 3. The performance gain of the proposed hierarchical P can be up to 9.6%, 9.5% and 9.7% on Y, U and V respectively. And the average performance gain is 4.3%, 7.0% and 7.3% respectively.

In order to further illustrate the efficiency of the proposed scheme, the RD-curves are provided in Fig. 5, Fig. 6 and Fig.7. Fig. 5 and Fig. 7 show the RD-curve comparisons for random access between the anchor and the proposed scheme with GOP size 16 and 24. It can be observed that the proposed scheme can achieve better performance under low and high bit rate. Meanwhile, for low delay test cases, the proposed scheme can also obtain better performance as illustrated in Fig. 7.

Table 3. Performance comparisons between the proposed scheme of hierarchical B with GOP size set 24 and RD4.0 anchor

Sequences	Y	U	V
Sunflower	-0.5%	-11.1%	-12.3%
Pedestrian_area	-3.7%	-9.0%	-5.9%
Kimono1	-1.0%	-9.9%	-12.5%
Average	-1.7%	-10.0%	-10.2%

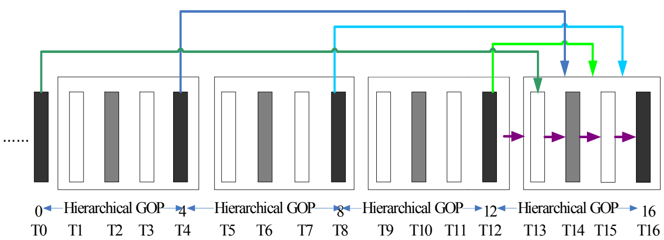
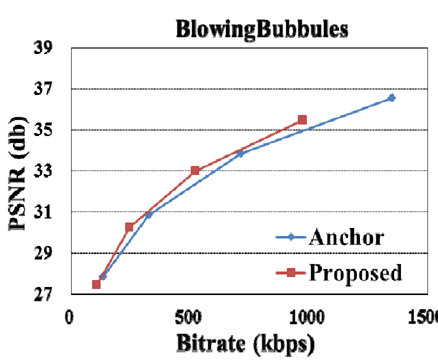


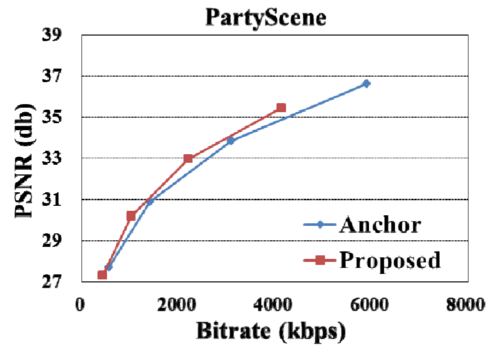
Fig.4. Hierarchical P coding structure

5. CONCLUSION

In this paper, a flexible reference management scheme is proposed, which can satisfy the needs for different sizes of GOP and can vary to adapt different sceneries. Through decoding the reference information transmitted in bit stream, the decoder side can set up reference list and manage reference pictures and DPB efficiently. Experimental results demonstrate that the proposed scheme can get better performance in low delay P and random access configurations.

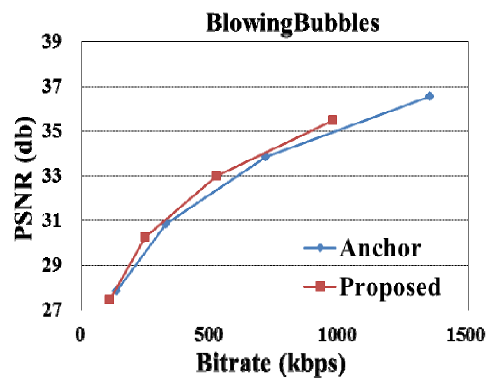


(a) BlowingBubbles_416x240_50

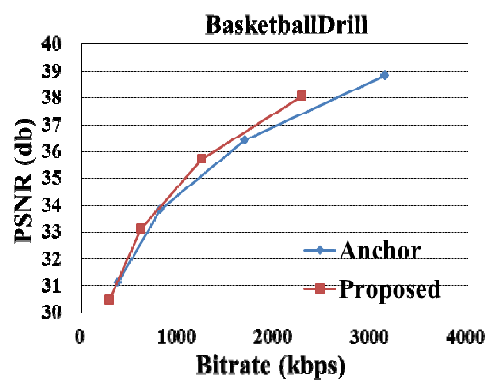


(b) PartyScene_823x480_50

Fig.5. RD curves of proposed scheme for hierarchical B coding structure in GOP size 16



(a) BlowingBubbes_416x240_50



(b) BasketballDrill_832x480_50

Fig.6. RD curves of proposed reference management scheme for hierarchical P coding structure in GOP size 4

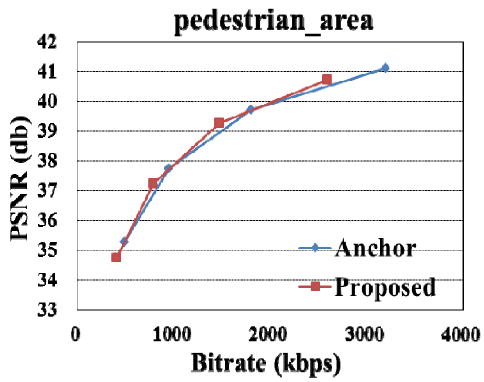


Fig.7. RD curves of proposed reference management scheme for hierarchical B coding structure in GOP size 24

6. ACKNOWLEDGMENT

This research is supported by the National Hightech R&D Program of China (863 Program, 2012AA010805), the National Science Foundation of China (61322106, 61103088, 61379100 and 61001108), which are gratefully acknowledged.

7. REFERENCES

[1] AVS Working Group Website: http://www.avs.org.cn.
[2] GB/T20090.2 information technology- Advanced Audio Video Coding Standard Part 2: Video, 2006.
[3] Draft of Second Edition of Audio Video Coding – Part 2: Video, AVS_N2019, 2014.
[4] AVS2-P2 common test condition, AVS_N2020, 2014.
[5] S. Ma, S. Wang and W. Gao, "Overview of IEEE 1857 Video Coding Standard", International Conference on Image Processing, September 15-18, 2013, Melbourne, Australia.
[6] Draft of Second Edition of Audio Video Coding – Part 2: Video, AVS_N1975, 2013.
[7] Z. He, Y. Ling, L. Yu, S. Lee, I. Kim, Y. Piao, "A reference picture management scheme based on reference layer ", AVS Doc. AVS-M3109, 2013.
[8] X. Zhang, S. Dong, P. Xing, T. Huang, "The reference picture optimization algorithm in AVS2 IPPP coding structure", AVS Doc. AVS-M2997, 2013.
[9] Q. Yu, S. Ma, K. Zhang, J. An, S. Lei, "A flexible configurable reference management algorithm", AVS Doc. AVS-M3003, 2013.
[10] Q. Yu, X. Zhang, S. Dong, S. Ma, K. Zhang, J. An, S. Lei, "A configurable flexible reference management scheme and the reference picture optimization algorithm in AVS2 IPPP coding structure", AVS Doc. AVS-M3052, 2013.
[11] Q. Yu, X. Jiang, X. Zhang, S. Dong, S. Ma, K. Zhang, J. An, S. Lei, "CE: A configurable flexible reference management scheme and the reference picture optimization algorithm in AVS2 IPPP coding structure", AVS Doc. AVS-M3178.
[12] RD Software Ver. 4.0, RD4.0 Available: ftp://124.207.250.92//incoming/video_codec/AVS2_P2/RD4.0.zip.