THIRD GENERATION AVS VIDEO CODING STANDARD (AVS3) : FEATURES AND APPLICATIONS

ABSTRACT:
This paper presents a comprehensive introduction of AVS3, which is developed by Audio and Video Coding Standard (AVS) Workgroup of China and is published as GY/T 368 by NRTA in 2023. In order to improve coding efficiency, AVS3 adopted many novel technologies including block partition, intra/inter prediction transform, quantization and entropy coding and has 50% bitrate savings compared with its predecessor AVS2. AVS3 has been used in China's 8K UHDTV channels and will have more applications worldwide in the future.

1. INTRODUCTION

In recent years, China has rapid development in UHD video industry. With higher resolution, framerate, bit depth, dynamic range and wider color gamut, the data amount of UHD video has increased largely. In order to meet the demand of 4K UHD video services, NRTA published the second generation AVS video coding standard (AVS2) in 2016 *[See GY/T 299.1–2016; High efficiency coding of audio and video Part 1: Video, National Radio and Television Administration, 2016]*. AVS2 has similar coding efficiency with H.265 *[See ITU-T H.265; High efficiency video coding, International Telecommunication Union, 2013]* and is adopted for 4K UHD video coding in China with the bitrate of 36Mbps under the configuration of 3840×2160@50p, 10bit and HDR. However, 8K UHD video brings more challenges and it is necessary to develop new video coding standard to further improve the coding efficiency.

AVS3's standardization process was started in December 2017. The aim of AVS3 is to have 50% bitrate savings with AVS2 under the premise of the same video quality. To promote industrial application, the development of AVS3 is divided into two phases: In the first phase, the primary consideration for the selection of coding tools is the limited decoder complexity, and it was finalized in July 2019 with two year's efforts. In the second phase, it is aimed to future improve coding efficiency based on the first stage. Finally, AVS3 was finalized in 2022. Through the comprehensive evaluation and testing procedure, AVS3 has been adopted as broadcasting industry standard in China and published by NRTA in March 2023 *[See GY/T 368–2023; Advanced and efficient video coding, National Radio and Television Administration, 2023]*. This paper gives an overview of AVS3 standard including key technologies and its application in China.

2. KEY CODING TOOLS OF AVS3

AVS3 adopts block-based hybrid video coding framework as shown in ***Figure 1***. A flexible block partition mechanism is put forward to adapt to higher resolution and novel intra/inter prediction, transformation, quantization, and entropy coding tools are also introduced to highly improve coding performance.

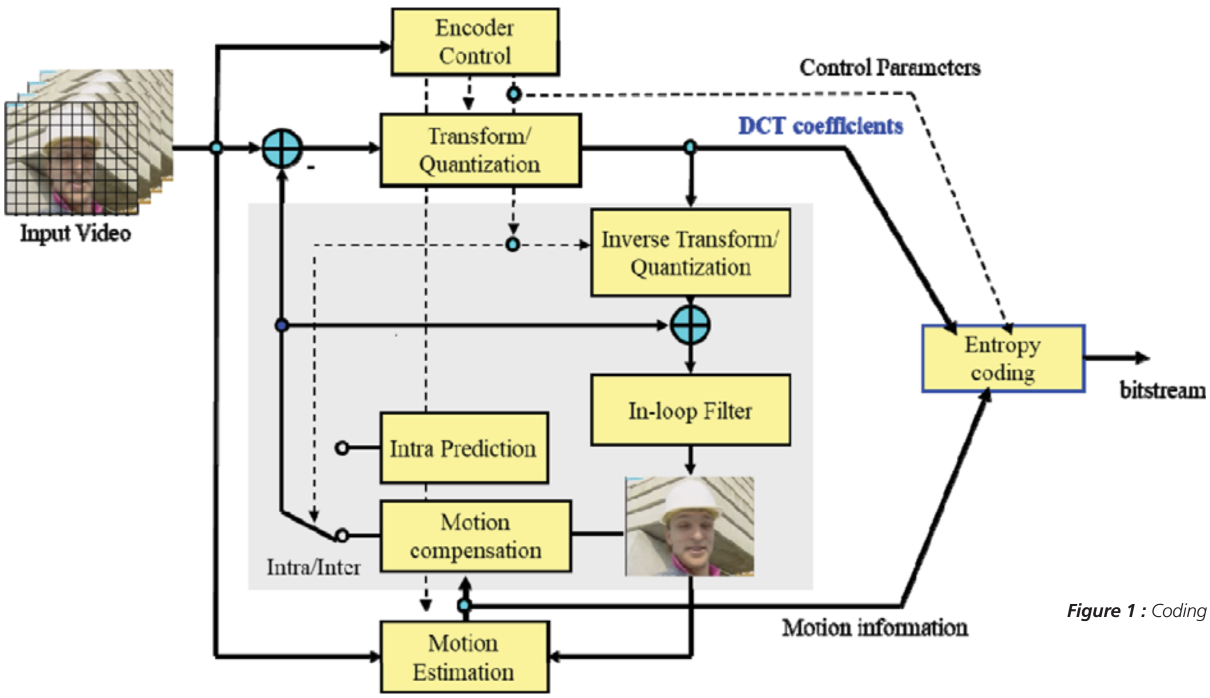

Figure 1 : *Coding framework of AVS3*

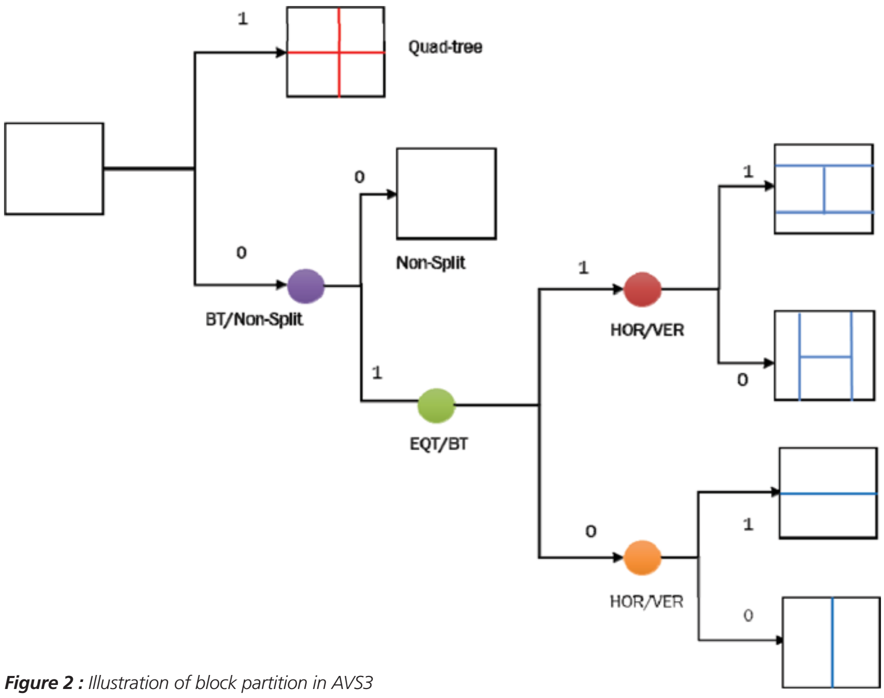

Figure 2 : *Illustration of block partition in AVS3*

2.1 FLEXIBLE BLOCK PARTITION STRUCTURE

AVS3 designs a more flexible block partition structure to adapt for different video resolutions. The coding picture is firstly divided into patches and then are divided into largest coding unit (LCU), the largest size for LCU is 128x128 which is more applicable for UHD video. LCU can be recursively split into coding units (CU) based on partition tree structure. AVS3 adopts quad-tree (QT), binary tree (BT) and extended quad-tree (EQT) partition to decide the final size of CU as shown in ***Figure 2***.

For each LCU, one bit is first used to indicate whether the LCU is divided into QT or not. If not, one more bit is then used to represent whether the LCU is divided into EQT or BT. Through this flexible partition process, CU can be non-square.

2.2 INTRA PREDICTION

AVS3 adopts 66 luma intra prediction including 62 angles prediction modes, DC mode, planar mode, PCM mode and bilinear mode. AVS3 introduces Intra Derived Tree (Intra-DT), intra prediction filter (IPF) and two step cross-component prediction mode (TSCPM) to improve intra prediction.

Intra Derived Tree defines intra coding unit partitioning mechanism vertically or horizontally. The current CU is divided into two or four prediction units (PU) when the current CU apply to intra-DT. Intra-DT has 6 split types.

IPF is a coding tool to improve the accuracy of the intra predicted pixels. The predicted value near the top and left boundary of current CU is filtered to enhance the prediction accuracy.

TSCPM is designed for intra chroma coding, which assumes the linear correlation between the prediction results of luma and chroma components. The linear regression function is first applied to the originally reconstructed luma block to get an internal prediction block. Second, the internal prediction block is down-sampled to generate the final chroma prediction block.

2.3 INTER PREDICTION

In AVS3, Affine prediction mode considers more complex motion instead of translational motion. There are two types of affine motion prediction, one is affine direct mode while the other is affine inter mode. In affine direct mode, an affine candidate list including at most 5 affine models is derived from the motion information of neighboring blocks, an affine model candidate is a 4-parameter affine model or 6-parameter affine model, represented by 2 or 3 motion vectors. When the CU is affine inter mode, it is 4-parameter affine motion model.

Adaptive Motion Vector Resolution (AMVR) is adopted to reduce the coding bits for the motion vector difference (MVD) coding. The MVD is usually coded with 1/4-pel precision. While AMVR allows various resolutions for MVD coding. For instance, 1/4–pel, 1/2–pel, 1–pel, 2–pel and 4–pel resolution are supported.

Ultimate motion vector expression (UMVE) technology is an effective way to express motion vector for skip and direct modes in AVS3. According to the base motion vectors derived from the information of spatial or temporal neighboring candidates, this technique offsets these motion vectors to obtain the motion vectors that are more suitable for the current coding blocks.

Based on the above HMVP and AMVR, an enhanced motion vector resolution (EMVR) mode is carefully designed to avoid the shortcoming of one motion vector prediction (MVP) compared to multiple MVP. EMVR is proposed to bind the AMVR and HMVP, which corresponds the existing AMVR index of current CU with the motion information in HMVP table, and then conducted inter-frame prediction.

Symmetry motion vector difference (SMVD) is used for motion vector difference in inter Bi-directional prediction mode. One bit is used to indicate whether SMVD is adopted when the prediction mode of current CU is Bi-directional prediction mode. When SMVD is used, the motion information in List1 is symmetrical with respect to List0.

2.4 TRANSFORM

Position based transform (PBT) is introduced in AVS3, DST-VII and DCT-VIII are added to transform module. The selection between DCT-II, DCT-VIII and DST-VII are depended on signaling bit and intra block's position. AVS3 introduces a new transform method which uses pre-designed transform sets for four sub-blocks according to their positions. A new CU-level syntax element, PBT flag, is introduced. When PBT flag is 0, the traditional DCT-II transform is used for current CU. When PBT flag is 1, PBT splits the CU into four sub-blocks, and use pre-designed transform set according to their position. In each sub-block, its position number relates to a pre-designed transform set.

2.5 IN-LOOP FILTERING

Enhanced sample adaptive offset (ESAO) is applied to consider the textural and edge directional features. There are two feature descriptors in the sample classification process. The first one is calculated based on local binary patterns by comparing the intensities of the current sample with its 8 neighboring samples. The second descriptor is derived based on the amplitudes of the current reconstructed samples.

Cross-component sample adaptive offset (CCSAO) is only applied for chroma components. There are two feature descriptors in CCSAO classification process. The first one is based on the difference between the corresponding luma sample and its neighboring samples. The second descriptor is based on the amplitude of the current sample or its corresponding luma sample.

Besides, AVS3 improves adaptive loop filter (ALF), a larger filter shape is enabled with 15 coefficients and 29 taps. To make the filter more adaptive, the reconstructed frame is divided into 16 or 64 categories.

3. CODING PERFORMANCE

In order to evaluate the coding performance of AVS3, a subjective assessment between AVS3, AVS2 and Versatile Video Coding (VVC) was conducted for 8K testing sequences with the parameters of 7680×4320@50P, 10bit, BT.2020, HLG. And the coding bitrate is set 72Mbps, 144Mbps and 72Mbps for these three standards.

At the given encoding rate, AVS3 has 5.3% quality reduction compared with the original sequences, while AVS2 and VVC have 5.0% and 5.3% quality reduction separately. The testing results shows that AVS3 has similar coding performance with VVC and can reach 50% bitrate savings with AVS2.

TABLE 1 Subjective Assessment of AVS3

8K Test sequences	Quality Reduction		
	AVS3@72Mbps	AVS2@144Mbps	VVC@72Mbps
1	4.6%	3.6%	4.0%
2	6.7%	6.7%	7.7%
3	4.3%	3.5%	4.6%
4	5.7%	5.6%	4.1%
5	7.5%	6.2%	7.2%
6	3.4%	5.0%	5.6%
7	6.3%	5.5%	4.4%
8	3.8%	4.0%	5.0%
Average	5.3%	5.0%	5.3%

4. INDUSTRY APPLICATIONS

AVS3 has attracted tremendous attention from industry ever since the finalization of AVS3 Main Profile in March, 2019.The designation and performance-complexity trade-off naturally fulfill the demands and requirements of different use cases. AVS3 has deployed from industrial-grade 8K real-time software codecs to 8K decoder chip for smart-phone, from 8K UHD TV broadcasting service to worldwide large-scale events support.

Based on the Intel scalable video technology (SVT) architecture, a novel AVS3 8K real-time software encoder has been presented, namely SVT-AVS3 *[See H. Ren et al, SVT-AVS3: An Open-Source High-Performance AVS3 Encoder With Scalable Video Technology; IEEE Transactions on Multimedia, vol. 26, pp. 3291-3301, 2024]*. The SVT-AVS3 encoder achieves an optimal trade-off among coding performance, encoding speed, and perceptual visual quality.

AVS3 all-platform open-source decoder "MERAK" has been released. The decoder was developed by the team of Peking University Shenzhen Graduate School, code-named uavs3d, and supports up to 8K/60P real-time video decoding.

In September 2019, Hisilicon Inc. has officially released a new end-to-end solution for 8K contents based on the world's first AVS3 8K decoder chip Hi3796CV300 in International Broadcasting Convention (IBC), which greatly promotes the 8K applications of AVS3.

On February 1, 2021, the China Media Group launched an 8K UHD test channel using the AVS3 video standard. On January 1, 2022, the Beijing TV Winter Olympics Documentary Channel was broadcast using the AVS3 video standard. On January 25, 2022, the 8K UHD channel of China Media Group as broadcast using the AVS3 video standard.

In July 2022, AVS3 officially became one of the next-generation video codec standards in the DVB standard system *[See ETSI TS 101 154 V2.8.1; Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications, European Telecommunications Standards Institute, 2023]*, which means that AVS3 video standard, as an option of video encoding of international organizations, will be of great significance to supporting the future development of the global UHD video industry.

5. CONCLUSION

AVS3 is now opening a new era for 8K UHD video coding in which end-to-end solutions and commercial-level AVS3 codecs have been incorporated and established. Several directions are currently under investigation for future research and applications, including neural-network based video coding technology and machine-vision oriented video coding.

ABOUT AUTHOR

Zhou Yun, Senior Engineer, Academy of Broadcasting Science (ABS), National Radio and Television Administration (NRTA), China

Zhou Yun, received the B.S. and M.S. degree from Beijing University of Posts and Telecommunications. She is currently a senior engineer at Academy of Broadcasting Science, NRTA. Her research interests include video and audio coding, UHDTV and immersive media. ■