DOI: 10.3969/j. issn. 1673-5188. 2016. 01. 007
http://www.cnki.net/kcms/detail/34.1294.TN.20160122.1015.002.html, published online January 22, 2016

Introduction to AVS2 Scene Video Coding Techniques

Jiaying Yan[1,2,3], Siwei Dong[1,3], Yonghong Tian[1,3], and Tiejun Huang[1,3]
(1. National Engineering Laboratory for Video Technology, School of EE & CS, Peking University, Beijing 100871, China;
2. School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, Shenzhen 518055, China;
3. Cooperative Medianet Innovation Center, Beijing 100871, China)

Abstract

The second generation Audio Video Coding Standard (AVS2) is the most recent video coding standard. By introducing several new coding techniques, AVS2 can provide more efficient compression for scene videos such as surveillance videos, conference videos, etc. Due to the limited scenes, scene videos have great redundancy especially in background region. The new scene video coding techniques applied in AVS2 mainly focus on reducing redundancy in order to achieve higher compression. This paper introduces several important AVS2 scene video coding techniques. Experimental results show that with scene video coding tools, AVS2 can save nearly 40% BD-rate (Bjøntegaard-Delta bit-rate) on scene videos.

Keywords

AVS2; scene videos coding; background prediction

1 Introduction

The primary application of AVS2 is in ultrahigh-definition videos, especially scene videos. Scene videos are usually captured by stationary cameras and include videos from surveillance systems all over the world and from other applications, such as video conference, online teaching and remote medical. Scene videos have huge temporal and spatial redundancy for the background regions appear frequently and AVS2 can utilize the background information to compress the scene videos efficiently.

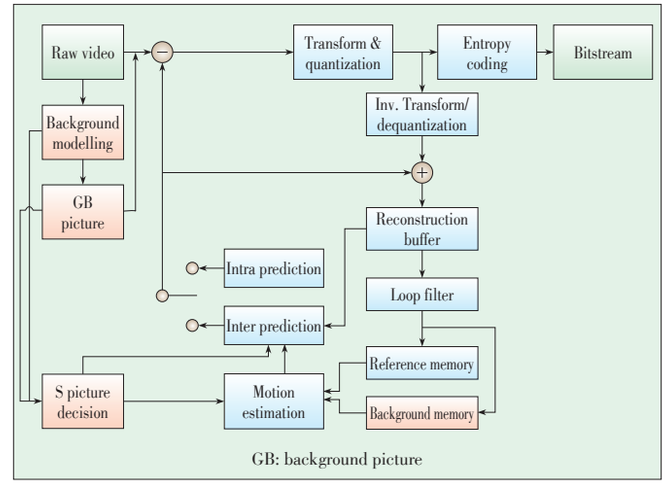
Similar to previous coding standards, AVS2 still adopts the classic block-based hybrid video framework. However, in order to improve coding efficiency, in the AVS2 coding framework, a more flexible coding unit (CU), prediction unit (PU) and transform unit (TU) based structure is adopted to represent and organize the encoded data. With the quad-tree structure, the sizes of CUs are various from 8×8 to 64×64. At the same time, the PUs are not limited to symmetric partition while asymmetric PUs are also available. To make coding more flexible, the size of TUs is independent from the size of PUs. Moreover, creative techniques are adopted in AVS2 modules of prediction, transform, entropy coding, etc. [1]. **Fig. 1** describes the video coding architecture.

This work is partially supported by the National Basic Research Program of China under grant 2015CB351806, the National Natural Science Foundation of China under contract No. 61425025, No. 61390515 and No. 61421062, and Shenzhen Peacock Plan.

The rest of the paper is organized as follows. The related works are briefly discussed in section 2. Scene video coding techniques are introduced in section 3. Section 4 contains the experimental results of AVS2 scene video coding. The paper is concluded in section 5.

2 Related Works

Some research has been done to improve the compression ef-



▲**Figure 1. The architecture of AVS2 scene video coding.**

ficiency in scene videos.

One of the most direct solutions for surveillance and conference videos is the object-based coding. In the object-oriented analysis-synthesis coding method, each video was coded with motion and shape of objects, color information and prediction residuals. However, object-based coding has three main challenges: accurate foreground segmentation, low-cost object representation, and high-efficiency foreground residual coding [2].

In the traditional hybrid coding framework, hybrid block-based methods are used to encode each picture block by block. The main types of these methods include the following aspects: 1) Region-based coding and 2) Background prediction based coding. The former aimed at achieving better subjective quality of foreground regions with low coding complexity. Instead, with the assumption that in scene videos, there might be one background picture that remains unchanged for a long time, the second method improves the objective compression efficiency by utilizing one background picture as the reference for the following pictures.

However, there are some regions that may appear in the current frame but are covered by objects in the recent reference frames or the key frame. Thus, it is hard to compress the regions efficiently by using the key frame as the background. To address this problem, several background modeling based methods were proposed, for example, using the reconstructed pictures to model the background or utilizing the background picture that was modeled from the original input frames as the reference for more efficient background prediction.

3 AVS2 Scene Video Coding Techniques

As we know, the key to improve the coding performance efficiently of scene videos is reducing the background redundancy. AVS2 adopts the long-term reference technique and S picture to reduce the background redundancy to improve the coding performance efficiently [3].

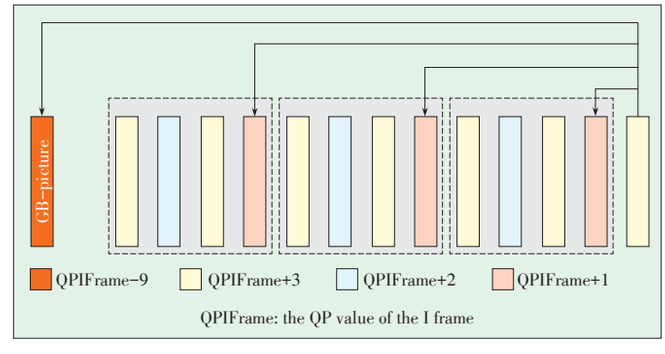
3.1 The Long-Term Reference Technique

Traditionally, the current frame can only be inter-predicted by the previous frames in the group of pictures (GOP). Thus, the distance between the current picture and the reference frame is only relatively short, which means that the reference frame may not be able to provide abundant prediction in the background regions. In order to provide better reference for background regions, AVS2 adopts a long-term reference frame named background picture (GB picture) [1], [4].

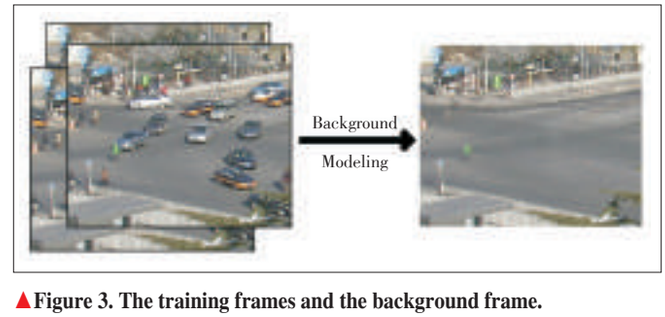
As shown in **Fig. 2**, GB picture is a background picture where the whole picture is background regions, so the background regions of each subsequent inter-predicted frame can always find the matching regions in GB picture. When encoding the GB picture, only intra mode is utilized, and smaller Quantization Parameter (QP) is selected to obtain a high quality GB picture. When the P picture is uses the long-term reference technique, the reference picture chain comprises general reference sequence and the long-term GB picture. Thus when the subsequent inter-predicted frames choose reference frames for their background regions, there is high possibility to select GB picture. Although the GB picture takes a lot of bits, more bitrate will be saved when the following frames refer GB picture because of the high quality of GB picture. As a result, the total performance becomes better.

Although the AVS2 standard does not limit the way a GB picture is generated, it chooses the segment-and-weight based running average (SWRA) to generate the GB picture in AVS2 Reference Design (RD). By weighting the frequent values more heavily in the averaging process, SWRA can generate pure background. The specific process is as follows. Technologically, SWRA divides the pixels at a position in the training pictures into temporal segments with their own mean values and weights and then calculates the running and weighted average result on the mean values of the segments. In the process, pixels in the same segment have the same background/foreground property, and the long segments are more heavily weighted. Experimental results [5] show that SWRA can achieve good performance yet without suffering a large memory cost and high computational complexity, which can meet the requirement of real-time transmission and storage for scene videos. An example of the constructed background frame and the training frames are shown in **Fig. 3**.

Once a GB picture is obtained, it is encoded, and the reconstructed picture is stored in the independent background memory and updated only if a new GB picture is selected or generated. The update mechanism guarantees the effectiveness of the



▲Figure 2. The long-term reference technique.



▲Figure 3. The training frames and the background frame.

GB picture.

3.2 S Picture for Random Access

To ensure random access ability, the picture at the random access point is decoded independent of the previous frames. In previous coding frame standards, I picture can be used as the random access point. However, the performance of I picture is not very well because it only adopts intra prediction and the performance of intra prediction is not equal to inter prediction. Along with GB picture and the long-term reference technique, another picture type called S picture is designed for balancing the coding performance and purpose of random access.

S picture is similar to P picture, which can only be predicted from a reconstructed GB picture and has no motion vector, so only three modes including Intra, Skip and 2N×2N are available in S picture. These characteristics make it possible for an S picture to be an ideal replacement for an I picture. Before the S picture is generated, the GB picture if first obtained to ensure the decoding independence of S picture. Zero motion vector in the S picture also makes sure that there is no need in consideration of the motion vector prediction (MVP). Thus the relative independence of the S picture makes sure it can be set as the random access point and can present better performance than I picture.

3.3 Improvement of Motion Vector Derivation

The BlockDistance is the distance between the current block and the reference block pointed by the motion vector, which is associated with the picture order count (POC) of reference picture. In the case of one reference with two motion vectors, such as F picture, one of the motion vectors is calculated by the other motion vector. When we introduce GB picture to AVS2, the problem comes. Because there is no POC existing in GB picture, the BlockDistance between current block and its reference block is unavailable if the reference block is from GB picture. In order to solve the problem, AVS2 provides a strategy in this situation. If one of the reference pictures is GB picture, the BlockDistance between current block and the block in GB picture is restricted to 1. By doing so, the motion vector derivation is available all the time no matter GB picture is involved in or not.

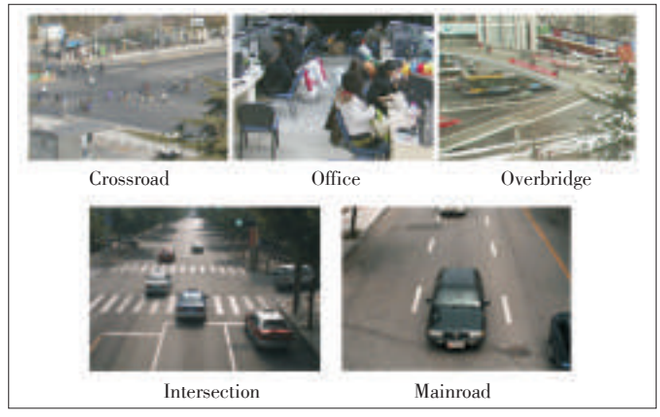
4 Performance Evaluation of AVS2 Scene Video Coding

4.1 Common Test Sequences and Conditions

There are five typical scene videos selected as the common test sequences [6], [7]. Three are 720×576 surveillance videos, and the other two are 1600×1200 ones (**Table 1**). From **Fig. 4**, these five surveillance videos cover different monitoring scenes, including bright and dusky lightness (BR/DU), large and small foreground (LF/SF), fast and slow motion (FM/SM).

▼**Table 1. The common test sequences of AVS2 scene video coding**

Resolution	FrameRate	Sequence	FramesToBeEncoded
720×576	30	Crossroad	600
		Office	
		Overbridge	
1600×1200	30	Intersection	600
		Mainroad	



▲**Figure 4. The common test sequences for scene video coding in AVS2.**

To evaluate the coding performance of the scene video compression of AVS2 (RD 12.0.1 Scene), the latest released reference software for AVS2 keeping the scene video coding techniques disabled (RD 12.0.1 General) is used as the basic experimental platform. Here, our objective is to evaluate the improvement in efficiency and reduction in complexity that AVS2 scene video coding can achieve over AVS2 General.

Four configurations are adopted to perform the experiment [8]. They are:

1) Low delay (LD);
2) Random Access with B slices (RAB);
3) Random Access with F slices (RAF);
4) Random Access with P slices (RAP).

The F frame is a bidirectional reference frame. Unlike the B frame, one motion vector of the F frame is derived from the other motion vector.

Table 2 shows the common test conditions of AVS2 scene video coding.

4.2 Performance Evaluation

The coding performance between RD 12.0.1 Scene and RD 12.0.1 General is shown in **Table 3**. According to the experimental result, RD 12.0.1 Scene reduces 24.33% (LD), 44.11% (RAB), 40.25% (RAF) and 40.56% (RAP) bitrates in average against RD 12.0.1 General on 720×576 videos and 42.07% (LD), 39.24% (RAB), 38.36% (RAF) and 37.90% (RAP) on 1600×1200 videos. Among the video sequences, Office and Intersection have large foreground objects and they are hard to generate clear background picture, so the coding performance

▼Table 2. The common test conditions of AVS2 scene video coding

Parameter	LD	RAB	RAF	RAP
QPIFrame	27, 32, 38, 45			
QPPFrame	QPIFrame+1			
QPBFrame	-	QPIFrame+4	-	-
SeqHeaderPeriod	0	1	1	1
IntraPeriod	0	32	32	32
NumberBFrames	0	7	0	0
FrameSkip	0	7	0	0
BackgroundQP	QPIFrame-9			
BackgroundEnable	1			
FFRAMEEnable	1	1	1	0
ModelNumber	120			
BackgroundPeriod	900	112	900	900

LD: low delay
RAB: random access with B slices
RAF: random access with F slices
RAP: random access with P slices

▼Table 3. The coding performance comparison between RD 12.0.1 Scene and RD 12.0.1 General

Resolution	Sequence	RD 12.0.1 Scene vs. RD 12.0.1 General (BD-Rate)			
		LD	RAB	RAF	RAP
720×576	Crossroad	-25.64%	-41.99%	-37.48%	-38.07%
	Office	-12.66%	-26.77%	-23.77%	-24.10%
	Overbridge	-34.10%	-63.58%	-59.50%	-59.51%
	Average	-24.13%	-44.11%	-40.25%	-40.56%
1600×1200	Intersection	-22.46%	-22.06%	-21.19%	-19.91%
	Mainroad	-61.68%	-56.42%	-55.52%	-55.90%
	Average	-42.07%	-39.24%	-38.36%	-37.90%
All	Average	-31.31%	-42.16%	-39.49%	-39.50%

BD: Bjøntegaard-Delta
RD: Reference Design
LD: low delay
RAB: random access with B slices
RAF: random access with F slices
RAP: random access with P slices

is relatively lower than others. In average, RD 12.0.1 Scene can obtain 31.31% (LD), 42.16% (RAB), 39.49% (RAF) and 39.50% (RAP) bitrate savings on all common test sequences.

5 Conclusions

Based on the classic block-based hybrid video framework, AVS2 is the latest coding standard with efficient scene video coding techniques and is designed for high efficiency video coding of scene videos. This paper introduces several representative techniques adopted in AVS2, including the long-term reference technique and S picture.

By adopting the techniques of scene video coding mentioned above, AVS2 can gain 31.31% (LD), 42.16% (RAB), 39.49% (RAF) and 39.50% (RAP) BD-rate saving in coding efficiency on scene videos. The excellent coding performance of AVS2 in scene videos coding will bring a bright prospect in video coding research and industrial fields.

References

[1] L. Zhao, S. Dong, P. Xing, and X. Zhang, "AVS2 surveillance video coding platform," AVS M3221, Dec. 2013.
[2] X. Zhang, Y. Tian, T. Huang, S. Dong, and W. Gao, "Optimizing the hierarchical prediction and coding in hevc for surveillance and conference videos with background modeling," *IEEE Transaction on Image Processing*, vol. 23, no.10, pp. 4511–4526, Oct. 2014. doi: 10.1109/TIP.2014.2352036.
[3] F. Liang, "Information technology—advanced media coding part2: video (FCD4)," AVS N2216, Sept. 2015.
[4] R. Wang, Z. Ren, H. Wang, "Background-predictive picture for video coding," AVS M2189, Dec. 2007.
[5] X. Zhang, Y. Tian, T. Huang, and W. Gao, "Low-complexity and high-efficiency background modelling for surveillance video coding," in *Proc. IEEE International Conference on Visual Communication and Image Processing*, San Diego, USA, Nov. 2012, pp. 1–6. doi: 10.1109/VCIP.2012.6410796.
[6] S. Dong, L. Zhao, "AVS2 surveillance test sequences," AVS M3168, Sept. 2013.
[7] L. Yu, "Meeting summary of AVS2 video coding subgroup," AVS N1998, Sept. 2013.
[8] X. Zheng, "Common test conditions of AVS2-P2 surveillance profile," AVS N2217, Sept. 2015.

Manuscript received: 2015-11-25

Biographies

Jiaying Yan (yanjiaying@pku.edu.cn) received the BS degree from Beijing Institute of Technology, China in 2014. He is currently pursuing the MS degree with the School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, China. His research interests include surveillance video coding and multimedia learning.

Siwei Dong (swdong@pku.edu.cn) received the B.S. degree from Chongqing University, China in 2012. He is currently pursuing the PhD degree with the School of Electronics Engineering and Computer Science, Peking University, China. His research interests include video coding and multimedia learning.

Yonghong Tian (yhtian@pku.edu.cn) is currently a professor with the National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, China. He received the PhD degree from the Institute of Computing Technology, Chinese Academy of Sciences, China in 2005, and was also a visiting scientist at Department of Computer Science/Engineering, University of Minnesota, USA from November 2009 to July 2010. His research interests include machine learning, computer vision, video analysis and coding, and multimedia big data. He is the author or coauthor of over 110 technical articles in refereed journals and Conferences. Dr. Tian is currently an associate editor of *IEEE Transactions on Multimedia*, a young associate editor of the *Frontiers of Computer Science*, and a member of the IEEE TCMC-TCSEM Joint Executive Committee in Asia (JECA). He was the recipient of the Second Prize of National Science and Technology Progress Awards in 2010, the best performer in the TRECVID content-based copy detection (CCD) task (2010–2011), the top performer in the TRECVID retrospective surveillance event detection (SED) task (2009–2012), and the winner of the WikipediaMM task in ImageCLEF 2008. He is a senior member of IEEE and a member of ACM.

Tiejun Huang (tjhuang@pku.edu.cn) is a professor with the School of Electronic Engineering and Computer Science, the chair of Department of Computer Science and the director of the Institute for Digital Media Technology, Peking University, China. His research areas include video coding and image understanding, especially neural coding inspired information coding theory in last years. He received the PhD degree in pattern recognition and intelligent system from the Huazhong (Central China) University of Science and Technology in 1998, and the master's and bachelor's degrees in computer science from the Wuhan University of Technology in 1995 and 1992, respectively. Professor Huang received the National Science Fund for Distinguished Young Scholars of China in 2014. He is a member of the Board of the Chinese Institute of Electronics, the Board of Directors for Digital Media Project and the Advisory Board of IEEE Computing Now.