May 2006, Vol.21, No.3, pp.310–314 J. Comput. Sci. & Technol.

Performance Comparison of AVS and H.264/AVC Video Coding Standards

Xin-Fu Wang (王新夫) and De-Bin Zhao (赵德斌)

Department of Computer Science, Harbin Institute of Technology, Harbin 150001, P.R. China

E-mail: {xfwang, dbzhao}@jdl.ac.cn

Received October 18, 2005; revised March 14, 2006.

Abstract A new audio and video compression standard of China, known as advanced Audio Video coding Standard (AVS), is emerging. This standard provides a technical solution for many applications within the information industry such as digital broadcast, high-density laser-digital storage media, and so on. The basic part of AVS, AVS1-P2, targets standard definition (SD) and high definition (HD) format video compression, and aims to achieve similar coding efficiency as H.264/AVC but with lower computational complexity. In this paper, we first briefly describe the major coding tools in AVS1-P2, and then perform the coding efficiency comparison between AVS1-P2 Jizhun profile and H.264/AVC main profile. The experimental results show that the AVS1-P2 Jizhun profile has an average of 2.96% efficiency loss relative to H.264/AVC main profile in terms of bit-rate saving on HD progressive-scan sequences, and an average of 28.52% coding loss on interlace-scan sequences. Nevertheless, AVS1-P2 possesses a valuable feature of lower computational complexity.

Keywords video coding, image compression, AVS, H.264/AVC

1 Introduction

The AVS standard is the newest compression standard of China. It is developed by the AVS working group, which was approved and established by the Science and Technology Department of Ministry of Information Industry of China in June 2002. The mandate of the working group is to establish the standards for compression, decompression, manipulation and display in digital audio and video multimedia equipments and systems. Its major standardization fields include system, audio, video and digital copyright management. AVS1-P2 video coding standard[1], as one part of AVS-video, is the earliest and most important part of the AVS standard. The project of AVS1-P2 is started in June 2002, with the target to establish a standard that can achieve similar coding efficiency as H.264/AVC with much lower computational complexity on SD and HD videos. After one and a half years' hard work, the draft design of AVS1-P2 was finalized for formal approval submission in December 2003. AVS1-P2 adopts the classical block-based prediction with transform coding framework. With advanced coding tools and low complexity consideration, AVS1-P2 video coding standard supplies high coding efficiency for SD and HD video compression with low computational complexity and low storage requirement.

H.264/AVC[2] is the newest video coding standard of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). It has been proved to outperform all of the previous video coding standards in terms of coding efficiency. To demonstrate the performance of AVS1-P2, it is necessary to compare it with H.264/AVC.

This paper is organized as follows. Section 2 gives a brief description of the key techniques in AVS1-P2. Section 3 shows the coding efficiency comparison between AVS1-P2 and H.264/AVC video coding standards with the comprehensive experimental results. An analysis of performance gap between AVS1-P2 and H.264/AVC on interlace-scan sequences is also provided in this section. At last the paper is concluded in Section 4.

2 Key Techniques of AVS1-P2 Video Coding Standard

Similar to many previous video coding standards such as MPEG-1, MPEG-2 and H.263, AVS1-P2 also employs the so-called block-based hybrid video coding scheme, which has been demonstrated to be a very successful video compression framework. In this framework, the input video signal is first split into macroblocks and then each macroblock is processed and coded. To acquire better tradeoff between coding efficiency and complexity, advanced coding tools for intra- and inter-prediction, transform, quantization, entropy coding and deblocking filter are adopted by AVS1-P2. They are briefly described as follows.

- *Intra prediction.* In AVS1-P2, only 8×8 block-based intra prediction[3] is used for both luma and chroma components of the regions that are coded as intra. Intra prediction employs the technique of directional spatial prediction, which means prediction is always conducted in the spatial domain by referring to neighboring samples (without being deblocking filtered) of already coded blocks. For intra luma prediction, four directional prediction modes are specified namely vertical, horizontal, down_left and down_right. Except these four direction-based modes, another prediction mode named DC is also designed for luma component. The samples used for DC

Short Paper

Supported by the National Natural Science Foundation of China under Grant No. 60333020 and the Natural Science Foundation of Beijing under Grant No. 4041003.

prediction mode and four directional prediction modes are shown in Fig.1. Intra chroma prediction enables four modes including vertical, horizontal, DC and plane modes. It should be noticed that the neighboring samples used in the DC prediction mode, down_left mode and down_right mode are first made “softer” by applying low pass filtering with the tap values (1/4, 2/4, 1/4) before directional prediction.

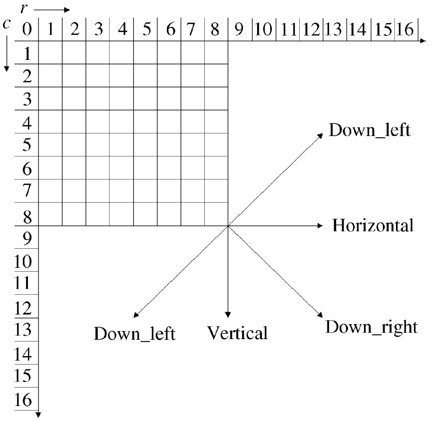

Fig.1. Samples used for DC prediction mode and four directional prediction modes for luma component.

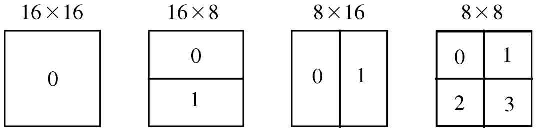

Fig.2. Segmentations of the macroblock for motion compensation.

- *Variable block-size motion compensation.* AVS1-P2 supports multiple block size inter-frame prediction with block size from 16×16 to 8×8 to better exploit inter-frame redundancy. Our experiments have shown that, for most HD sequences, block sizes smaller than 8×8 have little contribution on inter-frame prediction. So AVS1-P2 selects 8×8 as the smallest block size, which also can reduce encoding and decoding complexity. The segmentations of the macroblock of AVS1-P2 for motion compensation are illustrated in Fig.2.

- *Quarter-sample-accurate motion compensation.* In AVS1-P2, motion compensation is performed in quarter-pixel accuracy. A 1/4-pixel accuracy 2-D separatable interpolation method named Two Steps Four Taps interpolation (TSFT)[4] is adopted in AVS1-P2. The fractional samples interpolation is illustrated in Fig.3. According to TSFT, the luminance prediction values at half-pixel locations b, h are interpolated by applying a 4-tap cubic convolution interpolation filter with tap values ($-1/8$, $5/8$, $5/8$, $-1/8$) on the values at integer-pixel positions, and j is interpolated by applying the filter on the values at half-pixel locations. The prediction values at quarter-pixel locations a, c, d, f, i, k, n and q are achieved by applying a 4-tap cubic spline filter with tap values (1/16, 7/16, 7/16, 1/16) on the values at integer-pixel and half-pixel locations. The prediction values at quarter-pixel locations e, g, p, and r are achieved by applying a bi-linear filter on the values at half-pixel poison j and at integer positions D, E, H, I, respectively. Compared with the method used in H.264/AVC, this scheme reduces spatial complexity and computational complexity, and a little more average SNR is gained on HD format sequences.

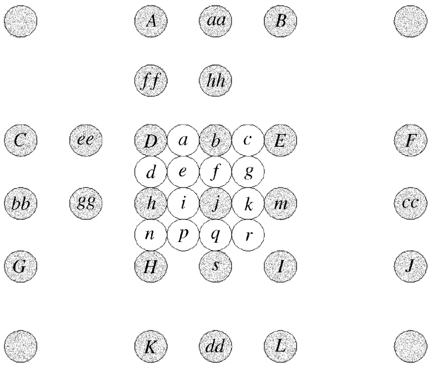

Fig.3. Filtering for fractional-sample accurate motion compensation. Upper-case letters indicate samples on the full-sample positions, while lower case samples indicate samples at fractional-sample positions.

- *Multiple reference picture motion compensation.* Multi-frame motion-compensated prediction is supported in AVS1-P2. The maximal number of reference frames is restricted to two to reduce the storage requirement and the computational complexity. P frames can make use of at most two previous pictures for motion search. In case of bi-predictive frames, no more than one forward picture and one backward picture are used for motion search.

- *Motion compensation modes.* In AVS1-P2, there are five motion-compensated macroblock modes for P-slice, which are skip, 16×16, 16×8, 8×16, and 8×8 modes. For B slices, five different types of motion compensation modes are supported: forward, backward, symmetry-based bi-predictive, skip, and direct. For the symmetric prediction mode[5], only forward motion vectors are coded for each block, and the backward motion vectors can be derived from the forward ones. The direct mode used in AVS1-P2[5] is different from that in H.264/AVC. In H.264/AVC, when the co-located block in the backward reference picture is coded as intra mode, the corresponding motion vectors for scaling will be set to zero. However, this method is inefficient in most cases. While the direct mode used in AVS1-P2 solves the problem by Spatial Direct Mode technique. Motion vector prediction for differential motion vector coding exploits the motion information of the spatially adjacent blocks and further reduces the correlation with adjacent blocks.

- *Transform and quantization.* Similar to previous video coding standards, AVS1-P2 also utilizes transform coding for the prediction residuals. An 8×8 integer DCT transform technique named Pre-Scaled Integer Transform (PIT)[6], which can be implemented by simple addition and shift operations, is designed. Fig.4 illustrates the conventional transform scheme in H.264. The scaling operations are needed on both encoder and decoder sides. While with PIT, the inverse scaling is moved to encoder side (as depicted in Fig.5) and combined with forward scaling and quantization as one single process, so the dequantization matrix is not needed to be stored in decoder side. With PIT, the storage of the expanded matrix and the look-up operations required in the implementation of conventional Integer Cosine Transform (ICT) are prevented. As a result, the spatial complexity and computational complexity of decoder side are further reduced compared to conventional ICT. For the quantization of transform coefficients, scalar quantization is applied. One of 64 quantization levels with a QP period of approximate 8 can be selected.

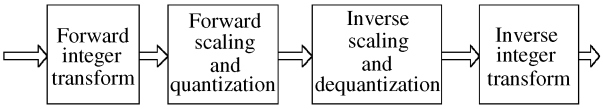

Fig.4. Block diagram of ICT in H.264.

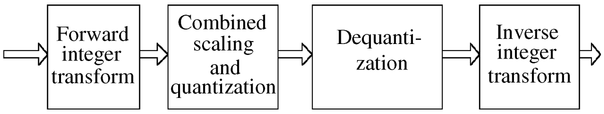

Fig.5. Block diagram of ICT in AVS1-P2.

- *Entropy coding.* In AVS1-P2, all syntax elements except the quantized transform coefficients are coded with Exponential-Golomb codes or fixed-length codes. For quantized transform coefficients, a more sophisticated method called context-based 2D-VLC is employed[7]. In this method, three groups of multiple tables are used for inter_luma, intra_luma and chroma coefficients, respectively. Each of multiple 2D-VLC tables are designed and used context-adaptively to better match different level/run combination's probability distribution when a block is coded. Another key feature is that each 2D-VLC table's codewords are constructed based on Exponential-Golomb codes, which keep the multiple tables with low memory requirement. This method achieves higher coding efficiency than one-table-for-one-block coding method.
- *In-loop deblocking filtering.* Blocking artifacts are inherent shortcoming of block-based video coding. These artifacts originate from both the prediction and quantization process. They can heavily degrade the subjective quality of pictures and decline the coding efficiency. In AVS1-P2, an 8 × 8 block-based adaptive in-loop deblocking filter is designed to remove the artifacts. It achieves lower computational complexity than that used in H.264/AVC due to fewer boundaries, fewer boundary strengths and simpler judgment process.
- *Interlace coding tools.* For interlace-scan sequences, a set of interlace coding tools are provided in AVS1-P2. One frame can be coded as one unit or be split into two parts and then coded separately. To reduce complexity, only picture-adaptive frame/field coding is supported.
- *Profile and level.* In AVS1-P2, one profile named Jizhun profile is defined. There are five levels for this profile. Each level specifies upper limits for the picture size, the maximal video bit-rate, the BBV buffer size and so on. Level 2.0, 4.0, 4.2, 6.0 and 6.2 target streaming video, SD video, SD video with 4 : 2 : 2, HD video and HD video with 4 : 2 : 2, respectively.

3 Performance Comparison

Although some initial performance test results have been reported by the AVS working group during the development of AVS1-P2, independent performance tests are still necessary for performance analysis and comparison. In this section, we compare the performance of AVS1-P2 Jizhun profile with H.264/AVC main profile in terms of rate-distortion efficiency, since they have similar application areas. We make use of the RM5.2c reference encoder[8] to generate the AVS1-P2 coding results and the JM7.6 test model encoder[9] to generate the H.264/AVC coding results.

To provide sufficient and comprehensive experimental results, three sets of sequences are selected for the comparison. All of these sequences are original. They are provided by the Joint Video Team (JVT) and most of them are used during the development of H.264/AVC. The first set (Set-1) consists of five interlace-scan SD sequences at the resolutions of 720 × 576 pixels (25Hz) including Basketball, FlowerGarden, MobileCalendar, Src3, and Src5. The second one (Set-2) consists of five progressive-scan HD sequences at the resolutions of 1920 × 1080 pixels (25Hz) including PedestrianArea, RushHour, Station2, Toys & Calendar, and VintageCar. The third one (Set-3) consists of two interlace-scan HD sequences at the resolutions of 1920 × 1080 pixels (30Hz) including Kayak and Flamingo. 200 frames of each video sequence of Set-1 and Set-2 and 60 frames of each video sequence of Set-3 are used in our experiments.

Two kinds of experiments are performed. The first one is used for the general performance comparison of AVS1-P2 Jizhun profile and H.264/AVC main profile. The second one is used to analyze the performance gap between AVS1-P2 and H.264/AVC on interlace-scan sequences.

3.1 General Performance Comparison

To provide sufficient test results, all of the three sets of sequences are used for evaluating the general performance of AVS1-P2. The coding options, except interlace coding tools, entropy coding tools and the number of reference frames, are similar for both standards in the simulations. The detailed test conditions are listed in Table 1.

Figs. 6–9 show the rate-distortion curves for FlowerGarden, SRC5, PedestrianArea and Kayak, respectively. The rate-distortion curves are produced according to the experimental results and they provide qualitative comparison results on each sequence. The curves marked with H.264_bm5 in Figs. 6 and 7 are the rate-distortion curves for H.264/AVC according to the test results in this subsection.

The quantitative performance comparison is based on the output bit-rate and PSNR of each encoded video sequence. PSNR is used only for the luminance component. For each sequence, to evaluate the average PSNR vs. bit-rate, we employ the method described in [10], which is widely used during H.264/AVC development.

The simulation results are summarized in Table 2 that shows that AVS1-P2 is inferior to H.264/AVC by 2.96% on average in terms of bit-rate saving for progressive-scan sequences. Though H.264/AVC adopts many advanced techniques that make it much more complex than AVS1-P2, such as CABAC, five reference frames, and small block-size inter prediction, AVS1-P2 still acquires similar coding efficiency on progressive-scan HD sequences relative to H.264/AVC. For interlace-scan sequences, Table 2 shows that AVS1-P2 has an average of 28.52% efficiency loss in terms of bit-rate saving, or equally −1.04dB in terms of PSNR gain relative to H.264/AVC. The large coding performance gap will be analyzed in the following subsection.

Table 1. Test Conditions

	AVS1-P2	H.264/AVC
MV resolution	1/4 pel	1/4 pel
Hadamard	ON	ON
Search range	±32	±32
Reference frames	2	5
B frames inserted	2	2
Entropy coding	2D-VLC	CABAC
RD optimization	ON	ON
Interlace coding option	PAFF	MBAFF
Loop filter	ON	ON
Rate control	ON	ON
Bit rate for SD	1M, 1.5M, 2M, 2.5M	1M, 1.5M, 2M, 2.5M
Bit rate for HD	8M, 10M, 12M, 16M	8M, 10M, 12M, 16M

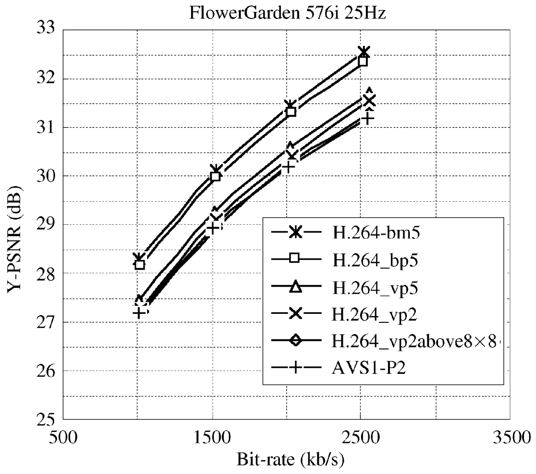

Fig.6. Rate-distortion curves of FlowerGarden sequence.

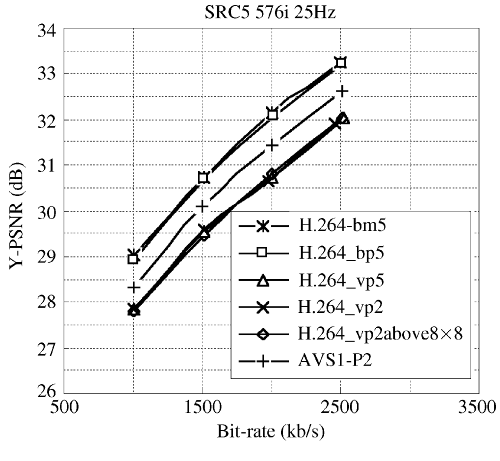

Fig.7. Rate-distortion curves of SRC5 sequence.

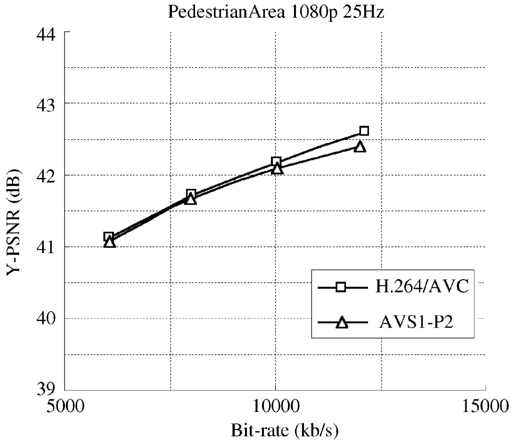

Fig.8. Rate-distortion curves of PedestrianArea sequence.

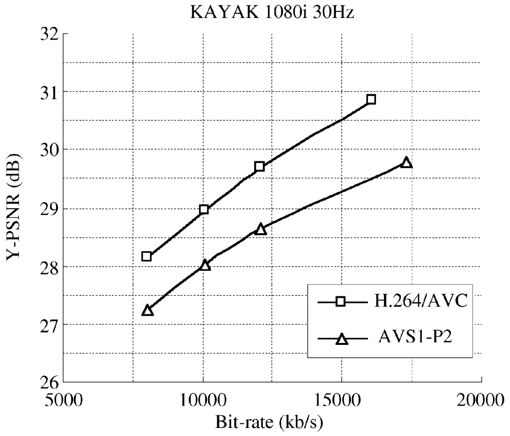

Fig.9. Rate-distortion curves of KAYAK sequence.

Table 2. Coding Efficiency Comparison of AVS1-P2 and H.264/AVC

Sequence	Ave PSNR gains (dB)	Ave Bit-rate savings
A:576i, 25Hz		
Basketball	−0.89	−19.97%
FlowerGarden	−1.18	−30.15%
MobileCalendar	−2.01	−61.77%
SRC3	−0.88	−26.47%
SRC5	−0.65	−14.82%
B:1080p, 25Hz		
PedestrianArea	−0.07	−3.32%
RushHour	−0.19	−13.78%
Station2	0.17	9.94%
Toys&Calendar	−0.13	−9.12%
VintageCar	0.03	1.50%
C:1080i, 30Hz		
Kayak	−1.04	−33.00%
Flamingo	−0.62	−13.49%

3.2 Analysis of Coding Performance Gap Between AVS1-P2 and H.264/AVC on Interlace-Scan Sequences

To study the performance gap mentioned in previous subsection, another four experiments are designed. Four coding tools namely entropy coding tool, interlace coding tool, multi-reference picture motion compensation and small block-size inter prediction are investigated in the simulations. These coding tools most probably cause the performance gap. The four experiments are named as H.264_bp5, H.264_vp5, H.264_vp2, and H.264_vp2A8 respectively. The experiment for H.264/AVC in Subsection 3.1 is named as H.264_bm5. All the coding options of the four experiments except that listed in Table 3 are the same as that in Table 1. Only the sequences in Set-1 are used in these experiments.

Table 3. Different Test Conditions of Experiments H.264_bp5, H.264_vp5, H.264_vp2, and H.264-vp2A8

	H.264_bp5	H.264_vp5	H.264_vp2	H.264_vp2A8
Entropy coding	CABAC	CAVLC	CAVLC	CAVLC
Interlace coding	PAFF	PAFF	PAFF	PAFF
Reference frames	5	5	2	2
Smallest block-size for ME/MC	4 × 4	4 × 4	4 × 4	8 × 8

The rate-distortion curves in Figs. 6 and 7 show qualitative comparison results on FlowerGarden and SRC5 sequence, respectively. The quantitative simulation results are summarized in Table 4. The comparison between each pair of experiments can be found in this table. The values on the diagonal positions of Table 4 provide the comparison results of single coding tool, and they indicate that MBAFF can achieve an average of 2.32% coding gains over PAFF, CABAC can achieve better performance than CAVLC by 16.93% on average, five reference frames can achieve an average of 4.60% coding gains over two reference frames, inter prediction with block-size as small as 4 × 4 can achieve an average of 2.80% coding gains over inter prediction with block size as small as 8 × 8 and the performance gap between AVS1-P2 and H.264_vp2A8 is about 0.56% in terms of bit-rate saving. So we come to the conclusion that AVS1-P2 can achieve similar coding performance with H.264/AVC on interlace-scan sequences in case they work on similar complexity level. The performance gap shown in previous subsection due to the lack of the complex coding tools such as CABAC, PAFF, five reference pictures for ME/MC and inter-prediction with block-size as small as 4 × 4 adopted in H.264/AVC.

Table 4. Performance Comparison of the Six Experiments in Terms of Bit-Rate Saving

	Average bit-rate savings relative to:				
	H.264_bp5	H.264_vp5	H.264_vp2	H.264_vp2A8	AVS1-P2
H.264 bm5	2.32%	19.57%	25.21%	–	30.64%
H.264_bp5	–	16.93%	22.25%	–	27.37%
H.264_vp5	–	–	4.60%	–	9.43%
H.264_vp2	–	–	–	2.80%	4.17%
H.264_vp2A8	–	–	–	–	0.56%

4 Conclusion

In this paper we present a performance evaluation of the AVS1-P2 Jizhun profile and H.264/AVC main profile in terms of rate-distortion efficiency. The comprehensive experimental results show that AVS1-P2 Jizhun profile achieves similar coding efficiency to H.264/AVC main profile on progressive-scan HD sequences but with much lower computational complexity. While for HD and SD interlace-scan sequences AVS1-P2 Jizhun profile is inferior to H.264/AVC main profile by an average of 28.52% in terms of bit-rate saving. The performance gap is mainly caused by the lack of the coding tools such as CABAC, PAFF, five reference pictures for ME/MC, small block-size inter prediction used in H.264/AVC. On condition that these coding tools are shut off, which means AVS1-P2 and H.264/AVC work on similar complexity level, AVS1-P2 can achieve similar coding efficiency with H.264/AVC. In the next phase of AVS1-P2 (X profile), some advanced coding tools should be adopted for further coding efficiency improvement.

References

[1] Audio Video Coding Standard Workgroup of China (AVS). Video Coding Standard FCD1.0, Nov. 2003.

[2] Wiegand T, Sullivan G J, Luthra A. Draft ITU-T recommendation H.264 and final draft international standard 14496-10 advanced video coding. Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, Doc. JVT-G050rl, Geneva, Switzerland, May 2003.

[3] Zhang Nan, Yin Baocai *et al.* Spatial prediction based intra-coding. In *Proc. IEEE Int. Conf. Multimedia and Expo* (*ICME2004*), Taipei, June 2004, 1: 97–100.

[4] Ronggang Wang, Chao Huang *et al.* Sub-pixel motion compensation interpolation filter in AVS. In *Proc. IEEE Int. Conf. Multimedia and Expo* (*ICME2004*), Taipei, June 2004, 1: 93–96.

[5] Xiangyang Ji, Debin Zhao *et al.* New bi-prediction techniques for B pictures coding. In *Proc. IEEE Int. Conf. Multimedia and Expo* (*ICME2004*), Taipei, June 2004, 1: 101–104.

[6] Ci-Xun Zhang, Jian Lou, Lu Yu *et al.* The technique of pre-scaled integer transform. In *Proc. IEEE 2005 International Symposium on Circuits and Systems* (*ISCAS2005*), Kobe, Japan, May 2005, 1: 316–319.

[7] Qiang Wang, Debin Zhao, Siwei Ma *et al.* Context-based 2D-VLC for video coding. In *Proc. IEEE Int. Conf. Multimedia and Expo* (*ICME2004*), Taipei, June 2004, 1: 89–92.

[8] Audio video coding reference software. ftp://159.226.42.57/public/ avs_doc/avs_software.

[9] Joint video team reference software, version 76. http://bs.hhi.de/∼suehring/tml/download/.

[10] Bjontegaard G. Calculation of average PSNR differences between RD-Curves. Doc. VCEG-M33, Austin, Texas, USA, Mar. 2001.