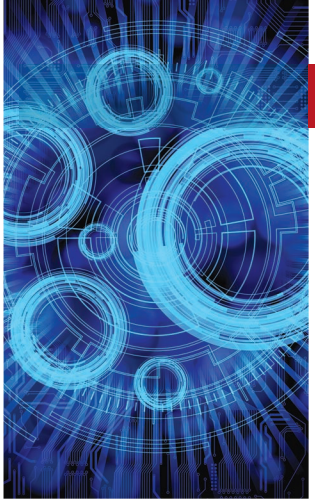

Editor: **Liuqing Yang**, Colorado State University, liuqingyang.ieee@gmail.com

IEEE 1857: Boosting Video Applications in CPSS

Tiejun Huang, Yonghong Tian, and Wen Gao, *Peking University*

Typically, a cyber-physical-social system (CPSS) features tight integration and coordination among computational (or cyber), physical, and human (social) elements.[1] In all types of information flows among the three spaces, video is definitely the one that takes the majority of traffic.[2] Yet, video hasn't received enough attention from the CPSS community. Nowadays, millions of surveillance cameras are being linked together via the broadband network, consequently leading to a "visual perception network" that's instituted virtually anywhere in the world. Such a world-scale perception network offers moment-by-moment pictures of both the physical world and human behavior over extended periods of time, providing a panoramic digital mapping of the running state of the world.

Obviously, there exists an enormous—and growing—gap between the amount of video data continuously collected by cameras and our ability to efficiently transmit, store, and intelligently analyze and digest this visual information. The IEEE 1857 standard released in June 2013, as the first standard that supports highly efficient surveillance video coding and objects-of-interest representation in the coding bitstream, can be used to narrow this gap, and consequently boost the various video applications in CPSS.

Bigger and Bigger Video Data in CPSS

From the silver-surfaced copper Daguerreotype plates introduced in 1839 to the electronic video camera tube invented in the 1920s, humans began a history of capturing and permanently recording the physical world. In the 1960s, closed-circuit televisions (CCTVs) and video recorders (VCRs) were first installed to monitor buildings, railways, and other public infrastructures. Years later, video surveillance systems appeared in banks and stores. By the 1990s, home security systems allowed users to remotely control a camera through a Web interface.

Recently, as the hosting cities of the 2008 and 2012 Olympic games, Beijing and London showcased the utility of large-scale video applications by deploying about 1 million surveillance cameras in public area. Intelligent transportation systems (ITS) is another typical CPSS field in which video cameras are employed to automatically monitor traffic flow, recognize vehicle license plates, detect violation behavior, and even identify people in the crowd. Today, we notice surveillance cameras in the elevator, on the ATM machine, along the sides of streets, in office buildings, and almost anywhere you can reach. A recent report from International Data Corporation (IDC) shows that half of the global Big Data in 2012 are surveillance video, and the percentage will increase to 65 percent in 2015.[3] Thus, in terms of fusing computational, physical, cognitive, and social domains, video is the most important—and also possibly the dominant—data that must be addressed in the CPSS study.

Challenges

For surveillance video, analysis can be done normally for purposes such as crime investigations, military intelligence, or consumer traffic pattern discovery. But from a CPSS viewpoint, surveillance video contains a great deal of information about society's operations day and night. Moreover, what began with manually monitored video displays has evolved into a variety of systems with automated processes to scan multiple video streams, detect events or objects of interest, and act on them. In the future, more intelligence will be embedded in the cameras and the surveillance cloud to automatically analyze what's happening in "video big data."

Technologically, there are many challenges involved in using CPSS to process video big data. Some of the foremost challenges involve how to

1541-1672/13/$31.00 © 2013 IEEE
Published by the IEEE Computer Society

efficiently transmit and store huge amounts of video data, and how to intelligently analyze and understand huge amounts of visual information. It's well known that there are many tools, algorithms, standards, and systems to deal with a single video sequence. But for millions of video streams captured around the clock, we need to rethink such challenges. For example, the data size of the surveillance videos captured each hour from the 1 million cameras in Beijing is larger than all the archived videos of China Central Television (the national TV station in China). According to IDC, surveillance video data will grow exponentially (more than 40 time the current amount) from 2010 to 2020.[3] However, our experience in the past three decades shows that, the video compression ratio of video coding standards doubles each decade—namely, from MPEG-2/H.262 in 1994 to AVC/H.264 in 2003 and then to HEVC/H.265 in 2013. Thus, if we still follow this technology roadmap, in the next decade, more than 20 times the storage will be needed to cope with the explosive growth of surveillance video data.

More importantly, the huge amount of surveillance videos could become another data tsunami. At this point, it's impossible for humans to monitor and analysis them. In the London underground bombing case in 2005, the suspect was identified after inspecting two weeks of surveillance video captured from the metro entries. In the recent Boston Marathon bombing case, the two suspects were identified from surveillance videos via a dramatically human-led endeavor. The fact is that many "wanted" people or criminal actions that were captured on video are missed, just because no sufficient human resource is devoted to scanning and analyzing all these videos. Thus, machine intelligence should be used to understand what's happening in surveillance videos, just as ITS does with license plate recognition and violation behavior detection.

IEEE 1857: New Features for CPSS

The newly released IEEE 1857 video coding standard is an effective attempt to address the two challenges for various video applications in CPSS and ITS. IEEE 1857, the Standard for Advanced Audio and Video Coding, was released in June 2013.[4] Despite consisting of several different groups, the most remarkable part of IEEE 1857–2013 is its surveillance groups, which not only achieve at least twice the coding efficiency on surveillance videos as H.264/AVC HP, but also are recognized as the most recognition-friendly video coding standards.

In contrast to other state-of-the-art standards, such as HEVC/H.265, IEEE 1857 surveillance groups are mainly inspired by the fact that most surveillance videos are often captured by stationary cameras that always stand toward the same scene for a long time. That is to say, there's usually similar background data among thousands of consecutive pictures. In this case, the background can be modeled and then exploited by the video codec to remove the "scenic redundancy" in consecutive pictures. In IEEE 1857 surveillance groups, the background model is represented as a specially encoded Intra-picture, or I-picture (called a *G-picture*, where the G is from backGround). As Figure 1 shows, instead of coding the residual data between the current frame and the recent reference frame, the IEEE 1857 surveillance groups code the difference data between the current frame and the G-picture, by using the difference data between the recent reference frame and the G-picture as a reference. Note that after being modeled in the initial stage, the G-picture will be updated regularly and thus remain approximately constant in the coding process of several groups of pictures (GOPs) that follow. Therefore, the total bit rate will be the same with the traditional coding method in the initial stage, and will decrease quickly in the coding process that follows. Overall, the IEEE 1857 surveillance group doubles the coding efficiency on surveillance videos compared to H.264/AVC.

Often, surveillance video is captured not for better visual experience in viewing or enjoying, but for analysis and recognition of attended objects or events. In this sense, the IEEE 1857 surveillance groups are the first video coding standard that provides the best support to video analysis and recognition. In the IEEE 1857 surveillance groups, the blocks could be classified as background units, foreground units, and background-foreground-hybrid units according to the similarity between them and the G-picture. The non-background units could be further analyzed and labeled as regions of interest (ROIs). The standard syntax of IEEE 1857 surveillance groups supports directly describing such ROIs in the coding bitstream. Then, these ROIs can be used for various video analysis tasks, such as vehicle detection, tracking, and activity analysis (see Figure 2). Additionally, the standard syntax also supports describing the camera's parameters, coordinates, and positioning data (such as GPS information) in the coding bitstream. Usually, these parameters can facilitate many challenging visual tasks, such as camera calibration, automatic vehicle localization, and tracking in multiple videos.

To further compress video data and analyze visual content, in our opinion,

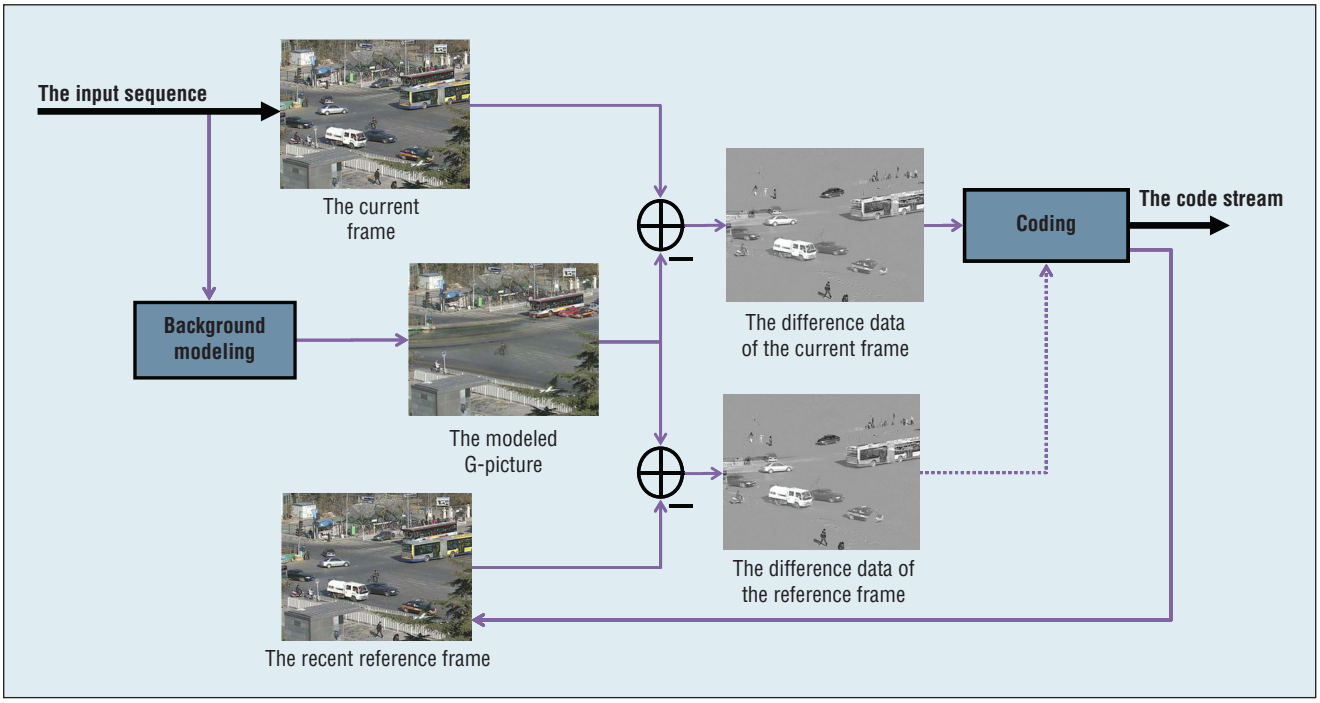

Figure 1. Doubling the coding efficiency of surveillance video by adding background modeling and the selective difference reference method in the coding framework. For an input surveillance video sequence, the encoding process is described as follows: a G-picture (a type of Intra-picture, where the G is from backGround) is first generated by the background modeling module for several groups of pictures (GOPs); then the G-picture and the recently decoded frame (called the recent reference frame) are naturally used as two references for coding each frame in these GOPs. Inter prediction can also be performed selectively between the two difference data that are calculated by subtracting the background data from the current coding frame and the recent reference frame.

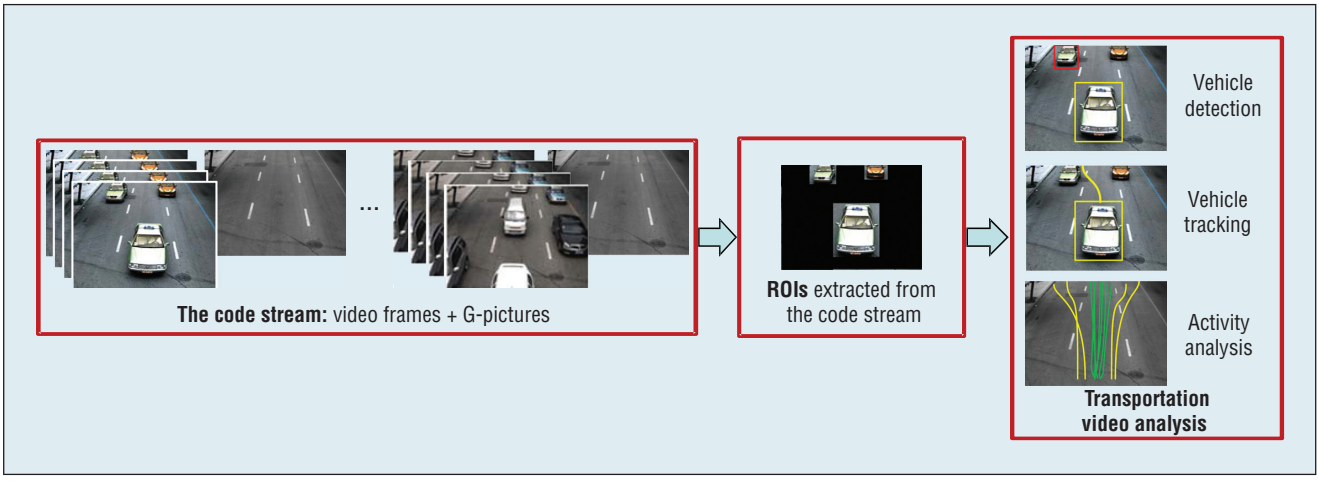

Figure 2. By utilizing the background frame in the video bitstream, the foreground objects (here, the moving cars) could be extracted directly and represented as regions of interest (ROIs). These ROIs are then used by intelligent transportation system (ITS) analysis tasks, such as vehicle detection, tracking, and activity analysis.

it's crucial to think long-term about video captured by surveillance cameras. What the IEEE 1857 standard has done is construct a pure background as a better reference frame for a group of frames (for example, 1,000 frames), while work in the future could involve building a background base that stores the most frequent backgrounds for a surveillance camera. Besides facilitating the compression, such a background base can also help understand

the scene's status—for example, daytime or nighttime, rain, snow, and so on. Once the machine knows more about the background, the moving objects could be detected and tracked more precisely. The camera can even identify or re-identify objects that frequently occur in the same scene. For example, a bus that periodically appears in a camera could be indexed, while a person who periodically appears in the view for several days could be regarded as a safe resident or a suspect to check in a potential crime location.

More generally speaking, it's clear that surveillance video is becoming a powerful bridge connecting the cyber, physical, and social worlds. Video systems are, in fact, mirror systems that map the physical world (including physical aspects of the social world) into cyberspace. Thus, the digitalized physical world (in the visual form) can be deeply analyzed in cyberspace. Then, these analysis results can in turn help optimize social systems and inspire the construction of future cyberspaces. With this in mind, video should be given more priority in future CPSS research and development. Despite various research results and tools that are available, more advances are still needed from the perspective of CPSS.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grants 61121002 and 61035001, and the National Basic Research Program of China under grant 2009CB320900.

References

1. F.-Y. Wang, "The Emergence of Intelligent Enterprises: From CPS to CPSS," *IEEE Intelligent Systems*, vol. 25, no. 4, 2010, pp. 85–88.
2. "Cisco Visual Networking Index: Forecast and Methodology, 2012–2017," white paper, Cisco, 29 May 2013; www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html.
3. J. Gantz and D. Reinsel, "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," *IDC iView*, Dec. 2012; www.emc.com/leadership/digital-universe/iview/index.htm.
4. *IEEE Std. 1857-2013, IEEE Standard for Advanced Audio and Video Coding*, IEEE CS, 4 June 2013.

Tiejun Huang is a professor in the School of Electrical Engineering and Computer Science at Peking University. Contact him at tjhuang@pku.edu.cn.

Yonghong Tian is a professor in the School of Electrical Engineering and Computer Science at Peking University. Contact him at yhtian@pku.edu.cn.

Wen Gao is a professor in the School of Electrical Engineering and Computer Science at Peking University. Contact him at wgao@pku.edu.cn.