This article has been accepted for publication in IEEE Intelligent Systems but has not yet been fully edited.
Some content may change prior to final publication.

IEEE 1857 Standard Empowering Smart Video Surveillance Systems

Wen Gao, Yonghong Tian, Tiejun Huang, Siwei Ma, Xianguo Zhang
School of EE & CS, Peking University

Abstract

IEEE 1857, *Standard for Advanced Audio and Video Coding*, was released as IEEE 1857-2013 in June 2013. Despite consisting of several different groups, the most significant feature of IEEE 1857-2013 is its surveillance groups, which can not only achieve at least twice coding efficiency on surveillance videos as H.264/AVC HP, but also should be the most recognition-friendly video coding standard till to now. This article presents an overview of IEEE 1857 surveillance groups, highlighting the background model based coding technology and recognition-friendly functionalities. We believe that IEEE 1857-2013 will bring new opportunities and drives to the research communities and industries on smart video surveillance systems.

Keyword: Surveillance video processing, video coding standard, background modeling, IEEE 1857

-- ★ --

The past seventies have seen an explosion in the use of video surveillance in many places from public security to traffic control and in-home nanny monitoring. Nowadays, different numbers of cameras are connected to recording and display devices via transmission networks, and constitute the so-called video surveillance systems at different scales.

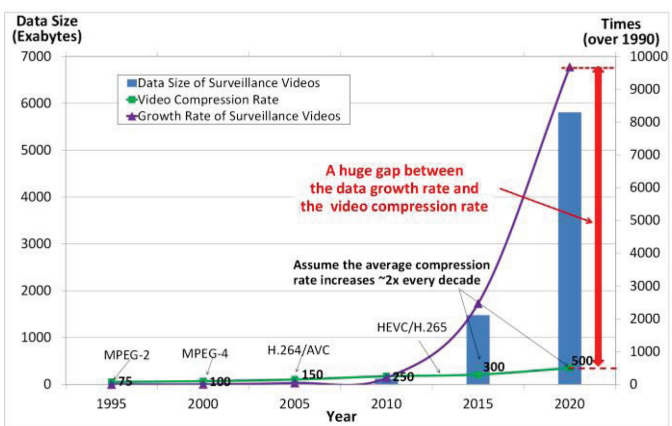


Fig. 1. The gap between the growth rate of surveillance videos and the video compression rate in recent three decades. In this figure, the growth rates of surveillance videos are approximately estimated based on the data from [1], while the compression rates in future several years are predicted by assuming that the average compression rate increases ~2x every decade.

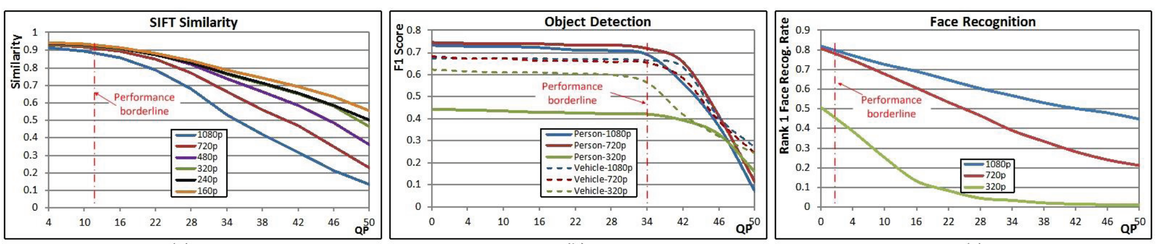
According to a recent report by IDC [1], by 2020, as much as 5,800 exabytes of surveillance videos will be saved, transmitted and analyzed. Traditionally, the industry employs the coding standards such as

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

This article has been accepted for publication in IEEE Intelligent Systems but has not yet been fully edited.
Some content may change prior to final publication.

MPEG.4 and H.264/AVC [2] that were originally designed for generic videos to compress surveillance videos. If we still follow this technology roadmap, in future several years, the growth rate of surveillance videos will be much higher than the video compression rate that H.264/AVC and even its next-generation development, HEVC/H.265 [5], can achieve. As shown in Fig. 1, this will lead to a huge gap between the two rates in future several years, consequently presenting an unprecedented challenge for high-efficiency and low-complexity surveillance video coding technology.

Besides, the resolution of surveillance cameras, the bandwidth of transmission network, the capacity of storage systems, as well as the recognition capability of video analysis software, all these may also influence the performance of a video surveillance system to varying degrees. In general, the higher the resolution of cameras, the better the quality of the captured videos or images, which in turn may enhance the visual experience in monitoring and improve the video analytics (e.g., face recognition). At the same time, however, such a video surveillance system would produce a larger amount of video data, leading to a much more expensive cost for the network bandwidth and the storage system. To solve this problem, most of current video surveillance systems often compress a high-definition video of 1.5 Gbps to 10Mbps by using H.264 high profile (H.264 HP for short) [2]. Some systems even set the compression rate to 300:1 or higher so as to further reduce the network bandwidth and the storage capacity. Nevertheless, high compression ratio will inevitably bring difficulties in visual feature extraction, consequently reducing the recognition accuracy of video analysis and recognition more or less.



(a) (b) (c)

Fig. 2. The effects of video compression quality on feature analysis, object detection and face recognition. In the experiments, one 1080p uncompressed video selected from the PKU-SVD-A dataset is resized to different definitions and compressed with different quantization parameters (QPs) by H.264 HP. For simplicity, we only use videos with three resolutions (i.e., 1080p, 720p and 320p) for object detection and face recognition experiments. (a) In the feature analysis experiment, SIFT features are extracted on video frames with different definitions and different QPs, and then the average similarity is calculated between SIFT features of the original 1080p frames and those of the corresponding transformed frames. (b) In the object detection experiment, the well-known deformable part model (DPM) trained by PASCAL VOC 2009 dataset [3] is used as the detector for pedestrians and vehicles, and the performance is measured by F1 that is calculated by precision and recall. (c) In the face recognition experiment, the face recognition algorithm [4] is used and the performance is measured by the face recognition rate over the top-rank result.

To evaluate the effects of video compression on typical analysis tasks, we conducted several experiments on one 1080p uncompressed video selected from the PKU-SVD-A dataset (later as shown in Fig. 6), which was then resized to different resolutions and compressed with different quantization parameters (QPs). Generally, QP regulates how much spatial detail is saved in the compression. When QP is very small, almost all that detail is retained; As QP is increased, some of that detail is aggregated so that the bit rate drops – but at the price of some increase in distortion and some loss of quality. As shown in Fig. 2, we can see that for all the three tasks, the performance is gradually going down with the increase of

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

QPs. In this case, the advantages of introducing high-definition cameras would be fully offset by the used video coding technology.

In tragic events such as London Underground bombings and the recent Boston Marathon bombing, surveillance videos have been proved to be an invaluable tool for the investigation. However, technology is far from where it needs to be. For example, after the Boston Marathon bombing, investigators were poring over the mountains of videos and still imagery shots from surveillance cameras and mobiles. It is, by all accounts, a dramatically human-led endeavor to identify the clear facial images of the two suspects. Actually, automatic video analysis techniques are on the verge of being able to slash the manpower needed for such investigations. When equipped with these techniques, video surveillance systems can be really used as a first pass to pare down huge volumes of videos and help investigators focus their search.

From the viewpoint of video coding, a “smart” video surveillance system will at least have the following two features:

(1) *Recognition-friendliness should be its first and most important objective*. In such a smart system, the majority of coding bits should be used to represent the objects or regions-of-interest (ROIs). It even can real-time identify the clear facial images, extract the features of the objects and track their trajectories in the coding process. These information are highly useful for further analysis and recognition, yet cannot be easily obtained in the traditional surveillance video systems.

(2) *High-efficiency and low-complexity video coding technologies should be adopted in the system*. To achieve higher compression rate, the special characteristics of surveillance videos (e.g., the relatively stationary background in the scene) should be utilized in the bitstream compression and coding. To make things better, semantic coding in the object and event layers could be used to support the structuring and summarization of surveillance video clips for both long-term storage and retrieval.

Nevertheless, it is not an easy job to empower a video surveillance system with such features. Fortunately, the recent IEEE 1857 standard is exactly a good choice which can help the developers to construct such a smart video surveillance system.

Overview of IEEE 1857 standard

IEEE 1857, *Standard for Advanced Audio and Video Coding*, was originally developed as early as 2006, and formally proposed to the IEEE Standards Association in March 2012. The proposal was supported by Standards Activities Board of IEEE Computer Society. The IEEE 1857 standard document was approved in March 2013, and then released as IEEE 1857-2013 in June 2013.

In IEEE 1857-2013, several groups (i.e., Main, Portable, Enhanced, Broadcasting, Surveillance Baseline and Surveillance) are defined to satisfy different requirements from some typical applications (see Fig. 3). Among them, the Main Group contains the general and basic high-efficiency hybrid video coding tools; some specific low-definition coding and error-resilient tools are included in the Portable Group; the Enhanced Group adopts some high-complexity-but-efficient coding tools for high-definition movie videos; the Broadcasting Group employs efficient coding tools for interlaced broadcasting videos so as to tradeoff between encoding complexity and efficiency; while the Surveillance Baseline Group and Surveillance Group are designed respectively for surveillance video transmission and storage.

The potential applications of IEEE 1857-2013 include but are not limited to Internet Protocol television

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

(IPTV), IP-based video conference, IP-based surveillance, user-generated multimedia content, and other video/audio-enabled services such as digital television broadcasting, digital storage media, and communication.

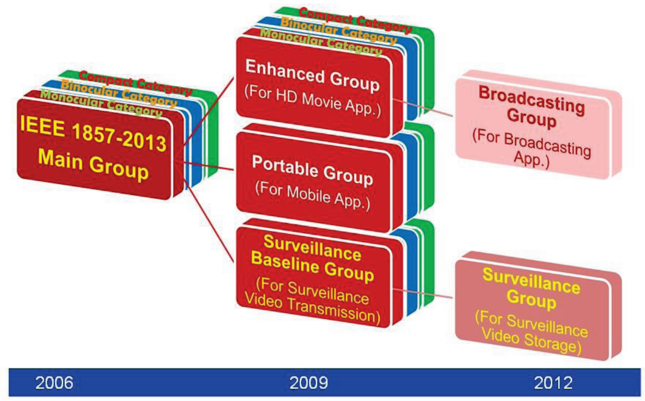


Fig. 3. The hierarchy of IEEE 1857-2013.

Despite consisting of several different groups, the most significant feature of IEEE 1857-2013 is its surveillance groups (i.e., Surveillance Baseline Group and Surveillance Group), which can achieve at least twice coding efficiency on surveillance videos as H.264/AVC HP, yet with a slightly additional encoding complexity. Moreover, they are highly recognition-friendly in that ROIs of attended objects can be coded into the code stream and thus can be directly extracted for further analysis and recognition. These features make the surveillance groups of IEEE 1857 one of enabling video coding standards and tools for the next-generation smart video surveillance systems. In the rest of this article, our discussion mainly focuses on IEEE 1857 surveillance groups.

Double the surveillance video coding efficiency

Different with other state-of-the-art standards such as HEVC/H.265 which improve the coding efficiency at the cost of significantly-increased coding complexity, IEEE 1857 surveillance groups are mainly motivated by the fact that most of surveillance videos are often captured by stationary cameras that always stand towards the same scene for a long time. That is, there are usually similar background data among hundreds and thousands of consecutive pictures. In this case, the background can be modeled and then utilized by the video codec to remove the "scenic redundancy" in these pictures, consequently leading to the remarkable decrease of the total bit rate (see Fig. 4(a)). In IEEE 1857 surveillance groups, the background model is represented as a specially-encoded I-picture (hereafter called *G-picture*) and can be encoded as a non-display frame into the steam to guarantee the decoding match. As shown in Fig. 4(b), the total bit rate should be the same with the traditional coding method in the initial stage, and decrease quickly in the following process. Overall, they can double the surveillance video coding efficiency compared with H.264/AVC HP.

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

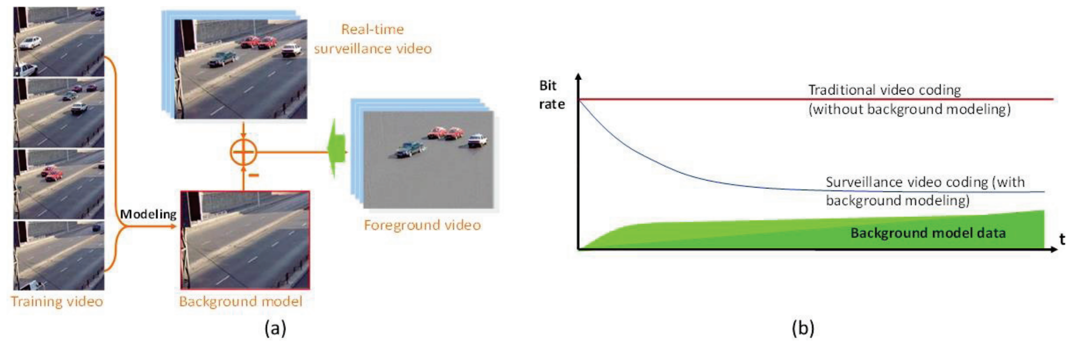


Fig. 4 (a) Surveillance video coding with background modeling, and (b) the ideal bit rate curve.

In particular, IEEE 1857 surveillance groups feature the following seven technological characteristics:

1) *A novel model-based coding framework:* IEEE 1857 surveillance groups add the background-based prediction and coding technology into the traditional hybrid entropy coding framework. As shown in Fig. 5, this novel coding framework include a background modeling and updating module, a background-based motion compensation module, a background buffer and their related control flows (marked by red color).

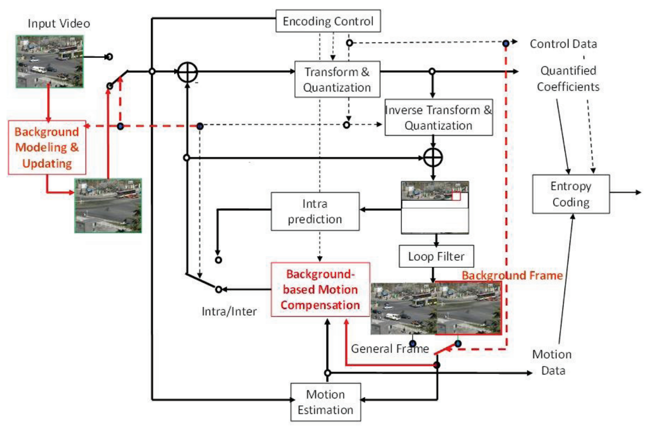


Fig. 5. The background model based coding framework used by IEEE 1857 surveillance groups.

2) *Low-complexity background modeling*: A segment-and-weight based running average (SWRA) is utilized to approximately calculate the background by assigning a larger weight on the frequent values in the averaging process. Technologically, SWRA divides the pixels at a position in the training frames 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 have larger weights. Experimental results show that SWRA can achieve good performance yet without suffering a large memory cost and high computational complexity [9].

3) *G-picture based background prediction*: To realize better prediction efficiency of the background pixels in the current frame, the G-picture can be quantified with a much smaller QP and be encoded as a non-display frame. Moreover, IEEE 1857 surveillance groups also adopt the S-picture as a special

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

P-picture which can be only predicted from the reconstructed G-picture. Therefore, each P frame can be predicted not only from two recent frames, but also from one recent frame and the G-picture.

4) *Optional difference coding for mixed macroblocks (MBs)*: IEEE 1857 surveillance groups can utilize the traditional entropy-based coding tool for each MB in a P-picture. In addition, it can also optionally use the background difference prediction strategy, namely, to predict the difference between the current MB and its corresponding background using the difference data between the recent reference frame and the G-picture. In such a way, MBs with both foreground and background can be encoded more efficiently.

5) *Improved motion vector prediction*: An algorithm is specially designed to derive the predicted motion vector (PMV) when using the not-display G-picture as direct or indirect prediction reference. That is, if one neighbor of the current MB utilizes or does not utilize the not-display G-picture as reference, the contribution of this neighboring block in deriving the final PMV of the current MB should be set 0. In addition, when the block co-located in the backward reference frame utilizes the not-display G-picture as reference, the spatial derivation instead of the temporal derivation of PMV should be adopted. In this way, dividing-zero error can also be avoided.

6) *Improved BBV buffer management*: A special mechanism is designed to deal with the not-display G-picture so as to support a no-delay, no-frame-drop and real-time encoding. It requires that the display distance between the frames before and after the not-display G-picture equals to 1. Moreover, the checking interval between each not-display G-picture and its next picture must be 0.

7) *Error-resilience coding tools*: Some error-resilience tools from other parts of IEEE 1857 are also adopted in IEEE 1857 surveillance groups, including the flexible slice set, core picture coding, the constrained DC intra-prediction, the non-reference P-picture, the optimized motion vector scaling for interlaced video coding and adaptive weighting quantization.

Overall speaking, compared with other video coding standards such as H.264/AVC, the most important contribution of IEEE 1857 surveillance groups is to seamlessly integrate the model-based coding (including background modeling and foreground coding) with the traditional hybrid entropy coding framework. As far as our knowledge, they are the first video coding standard specially designed for surveillance videos.

To evaluate the surveillance video coding performance, we constructed a large-scale dataset, called PKU-SVD-A, by collecting 73 videos with different resolutions (ranging from SD, 720p, 1600*1200, and 1920*1080) or at different weather and time conditions (e.g., dark, frog, rain,…). Some examples are shown in Fig. 6. This dataset will be online publically available soon for the research usage.

On this dataset, we conducted a set of extensive experiments to compare the performance of IEEE 1857 surveillance groups with H.264/AVC HP and IEEE 1857 main group. Under the common test condition for low-delay surveillance video coding (without B frames, while using 2 reference frames and other configurations in [6]), the experimental results are shown in Table I(a). We can see that in terms of the BD rate metric, IEEE 1857 surveillance groups outperform H.264 HP and IEEE 1857 main group on all testing sequences of different resolutions. We also notice that, there seems less bitrate reduction on SD/HD videos than 720p videos. This is because in the PKU-SVD-A, most of the SD/HD videos were collected from surveillance cameras on busy traffic roads, and thus the bit savings produced by background prediction are not as large as those on the 720p videos captured in the campus.

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

This article has been accepted for publication in IEEE Intelligent Systems but has not yet been fully edited.
Some content may change prior to final publication.

Fig. 6. Some examples in the PKU-SVD-A dataset. (a) Videos of different resolutions; (b) Videos for different weather and time conditions. Note that for better visualization, all pictures are resized to the same height but keeping the width-height ration unchanged.

Table I. The performance of IEEE 1857 surveillance groups (IEEE 1857-S) on the PKU-SVD-A dataset

(a) The bitrate reduction of IEEE 1857-S vs. H.264 HP and IEEE 1857 main group (IEEE 1857-M)

	SD	720p	1600x1200	1080p
IEEE 1857-S vs H.264-HP	38.9%	61.8%	35.5%	33.0%
IEEE 1857-S vs. IEEE 1857-M	39.6%	60.8%	30.7%	33.1%

(b) The bitrate reduction of IEEE 1857-S vs. H.264 HP on videos at different periods in one day

	dawn	morning	noon	afternoon	dusk	night
SD traffic videos	59.1%	37.5%	26.4%	33.8%	14.2%	15.1%
720P campus videos	87.95%	52.80%	46.60%	46.79%	63.01%	69.75%
1080p traffic videos	-	38.4%	45.4%	-	22.6%	-

(c) The bitrate reduction and time saving of the model-based HEVC coding over the HEVC test model HM with low-delay configurations

	SD	720p	1600x1200	1080p	AVG
Bitrate reduction	45.40%	53.47%	45.43%	39.24%	**45.89%**
Time saving	49.17%	63.64%	45.94%	24.68%	**45.86%**

Table I(b) shows the experimental results on videos at different periods in one day (from dawn to night). We can observe that for traffic videos, IEEE 1857 surveillance groups obtain the most bit-saving at dawn when the scene is least busy; while for campus videos, they achieve less bit-saving at noon and afternoon than at the other periods. This is mainly because in campus videos, more foreground objects appear at

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

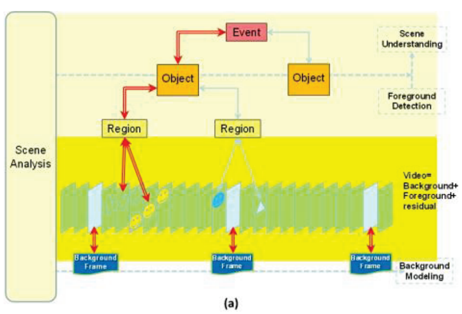
This article has been accepted for publication in IEEE Intelligent Systems but has not yet been fully edited.
Some content may change prior to final publication.

noon and afternoon, and there are no drastic lighting changes at the other time conditions; while in traffic videos, the lightness is heavily affected by the lights of cars at duck and night, making the background modeling not work so well. Clearly, whenever at different periods in one day, IEEE 1857 surveillance groups will achieve higher coding efficiency if the background modeling can better capture the unchanged data in the scene.

It also should be noted that, the model-based coding technology in IEEE 1857 surveillance groups can be embedded into the cutting-edge HEVC/H.265 framework, so as to optimize its encoder for further increasing the coding efficiency and at the same time reducing the encoding complexity. To evaluate this conjecture, we performed an experiment on eight video sequences with four resolutions (i.e., SD, 720p, 1600x1200, and 1080p) in the PKU-SVD-A dataset. As shown in Tab. I(c), the experimental results show that compared with the HEVC test model HM with low-delay configurations, this model-based HEVC coding technology can averagely save 45.89% bits and 45.86% encoding time on these surveillance videos.

Recognition-friendly schema

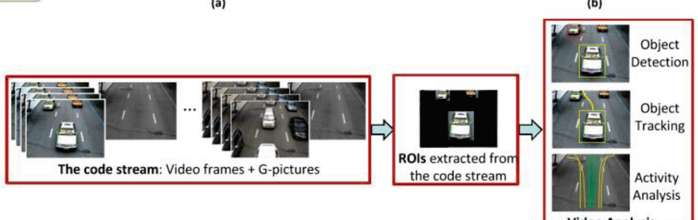
Often, surveillance video is captured not for better viewing experience, but for analysis and recognition of special objects or events. Traditionally, the principal objective of video coding is to reduce the compression rate of the whole frame while not distinguishing the background or ROIs in the frame. However, one of the most important prerequisites in video analysis and recognition is the relatively high fidelity (or clarity) of ROIs that often contain moving objects (e.g., pedestrians, vehicles) or faces. In this sense, IEEE 1857 surveillance groups have done a useful attempt to integrate the two seemingly contradictory subfields into a unified framework: On one hand, they utilize the background modeling in the coding loop to remove the "scenic redundancy" in consecutive frames, consequently achieving a higher coding efficiency; on the other hand, they should be the first video coding standard that provides the best supports to video analysis and recognition, including the syntax for describing ROI regions and camera parameters, high-clarity ROI coding, and object detection based on MB classification.



(a)

Syntax	Descriptor
region_extension() {	
extension_id	f(4)
region_number	ue(v)
for (i=0; i<RegionNum; i++) {	
marker_bit	f(1)
region_id[i]	ue(v)
region_event_id[i]	u(5)
marker_bit	f(1)
region_top_left_x[i]	u(10)
region_top_left_y[i]	u(10)
marker_bit	f(1)
region_bottom_right_x[i]	u(10)
region_bottom_right_y[i]	u(10)
}	
next_start_code()	
}	

(b)



(c)

Fig. 7. (a) The conceptual schema of the standard syntax in IEEE 1857 surveillance groups specially designed for video analysis and recognition tasks; (b) An example of the syntax for describing ROIs in the code stream; (c) A pipeline to demonstrate how to utilize ROIs extracted from the code stream in various video analysis tasks.

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

In IEEE 1857 surveillance groups, ROIs of attended objects can be coded into the code stream and thus can be directly extracted for further analysis and recognition tasks. Fig. 7(a) shows the conceptual schema of their standard syntax, which basically supports three levels of semantic descriptions of a video sequence, ranging from ROIs to objects and events. As an example, Fig. 7(b) illustrates the syntax that is used to describe ROIs in the code stream. These semantic descriptions, whatever are manually labeled or automatically extracted, can be used to video analysis tasks such as object detection and tracking (as shown in Fig. 7(c)). In addition, the standard syntax also supports the description of the parameters and positioning information (e.g., GPS) of the camera in the video stream. Usually, these data can facilitate many challenging visual tasks such as camera calibration, object tracking across cameras in a wide area.

Instead of utilizing some ROI-based resource allocation algorithms [7] to reach higher compression ratios without perceptually degrading the reconstructed ROIs in H.264/AVC, IEEE 1857 surveillance groups themselves can support the high-clarity coding for ROIs. That is, they can allocate more bits to the ROIs in the code stream without remarkably increasing the overall bitrate. With background modeling in the coding loop, the ROI detection and ROI-based bit allocation can be done automatically in the coding process. In this way, IEEE 1857 surveillance groups can effectively avoid the dilemma between high video compression ratios and high recognition accuracy both required for surveillance video systems.

Given G-pictures in the code stream, we can utilize an adaptive threshold to divide the MBs in the current frame into foreground units (FUs), foreground-background-hybrid units (HUs) and background units (BUs), where different kinds of units often correspond to different proportions of foreground. This MB classification can be used to improve the performance of object detection in both processing speed and accuracy. Our experimental results show that at the precision level of 90%, the object detection algorithm based on MB classification can obtain the improvement by recall of 6%, meanwhile saving 70% processing time.

In a broader sense, IEEE 1857 surveillance groups enable the possibility to embed more intelligence into the camera. Traditionally, cameras only act as the “eyes” of a video surveillance system, with the functionality of watching a given scene and continuously recording images about this scene. When equipped with model-based video coding, the cameras can begin to thinking about the scene by finding the “unchanged” data. As such, the static background (e.g., the street light poles) and periodic background (e.g., the scene changes each day) can be automatically learned from the data in a very long period of time (e.g., one or more months). Further with the help of more powerful object representation models, the camera even can identify or re-identify objects that frequently occur in the same scene. If it is true, such “compound eyes” will transform surveillance cameras from a data acquisition tool to information and intelligence acquisition systems, thus providing the whole smart surveillance systems with the ability to enhance situational awareness across multiple scales of space and time [8].

Conclusion

By exploiting model-based coding to remove the scenic redundancy, IEEE 1857 surveillance groups can achieve remarkably twice coding efficiency on surveillance videos over H.264/AVC HP. More importantly, they should be the most recognition-friendly video coding standard till to now, by enhancing their standard syntax for describing ROI regions and camera parameters and offering some powerful

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

technologies to support video analysis and recognition. Considering the amazing fact that surveillance videos will take nearly half of big data in the coming years [1], it is expected that IEEE 1857 surveillance groups can be used to a much wider range of surveillance video applications, thus presenting new opportunities and drives to the research communities and industries on smart video surveillance systems.

Acknowledgement

This work is supported by the National Basic Research Program of China under a grant No. 2009CB320900, and National Natural Science Foundation of China under a grant No. 61035001.

Reference

[1] J. Gantz and D. Reinsel, "The IDC Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," 2012.12. http://www.emc.com/leadership/digital-universe/index.htm.

[2] ITU-T and ISO/IEC JTC 1, "Advanced Video Coding for generic audio-visual services," in ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), May 2003.

[3] P. Felzenszwalb, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," *IEEE Trans Pattern Anal. Machine Intell.*, 32(9), Sep. 2010, 1627–1645.

[4] Y. Su, S.G. Shan, X.L. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Trans Image Processing*, 18(9), 2009, 1885-1896.

[5] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, 22(12), Dec. 2012, 1649-1668.

[6] T. K. Tan, G. J. Sullivan, and T. Wiegand, "Recommended Simulation Common Conditions for Coding Efficiency Experiments", *ITU-T Q.6/SG16, doc. VCEG-AA10*, Oct. 2005

[7] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-Interest Based Resource Allocation for Conversational Video Communication of H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, 18(1), 2008, 134-139.

[8] A. Hampapur, et al., "Smart Video Surveillance," *IEEE Signal Processing Magazine*, Mar 2005, 38-51.

[9] X. G. Zhang, Y. H. Tian, T. J. Huang, W. Gao, "Low-complexity and High-efficiency Background Modeling for Surveillance Video Coding," *Proc. IEEE Int'l Conf. Visual Communication and Image Processing*, Nov 2012, 1-6.

Wen Gao is a professor in the School of Electrical Engineering and Computer Science at Peking University, China. His research interests include image and video communication, computer vision, and artificial intelligence. Gao has a Ph.D. from the University of Tokyo. He is an IEEE fellow and an academician in the Chinese Academy of Engineering. Contact him at wgao@pku.edu.cn.

Yonghong Tian (corresponding author) is a professor in the School of Electrical Engineering and Computer Science at Peking University, China. His research interests include computer vision and multimedia content analysis. Tian has a Ph.D. in Computer Science from the Institute of Computing Technology at the Chinese Academy of Sciences, China. He is a senior member of IEEE and a member of ACM. Contact him at yhtian@pku.edu.cn.

Tiejun Huang is a professor in the School of Electrical Engineering and Computer Science at Peking University, China. His research interests include digital library and video coding and understanding. Huang has a PhD in Image

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE

Recognition and Artificial Intelligence from Huazhong University of Science and Technology, China. He is a senior member of IEEE and a member of ACM. Contact him at tjhuang@pku.edu.cn.

Siwei Ma is an associate professor in the School of Electrical Engineering and Computer Science at Peking University, China. His research interests include image and video coding, processing, and transmission. Ma has a Ph.D. in Computer Science from the Institute of Computing Technology at the Chinese Academy of Sciences, China. He is a member of IEEE. Contact him at swma@pku.edu.cn.

Xianguo Zhang is a researcher in MediaTec(Beijing) Inc. His research interests include video coding and processing. Zhang has a Ph.D. from the School of Electrical Engineering and Computer Science at Peking University, China. Contact him at zxgvideo@gmail.com.

Contact information

Wen Gao: Email: wgao@pku.edu.cn, Phone: 0086-10-62758116, Fax: 0086-10-62751638,
Address: Rm. 2615, Science Buildings 2, Peking University, 5 Yiheyuan Rd., Haidian, Beijing, 100871, China

Yonghong Tian: Email: yhtian@pku.edu.cn, Phone: 0086-10-62753817, Fax: 0086-10-62751638,
Address: Rm. 2641, Science Buildings 2, Peking University, 5 Yiheyuan Rd., Haidian, Beijing, 100871, China

Tiejun Huang: Email: tjhuang@pku.edu.cn, Phone: 0086-10-62758116, Fax: 0086-10-62751638,
Address: Rm. 2615, Science Buildings 2, Peking University, 5 Yiheyuan Rd., Haidian, Beijing, 100871, China

Siwei Ma: Email: swma@pku.edu.cn, Phone: 0086-10-62756172, Fax: 0086-10-62751638,
Address: Rm. 2641, Science Buildings 2, Peking University, 5 Yiheyuan Rd., Haidian, Beijing, 100871, China

Xianguo Zhang: Email: xgzhang@jdl.ac.cn, Phone: 0086-10-62753817, Fax: 0086-10-62751638,
Address: Rm. 2641, Science Buildings 2, Peking University, 5 Yiheyuan Rd., Haidian, Beijing, 100871, China

Digital Object Indentifier 10.1109/MIS.2013.101 0885-9000/$26.00 2013 IEEE