OVERVIEW OF THE IEEE 1857 SURVEILLANCE GROUPS

Xianguo Zhang, *Tiejun Huang*, *Yonghong Tian*, *Wen Gao*

Institute of Digital Media, Peking University, Beijing, 100871, P. R. China
{tjhuang, yhtian}@pku.edu.cn

ABSTRACT

Among the multiple application-oriented video groups of IEEE 1857 video part, surveillance groups are the first specific video coding standards targeting on the exploring surveillance system. In this paper, we firstly present an overview of the technical features and characteristics of the *Surveillance Baseline* and *Surveillance* Groups. The video coding technologies are then described in greater detail on three main directions, including the background modeling based prediction techniques for high-efficiency surveillance video coding, error resilience methods for channel-noisy surveillance video transmission and the high-level syntax for surveillance video analysis. The surveillance groups can provide a good support for kinds of video analysis applications of computer vision and make the video transmission more applicable for noisy channels. Moreover, experimental results show that the background modeling based prediction techniques can well exploit the special characteristics of surveillance video and double the traditional compression performance.

Index Terms—video coding; IEEE 1857; surveillance groups; background modeling; region of interest

1. INTRODUCTION

In IEEE 1857 video part [1], multiple application-oriented video groups (i.e., *Main*, *Surveillance Baseline*, *Surveillance*, *Portable*, *Enhanced* and *Broadcasting Groups*) are defined to satisfy different requirements from typical applications. These groups combine advanced video coding tools with trade-off between coding efficiency and encoder/decoder implementation complexity as well as functional properties. Among them, the *Surveillance Baseline* and *Surveillance Groups* focus on standardizing the solutions for the video surveillance applications. In these groups, the characteristics of surveillance videos are specially considered, including the redundant background data compression, transmission with random noise, less affordable encoding complexity and friendliness to events detection and searching.

The main characteristic of surveillance video can be described as: the cameras are always deployed towards the same scene for a long time. As a result, the scene shot seldom happens in surveillance videos and there are usually similar background data among hundreds and thousands of pictures. To compress the background data more efficiently (i.e., reducing the temporal background redundancy), novel inter prediction techniques are adopted in these groups, including background picture (G-picture) and background-predictive picture (S-picture) [2], background modeling, background reference selection, background difference prediction [3], Bitstream Buffer Verifier (BBV) management and motion vector prediction (MVP) for the not-display modeled background frame (MBG) [4]. Details for these methods are described in Sec. 2, and the encoding/decoding complexity increase is insignificant. Besides the methods for background redundancy reduction, other techniques without complexity increase are also utilized in surveillance groups, i.e., the adaptive weighting quantization [5] and optimized motion vector scaling for interlaced video coding [6].

Another characteristic of surveillance videos is the channel-noisy video transmission. To make video transmission more robust, surveillance groups also employ some error resilience techniques. Sec. 3 will present a description of the methods including the non-reference P picture [7], flexible slice set [8], core picture [9] and constrained DC intra prediction [10].

Table 1. The comparison of adopted tools in surveillance groups

Coding tools	Surveillance	Baseline
Entropy Coding	AEC & 2DVLC	2DVLC
S-picture Coding	Only Intra/P16×16/PSKIP, no motion vector	
	The nearest G-picture is the only reference	
Extension data	Region Extension	
	Sequence/Picture Display Extension	
	Camera Parameter & Copyright Extension	
G-picture built up	Background modeling	Key picture
G-picture Coding	Much smaller QP on MBG	Smaller QP
G-picture display	MBG not output/not display	Output/display
Managing BBV	MBG is special managed	as *Main* group
MVP of P/B-pictures	Improved when MBG is used	as *Main* group
	Optimization for interlaced video coding	
Reference pictures	MBG can be set 2nd reference	as *Main* group
Being Referenced by following G-pictures	P-picture can be set as non-reference	
	Pictures before core pictures cannot be referenced by those following core pictures	
Inter Prediction	Background difference Prediction can be used	as *Main* group
Slice	Flexible Slice Set can be used	
Intra Prediction	Constrained DC intra prediction can be used	
Quantization	MB-adaptive weighting quantization	

In addition, surveillance videos are always captured and stored for video analysis applications like event detection and searching in case investigation. Therefore, surveillance groups adopt novel high-level syntax to describe the region-of-interest, the camera parameters, copyright data and sequence/picture display description in forms of some syntax extensions on sequence and picture headers. Sec. 4 will give an overview of these extensions.

To sum up the relationships between these two surveillance groups, Table 1 lists which tools are utilized in each of them. As is seen, *Surveillance Group* enhanced the *Surveillance Baseline Group* on the prediction techniques better utilizing the MBG and AEC ([11], advanced entropy coding, which is adopted in *Enhanced* and *Broadcasting* groups, instead of the 2DVLC [12] in main group) for more efficient entropy coding.

The rest of this paper is organized as follows. Sec. 2 presents the background redundancy reduction methods in surveillance groups, and the error resilience tools are introduced in Sec. 3. The descriptions of the extension data and their utilization are discussed in Sec. 4. Extensive experiments are conducted in Sec. 5 to prove surveillance groups' efficiency. Sec. 6 concludes this paper.

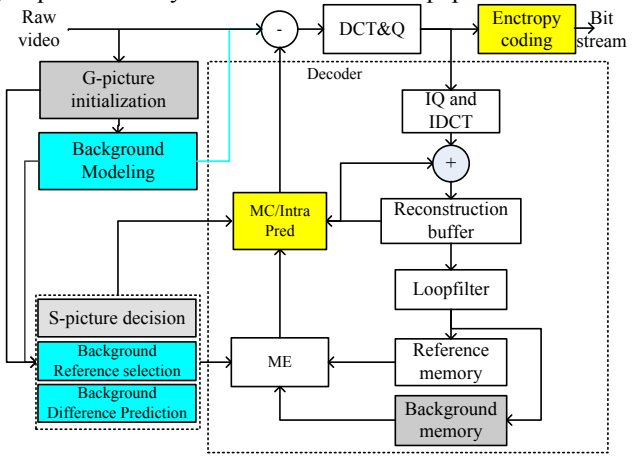

Fig. 1. Video codec architecture of surveillance groups. In this the gray modules are the common ones in surveillance groups; the blue modules are only adopted in *Surveillance Group*; the yellow modules are optimized by the *Surveillance Group*

2. TOOLS FOR REDUNDANCY REDUCTION

In order to reduce the background redundancy, surveillance groups adopt novel inter prediction tools based on G-picture. In this section, we firstly introduce the background prediction based video codec architecture shared by all the surveillance groups. Secondly, we will introduce the novel tools for better utilizing the MBG, i.e., the background reference selection, background difference prediction, BBV management and MVP related to MBG.

2.1. Video Codec Architecture

In all the surveillance groups, G-picture and S-picture are defined to further exploit the temporal redundancy and suppress background noise. The G-picture is a special I-picture and its reconstructed picture is stored in a separate background memory. The S-picture is a special P-picture which can be only predictable from reconstructed G-picture. In *Surveillance* group, the G-picture can be a non-display MBG which is modeled from input pictures and encoded into stream to guarantee the decoding match, and then each P-picture can also utilize G-picture as prediction reference or be encoded utilizing background difference prediction.

To well support such above method and accomplish a high-efficiency video codec, Fig. 1 shows the architecture for all surveillance groups. In this architecture, six additional modules compared with main groups are marked. G-picture is initialized by *G-picture initialization* and updated by *Background Modeling* with methods such as median filtering etc. In such way, the selected or generated G-picture can well represent the background of a scene with seldom or even no occluding foreground objects and noise. Once a G-picture is obtained and encoded, the reconstructed picture will stored into the *Background Memory* in encoder/decoder and updated only if a new G-picture is selected or generated. After that, S-picture can be involved in the encoding process by *S-picture Decision*. Only with G-picture as reference, the S-picture owns similar utilities as traditional I-picture, such as error resilience and random access. Therefore, the pictures which should be coded as traditional I-pictures are candidate S-pictures, such as the first picture of one GOP, scene change frame etc. To accomplish a fast random access and error resilience, only intra, SKIP and P16×16 modes with zero motion vectors are available for S-picture in surveillance groups.

2.2. Tools for better utilizing the MBG

In *Surveillance Baseline Group*, G-picture is selected from input pictures and only utilized to predict its following two pictures and the S-pictures. After supporting encoding MBG as G-pictures in *Surveillance Group*, more tools are developed to utilize MBG for better inter prediction.

1) A novel segment-and-weight based running average (SWRA) is utilized to generate the MBG by assigning a larger weight on the frequent values in the averaging process. SWRA divides the pixels at a position in the training frames into temporal segments with their own mean values and weights. And then, it 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 longer segments have larger weights. Experimental results show that SWRA can achieve good performance yet without suffering a large memory cost and high computational complexity [13].

2) MBG based background prediction: To obtain better prediction efficiency of the background pixels in the current frame, MBG can be quantized with a much smaller QP and be encoded as a non-display frame. Moreover, to realize better prediction efficiency of the background pixels in the current picture, *Surveillance Group* codec can indicate

whether the reconstructed MBG is the second reference on P/B-picture header. Because MBG has few foreground is pixels and is updated infrequently, it can be quantized with much smaller quantization parameter (QP) to enlarge the prediction efficiency of the background pixels of the frames with MBG as reference.

3) An improved algorithm is specially designed for the predicted motion vector (PMV) derivation when using MBG as direct or indirect prediction reference. That is, if one neighbor of the current macroblock (MB) utilizes or does not utilize the MBG as reference synchronously with the current MB, the contribution of this neighboring block in deriving the final PMV of the current MB should be set 0. In addition, when the co-located block in the backward reference frame utilizes the MBG as reference, the spatial derivation instead of the temporal derivation of PMV should be adopted for the current MB in B frames. In this way, dividing-zero error can also be avoided.

4) Improved BBV management: An improved mechanism is designed to deal with the MBG so as to support a no-delay, no-frame-drop and real-time encoding. It requires the display distance between the frames before the MBG and after it equal to 1. Moreover, the checking interval between each MBG and its next picture must be 0.

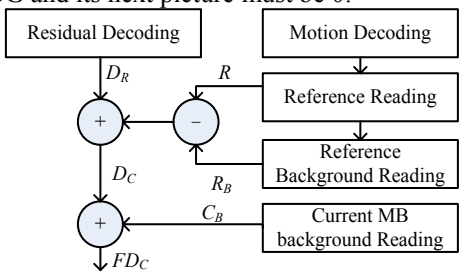

Fig. 2. Decoding process of an MB which adopts background difference prediction

5) Background difference prediction. While MBG is set as the second reference frame, the background difference prediction can also be selected to encode each MB. While using background difference prediction, the finally decoded result of current MB FD_C is calculated by

$$FD_C = D_C + C_B = D_R + R - R_B + C_B. \quad (1)$$

In this Equation, $R-R_B$ is the difference data between motion-vector-pointed reference data R and its background R_B, C_B is the background data in the MBG at the position of the current MB, and $D_C=D_R+R-R_B$ is the directly decoded data using $R-R_B$ to compensate the decoded residual D_R. Fig. 2 shows the decoding process of an MB utilizing background difference prediction. As for the encoding process, we firstly search the $C-C_B$ within *Ref−MBG* to get the motion vector and corresponding R. In this process, C is the input data of the current MB and *Ref* is the first reference frame. Afterwards, transforming, quantizing and entropy coding are conducted on the prediction residual $C-C_B-(R-R_B)$. In such way, MBs with both foreground and background are encoded more efficiently.

3. ERROR RESILIENCE TOOLS

Error resilience is very important for video transmission, especially when the network is erroneous. H.264/AVC is very promising for low bit-rate applications such as IP network and wireless communications. To adapt these applications, more error resilience tools are introduced into H.264/AVC, such as parameter sets, FMO, and redundant slice etc. As *Main*, *Enhanced* and *Broadcasting* groups mainly target to video broadcasting, few error resilience tools are defined except the slice partition which composes each slice of one or more rows of MBs. However, as surveillance groups may be used in wireless network or used to transport on erroneous channels, error resilience is very important. Inspired by this, tools of non-reference P-picture, core picture coding, flexible slice set and constrained intra prediction are further introduced.

Non-reference P-picture is defined to make one picture never being referenced by following frames, and then the picture following this non-reference P-picture utilizes two pictures before it as candidate reference frames. In this way, while channel-error happens, the codec can encode some non-reference P-pictures into the bit stream.

Core picture is a special inter-picture which can be only predicted from another core picture. The following P-pictures of one core picture can only refer to the frames after its nearest core picture as reference frames. If there is a feedback channel from decoder to encoder, only the core pictures that have been decoded should be taken as reference picture for core pictures. Usability of core picture is signaled in the sequence header.

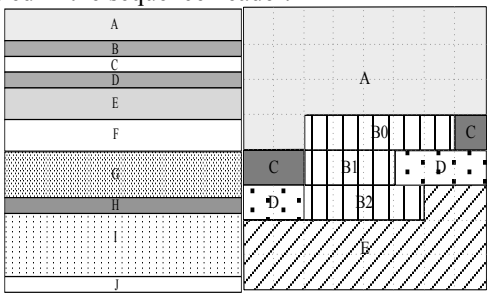

Fig. 3. Slice structure in *Main Group* and surveillance groups. Left: normal slice structure where slice only has continual lines of MBs; Right: flexible slice set allowing more flexible grouping of MBs.

Constrained DC intra-prediction is another error resilience tool defined in surveillance groups. One marker bit in sequence header and one marker bit in picture header of P and B picture indicates the use of constrained DC intra-prediction mode instead of normal DC mode. Constrained DC intra-prediction mode constrains the prediction values of DC mode to be fixed to 128 for luminance 8×8 blocks to avoid possible error accumulation.

Flexible slice set provides robustness to transmission. It allows that slices with the same index of slice group in one picture belong to the same slice group. One slice can refer to other slices in the same slice group. As shown in Fig. 3, slices $B0$, $B1$ and $B2$ belong to the same slice group and they can refer to each other. Flexible slice set is helpful for both coding efficiency and error robustness.

4. EXTENSION DATA FOR VIDEO ANALYSIS

In surveillance groups, five kinds of extension data are added after sequence and picture header in the high-level syntax, including region, sequence/picture display, camera parameter and copyright extension.

In region extension, region number, event ID, coordinates for top left and bottom right corners are included to show what number the region of interest (ROI) is, what event happens and where it lies. Based on these syntaxes, we can mark the ROI in bit stream after we detect objects with the help of background subtracting for surveillance videos. As shown in Fig. 6, following the steps of background modeling, background subtracting and foreground clustering, convergent ROI region can be automatically recognized.

In sequence/picture display extension, bits for video format indicate the video sequence play in PAL, NSTC, SECAM or MAC; there are also some bits indicating that the center of the reconstructed frame lies below and right of the center of the display rectangle. The copyright extension adds copyright information in the bit streams. Note that, these extension data have their separate start codes, consequently leading to no bit-rate increase, if the codec does not plan to use any extension data.

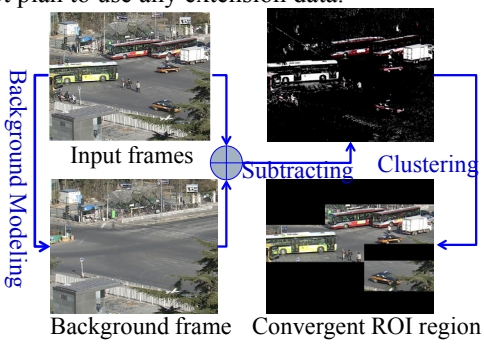

Fig. 6. Automatic ROI region clustering.

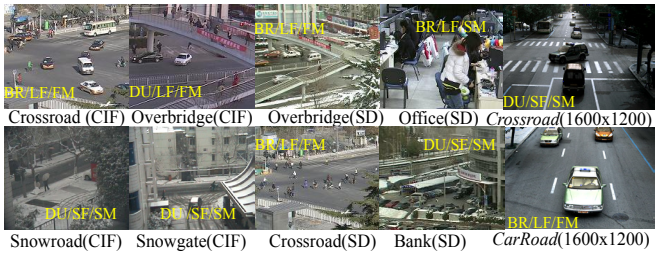

Fig. 7. Ten typical surveillance videos

5. EXPERIMENTS

To evaluate the surveillance video compression performance more extensively, ten CIF-HD surveillance videos with 1020 frames are utilized as the test sequences. As shown in Fig. 7, these ten videos cover different monitoring scenes, including bright and dusky lightness (BR/DU), large and small foreground (LF/SF), fast and slow motion (FM/SM).

In our experiments, H.264/AVC high profile (HP), IEEE 1857 main (MA) and broadcasting (BC) groups are utilized as the three anchors to compare with the groups of *Surveillance Baseline Group* (SBG) and *Surveillance Group* (SG). It should be noted that, all the codecs are commonly configured in Table 2. Moreover, the QP for MBG equal to that of I-picture minus 20.

Table 2. Common Configurations

Conf.	Value	Conf.	Value	Conf.	Value
Frame structure	*IPPP*	Fast search	*Enable*	Loop Filter	*Enable*
		GOP size	30	Reference No.	*2*
Rate control	*Disable*	Search range	64	Sub pixel	*Quarter*
RDOQ	*Disable*	Hadamard ME	1	I/P QP gap	1

Table 3 presents the comparison results between surveillance groups and the anchors. As is seen, nearly all surveillance groups can achieve bit-saving over the anchors on each sequence, and the only exception happens on SBG vs. BC. The exception is mainly because the AEC utilized in BC performs much better than the 2DVLC in SBG, and the gain from AEC exceeds the gain from G/S-picture. In summary, SBG averagely achieves 24.7/30.3/24.3% bit-saving than HP/MA/BC, and SG can save 51.4/55.4/51.2% of the total bitrate of HP/MA/BC on average. We can observe from the results that, the better utilization of MBG in SG greatly increase the total coding efficiency.

Table 3. Bitrate change of surveillance groups vs. the anchors (%)

Sequence	SBG vs.			SG vs.		
	HP	MA	BC	HP	MA	BC
Bank	-39.6	-41.6	-33.6	-74.2	-75.0	-70.9
Crossroad	-11.9	-19.3	-13.4	-40.9	-46.1	-42.1
Office	-7.6	-17.2	-11.0	-28.5	-36.0	-31.0
Overbridge	-46.0	-48.9	-43.3	-72.7	-74.5	-71.0
crossroad-cif	-4.5	-16.3	-10.8	-27.4	-36.3	-32.1
overbridge-cif	-2.8	-11.8	-6.7	-32.1	-38.3	-34.8
snowgate-cif	-57.6	-58.9	-55.2	-77.5	-78.3	-76.3
snowroad-cif	-48.5	-52.9	-49.2	-70.9	-73.4	-70.6
Crossroad-HD	-3.8	-8.2	**3.6**	-22.9	-25.8	-16.4
CarRoad-HD	-25.0	-27.8	-20.2	-66.9	-69.8	-66.8
Average	-24.7	-30.3	-24.0	-51.4	-55.4	-51.2

6. CONCLUSION

In this paper, we presented an overview of the adopted tools in surveillance groups of IEEE 1857 video part, including the codec architecture, the better utilization of background picture, the error-resilience tools and the extension data. These tools can well support kinds of requirements for surveillance video coding, transmission and analysis applications. Moreover, video codecs can double the surveillance video coding efficiency following these groups.

7. ACKNOWLEDGEMENT

This work is partially supported by grants from National Basic Research Program of China under contract No. 2009CB320906, the Chinese National Natural Science Foundation under contract No. 61035001, 61121002 and 61176139.

8. REFERENCES

[1] P1857/D1, "Draft Trial-Use Standard for Advanced Audio and Video Coding," July 2012

[2] R. Wang, Z. Ren, H. Wang, "Background-predictive picture for video coding," in AVS Doc. M2189, 2007

[3] X. Zhang, Y. Tian, T. Huang, et al., "Macro-block-level Selective Background Difference Coding for Surveillance Video," in Proc. IEEE Int. Conf. Multimedia Expo, pp. 1067-1072, July 2012

[4] X. Zhang, T. Huang, "Change for Testing time interval in AVS-P2 for surveillance video coding," in AVS Doc. M2886, 2011

[5] J. Zheng, X. Zheng, C. Lai, "Adaptive Weighting Technology for AVS Jiaqiang Profile," in Document AVS_M2427, 2008.

[6] Y. Lin, X. Zheng, M. Han, et al., "Motion vector scaling based on half-pixel motion compensation," in AVS Doc. M2329, 2008

[7] Q. Wang, X. Mou, Q. Liu, M. Li, "A video coding method and reference management satisfying temporal scalable video stream," in AVS Doc. M2384, 2008

[8] M. Zhen, Z. Wu, X. Xu, Y. He, "AVS-S flexible slice group," in AVS Doc. M2305, 2008.

[9] R. Chen, P. Yu, X. Huang, N. Wang, "A coding method for error resilience," in AVS Doc. M2192, 2007.

[10] C. Lai, X. Zheng, Y. Lin, J. Zheng, M. Han, "Constrained DC intra-prediction in AVS-S," in AVS Doc. M2464, 2008.

[11] Q. Wang, D. Zhao, W. Gao, "Context-Based 2D-VLC Entropy Coder in AVS Video Coing Standard," in Journal of Computer Science and Technology, 21(3), 315-322, 2006.

[12] L. Zhang, Q. Wang, N. Zhang, D. Zhao, X. Wu, W. Gao, "Context-based entropy coding in AVS video coding standard," in Signal Processing: Image Communication, 24(4), 263-276, 2009.

[13] X. G. Zhang, Y. H. Tian, T. J. Huang and W. Gao, "Low-complexity and High-efficiency Background Modeling for Surveillance Video Coding," Proc. 2012 IEEE Int'l Conf. Visual Communication and Image Processing, San Jose, Nov 2012